



Forecasting Spatio-Temporal Variation in Residential Burglary with the Integrated Laplace Approximation Framework: Effects of Crime Generators, Street Networks, and Prior Crimes

Maria Mahfoud¹ · Wim Bernasco^{2,3}  · Sandjai Bhulai⁴ · Rob van der Mei^{1,4}

© The Author(s) 2020

Abstract

Objectives We investigate the spatio-temporal variation of monthly residential burglary frequencies across neighborhoods as a function of crime generators, street network features and temporally and spatially lagged burglary frequencies. In addition, we evaluate the performance of the model as a forecasting tool.

Methods We analyze 48 months of police-recorded residential burglaries across 20 neighborhoods in Amsterdam, the Netherlands, in combination with data on the locations of urban facilities (crime generators), frequencies of other crime types, and street network data. We apply the Integrated Laplace Approximation method, a Bayesian forecasting framework that is less computationally demanding than prior frameworks.

Results The local number of retail stores, the number of street robberies perpetrated and the *closeness* of the local street network are positively related to residential burglary. Inclusion of a general spatio-temporal interaction component significantly improves forecasting performance, but inclusion of spatial *proximity* or temporal *recency* components does not.

Discussion Our findings on crime generators and street network characteristics support evidence in the literature on environmental correlates of burglary. The significance of spatio-temporal interaction indicates that residential burglary is spatio-temporally concentrated. Our finding that *recency* and *proximity* of prior burglaries do not contribute to the performance of the forecast, probably indicates that relevant spatio-temporal interaction is limited to fine-grained spatial and temporal units of analysis, such as days and street blocks.

Keywords Predictive analytics · Forecasting · Bayesian spatio-temporal models · Centrality measures · Residential burglary

✉ Wim Bernasco
WBernasco@nscr.nl

Extended author information available on the last page of the article

Introduction

Over the past decades, improvements in criminological theory, crime data, computing power, and statistical methods have spurred interest in crime forecasting. The interest is certainly not limited to academics. Many police forces in the United States and around the globe have started to apply quantitative crime forecasting tools to support law enforcement operations (Hvistendahl 2016; Perry et al. 2013), an activity that has been labeled *predictive policing*. These tools are informed by large-scale and fine-grained spatio-temporal crime data. Spatio-temporal crime data are data on individual crimes that are geo-referenced (contain information on where the crime occurred) and time-resolved (contain information on when it occurred, at what date and at what time).

The description and explanation of spatial crime patterns is a classic research topic (Guerry 1833; Shaw and McKay 1942) that is still a thriving area of research, with contemporary developments emphasizing small spatial units of analysis (Weisburd 2015). The analysis of crime trends is a longstanding research field as well, and has recently been augmented by group-based trajectory analysis (Weisburd et al. 2004). However, the simultaneous variation of crime over time and space, or spatio-temporal variation, has been under-researched until recent times (Luan et al. 2016; Ratcliffe 2010). New research has started to demonstrate, for example, how spatial patterns of robbery vary across the hours of the day, the days of the week (Bernasco et al. 2017; Haberman and Ratcliffe 2015; Hipp and Kim 2019) and the seasons of the year (Andresen and Malleson 2013; Breetzke and Cohn 2012; Brunson et al. 2009; Ceccato 2005).

Advanced traditional methods of spatial description, such as crime mapping and hot-spot analysis (Eck et al. 2005; Weisburd and Lum 2005), focus on *where* crimes are most likely to occur. Recently developed methods explain and forecast not only *where* but also *when* crimes are most likely to occur. Their purpose is to detect emerging crime trends and spatial crime concentrations (Ratcliffe 2008). Reliable identification of where and when a crime is most likely to occur allows police departments to efficiently allocate patrols and other resources (Groff and La Vigne 2002), and shorten response times, and ultimately to disrupt and prevent crime, or reduce its most harmful consequences.

The purpose of the present paper is to improve the explanation and prediction of spatio-temporal variations in residential burglary. To this purpose, we use a modeling and forecasting framework (INLA) that is relatively new in criminology, and apply it to monthly patterns of residential burglary in the neighborhoods of a single district in Amsterdam, the Netherlands.

The choice of months as temporal units and neighborhoods as spatial units is not based on features of the INLA approach, but is motivated by the practical need for a mid-range forecasting tool to support the planning of operational officer deployment. The Amsterdam police and other police forces in the Netherlands conduct planning of operational officer deployment on a 4-weekly basis, and their patrolling activities are allocated to districts (Camacho-Collados and Liberatore 2015; Todovic et al. 2015). The 4-week forecast range applies in particular to planning shift rosters, which by law are required to be set and communicated to personnel at least 4 weeks in advance. Mid-range predictions are also necessary for planning vehicle maintenance and campaigns and interventions that require more than a few days of preparation.

Obviously, addressing residential burglary is only one of the many tasks of police officers, and only a single factor in the allocation of police resources. The present article is limited to residential burglary for two reasons. First, it is perceived as a high impact crime and

its prevention and investigation has been high on the list of priorities of the Dutch police for many years (Blokdijk and Beijersbergen 2020). Second, the focus on a single crime facilitates interpretation and requires less space. The frequency of occurrence and relative unambiguity of its location make residential burglary a good starting point to evaluate the applicability of the INLA approach to crime forecasting.

To inform mid-range planning decisions, crime forecasts should reflect appropriate temporal and spatial ranges. Thus, whereas the aim of our application of the INLA approach is similar to the aims of most recently developed predictive policing tools (i.e., to be able to allocate available police resources where they are most effective), it is applied to somewhat larger spatial and temporal units: *neighborhoods* and *months*.¹

Mechanisms Underlying Spatio-Temporal Patterns

In recent literature, two mechanisms have been identified as the underlying causes of spatio-temporal crime patterns. The first mechanism emphasizes the temporal constraints that characterize human behavior, in particular, the organization of human activities across daily, weekly, and seasonal time cycles. Time geography (Hägerstrand 1970; Miller 2005) is a theory that describes three types of constraints that humans face when they engage in activities across time and space. *Capability constraints* are biological constraints, such as the need to sleep, and technological constraints, such as a lack of transportation tools. *Coupling constraints* require people to be at specific places at specific times because of the social roles they fulfill. Most people have fixed school hours or work schedules they must adhere to. *Authority constraints* are limits to accessibility because place owners do not allow access, either temporarily or permanently. These three types of constraints also limit the discretion of potential offenders and victims, and they can help explain spatio-temporal crime patterns (Ratcliffe 2006). In particular, crime tends to be clustered spatio-temporally around opening times near retail businesses and other facilities. For example, whereas street robberies can take place any time anywhere, their incidence is elevated around high schools, but only at daytime and not on weekends (Bernasco et al. 2017; Haberman and Ratcliffe 2015). On a larger temporal scale, some studies have demonstrated spatial variations of the relationship between seasonality and violent crime (Breetzke and Cohn 2012; Ceccato 2005), crime in general (Andresen and Malleson 2013) or police calls for service (Brunsdon et al. 2009). Just as with daily and weekly time cycles, the spatio-temporal seasonal variations are related to variations in routine activities of the population, which in turn are affected by seasonal variations in temperature, precipitation and other weather conditions, but also by the timing of school holidays and other social institutions. These insights imply that for predicting monthly crime, seasonal effects on crime are likely to play out differently in some neighborhoods than in others.

The second mechanism emphasizes individual learning and habit formation by offenders. It operates in linear time rather than cyclical time and has been proposed as an explanation of *repeat victimization* and *near-repeat victimization*, two specific forms of spatio-temporal clustering for which massive empirical evidence has emerged over the past decade. Repeat victimization refers to the phenomenon that in the wake of a crime, the risk

¹ Months do not align perfectly with 4-week periods. For practical reasons of data availability and because the research was not meant to have its results immediately implemented, we used monthly data rather than 4-weekly data.

of repeated victimization of the same target is temporally heightened (Pease et al. 1998). Near-repeat victimization refers to the extension of this heightened risk to targets in the proximity of the original target (Bowers and Johnson 2004; Townsley et al. 2003). Thus, near-repeat victimization implies that in the wake of a crime, potential targets near the initial target also have a heightened risk. The heightened risk decays over time back to its baseline, and in the case of near-repeats it also decays over space, i.e., the post-crime elevated risk is highest at the target of the initial crime, and decreases with distance from the initial target (Johnson et al. 2007). The near-repeat pattern has also been labeled *communicability of risk* (Johnson and Bowers 2004), *selfextinguishing point process* (Mohler et al. 2011) and *spatio-temporal interaction* (Grubestic and Mack 2008).

Empirical research on detected burglary cases has revealed that in repeat and near-repeat victimizations, the same offenders are often involved in both the primary and the subsequent burglaries (Bernasco 2008; Everson and Pease 2001; Johnson et al. 2009). Moreover, many subsequent burglaries are committed by other offenders who are connected to the initial offenders of the primary burglary (Lantz and Ruback 1997), and who may have been informed by the initial offenders about local opportunities. The spatio-temporal clustering of burglary thus appears to be driven by the tendency of offenders to commit subsequent burglaries against previously targeted properties and against properties in the immediate vicinity. The probability of return increases with the number of prior burglaries and decreases with the duration since the latest burglary (Bernasco et al. 2015; Lammers et al. 2015). This behavioral tendency of burglars conforms to a recently proposed general model of human mobility, which states that individual whereabouts are driven by two generic mechanisms. The first is *preferential return*, a propensity to return to locations visited before. The second is *spatial exploration*, the tendency to explore additional locations in the vicinity of locations visited before (Song et al. 2010). These spatial preferences may be examples of the more general behavior principle that successful choices are likely to be repeated and lead to the formation of habits and routines.

The heightened risk of burglary in the wake of and nearby a prior burglary implies that when forecasting monthly burglary frequencies, it may be important to take into account how many crimes have been committed in the preceding months in the focal neighborhood and in nearby neighborhoods, as these crimes may elevate the current burglary risk in the focal neighborhood.

Modeling Spatio-Temporal Crime Patterns

Two broad classes of spatio-temporal analysis methods can be distinguished, testing-based methods and model-based methods (Luan et al. 2016). The purpose of testing-based methods is to assess whether spatio-temporal interaction is present in the phenomenon under study. They have been developed in the context of epidemiology, where space-time interaction in disease outbreaks is potentially indicative of contagiousness, and provide a single global test statistic that indicates whether significant spatio-temporal interaction is detected. Examples in the study of crime patterns include the Knox test (Townsley et al. 2003) and the Mantel test (Johnson and Bowers 2004). The purpose of model-based methods is to explain and potentially forecast spatio-temporal variations in the phenomenon under study. They are conceptually similar to regression models (Luan et al. 2016) as they estimate the effects of spatial and temporal variables and space-time interactions on the outcome variable of interest, i.e., crime. Examining space-time interactions might help to uncover crime patterns, which can be overlooked in a purely spatial or temporal analysis

and which can possibly uncover long-term changes in local social-economic conditions, changes to drug markets, or changes in policing strategy.

Bayesian models have become a primary tool for explaining simultaneous variation across space-time (Opitz 2017). The application of Bayesian methods to spatio-temporal data started to appear around the year 2000 following the development of the Markov Chain Monte Carlo (MCMC) simulation (Blangiardo and Cameletti 2015), and combine observed data with prior information to estimate posterior probability distributions of model parameters, including the space-time interactions (Luan et al. 2016). A couple of previous studies applied Bayesian MCMC methods to property crime (Law et al. 2014), violent crime (Law et al. 2015) and burglary (Hu et al. 2018).

The main advantage of the Bayesian method lies in its flexibility, its capabilities of dealing with issues, for example, missing data, sparse data, measurement errors and incompatible data (Law et al. 2014), and its taking into account uncertainty in the estimates/prediction (Blangiardo and Cameletti 2015). However, the downside to the MCMC is that it is computationally challenging because it requires extensive simulation (Luan et al. 2016). For example, a large dataset will need a number of days computing time to perform Bayesian inference via MCMC (Blangiardo and Cameletti 2015).

The purpose of the present study is to explore spatial, temporal and spatio-temporal effects in the incidence of residential burglaries in Amsterdam, the Netherlands, by using an alternative and computationally less challenging Bayesian method, the Integrated Nested Laplace Approximation (INLA), and to evaluate its efficacy as a method of forecasting the monthly frequency of burglaries per neighborhood.

Data

The study area is police district 'West' of the city of Amsterdam, the Netherlands. The district consists of twenty neighborhoods (see Fig. 1, where each neighborhood is labeled with a 4-digit number). These twenty neighborhoods are the spatial units in the present analysis.

Our analysis involves a combination of three geo-referenced data sources. The first contains information on crime incidents, the second on land use and population, the third on street network structure.

Each source features different geometrical elements (points, polygons and lines, respectively), but their aggregation to the neighborhood level is straightforward. The next three subsections provide brief descriptions of these three datasets and how their contents were aggregated to the neighborhood level.

Crime

Crime data were obtained from the Dutch Police. They contained monthly numbers of residential burglaries, assaults, street robberies and commercial robberies that had been recorded from January 2009 until December 2012 in the 20 neighborhoods shown in Fig. 1. In our analysis, residential burglary is the dependent variable. Some of the other crime variables were considered as covariates in the analysis.

To protect the privacy of victims, and before giving us access to the data, the police aggregated the coordinates of the exact crime locations into grid cells of 125 by 125 m. Here we use the midpoints of the grid cells as the approximate crime locations. The crime

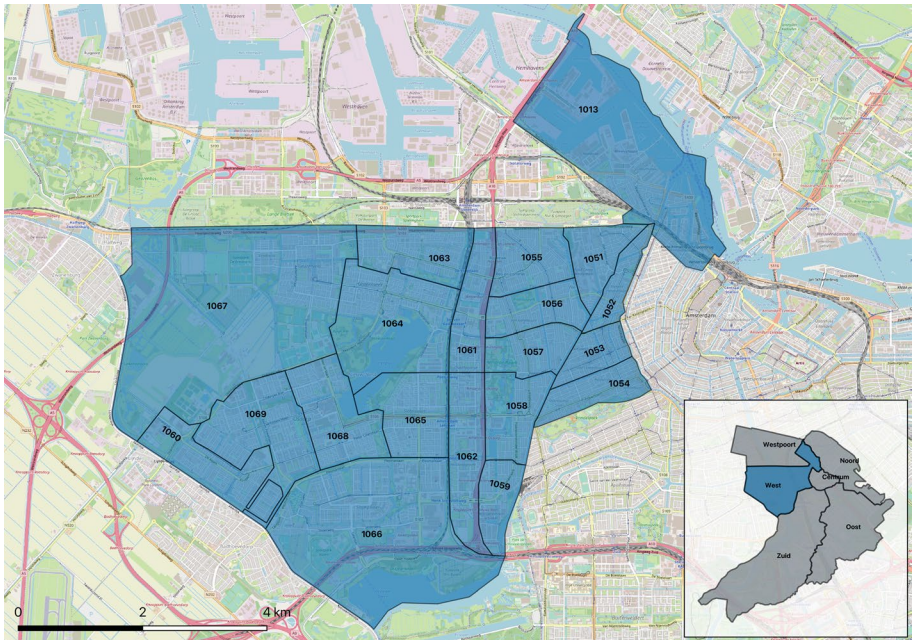


Fig. 1 Map of the neighborhoods in Amsterdam West

locations are thus ‘snapped’ onto the nearest grid cell midpoint, which leads to small geocoding errors: the mean geocoding error is 48 m and the maximum geocoding error is 88 m. The geocoding error is mostly inconsequential because all crime locations were aggregated to the neighborhood level.

In Fig. 2, which we include to illustrate the spatial aggregation of crime locations and street intersections to neighborhoods, the stars represent the grid points of these approximate crime locations.

Land Use and Population

Data on land use and population were obtained from the Municipality of Amsterdam and Statistics Netherlands (CBS). It includes demographic and socio-economic characteristics of the residential population, such as average monthly income, and land use indicators, such as numbers of schools, retail stores and restaurants.

The spatial unit of the land use and population data is the (six-digit) postal code area. On average, a postal code area in Amsterdam is roughly the size of a football field and contains approximately 18 residential properties and 40 residents. As the postal code system was created to facilitate post delivery, a single postal code is nearly always on the same street, applies to adjacent properties, and is not subdivided by physical barriers that impede pedestrian or car transportation (Bernasco 2010). The majority of postal codes refer to either one or both of the two-block faces on both sides of a street between two intersections (i.e., street segment, street block or face block, Weisburd et al. 2004).

Neighborhoods are supersets of six-digit postal code areas (the first 4 of the 6 digits indicate the neighborhood). To aggregate the population and land use characteristics of

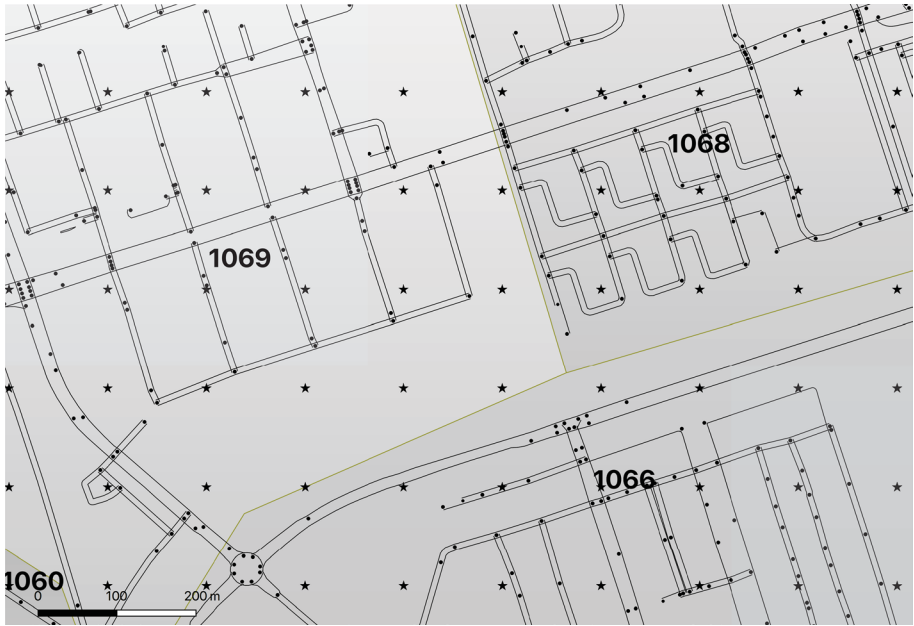


Fig. 2 An illustration of the different spatial units of the data. Stars represent approximate crime locations (grid points), the dots represent street intersections. Shades of grey and 4-digit labels indicate neighborhoods

postal code areas to neighborhoods, we used the unweighted sums of count variables (e.g., number of retail stores) and the unweighted averages of continuous variables (e.g., average household income).

Street Network Centrality

The third dataset contains two network centrality measures calculated on the street network of Amsterdam. In the analysis of transportation, the urban infrastructure of streets is modeled in terms of a network or *graph* that consists of *nodes* (also referred to as *vertices*) connected to each other by *edges*. Centrality measures are generic network measures that allow the analyst to quantify the relative importance of each node in the network. In transportation research, they quantify the expected size of traffic flow through a street segment or an intersection (Mahfoud et al. 2018). In the present paper, we use *closeness centrality* and *betweenness centrality*.

First we will describe how the centralities are created. This was done using two maps, the Google place id map received from the municipality of Amsterdam and the Hoofd en plusnetten map from the Ministry of Infrastructure and Water Management. The Google place id map was used to provide information on the directionality of the road segments, while the Hoofd en plusnetten map provided information on the maximum road speed. These maps were filtered to display only roads accessible for cars,

leaving out footpaths, bicycle and public transport lanes. Then, using a map matching algorithm, the maps were combined.

Using this map (for an example, see Fig. 2), a network was constructed combining all the street segments by use of an algorithm that forms links between two intersections using the following five steps:

1. Intersections are created by aggregating segments by longitude and latitude values of starting and ending points, and by the number of connecting segments to each of those starting and ending points.
2. To define where street intersections or nodes are, all unique segments less than or equal to 2 are filtered.
3. The single segments that connect two street intersections directly are counted as links.
4. The segments that connect street intersection A to street intersection B and then back to street intersection A are counted as links.
5. The remaining segments are connected one by one iteratively to a link starting from a street intersection until another street intersection is reached.

This procedure is continued until all segments are assigned to links and the street network is created. A more elaborated explanation of the construction of the network is provided by Erkin (2017).

In Fig. 2, which illustrates the spatial aggregation of crime locations and street intersections to neighborhoods, the small dots represent street intersections.

Closeness centrality (CC),² introduced by Bavelas (1950), is used to identify influential nodes in the network. In communications research, it measures how long it will take to spread information from a specific node to all other nodes in the network. In transportation networks, closeness can be regarded as a measure of the accessibility of a node from all other nodes. For example, if nodes define street segments, a street segment that is close to many other street segments will have a high closeness value and will likely experience more traffic than a street segment in the periphery of the network and thus be relatively isolated.

Betweenness centrality (BC),³ introduced by Freeman (1977), quantifies the relative importance of a node by counting how often it forms a bridge along the shortest path between two other nodes. In communications, it reflects the amount of control an individual has on communications between others in the network. In transportation, it measures whether a street segment or intersection forms a bridge that would partition the transportation infrastructure and constrain traffic when removed (Mahfoud et al. 2018). Formal

²

$$CC(v) = \frac{1}{\sum_{d(u,v) < \infty} d(u, v)}, \quad (1)$$

where $d(u, v)$ is the distance between the nodes u and v .

³

$$BC(v) = \sum_{u \neq w \neq v \in V} \frac{\sigma_{u,w}(v)}{\sigma_{u,w}}, \quad (2)$$

where $\sigma_{u,w}$ is the total number of shortest paths between node u and w . And $\sigma_{u,w}(v)$ is the total number of shortest paths between node u and w that pass through v .

definitions of closeness centrality and betweenness centrality as used in this paper are provided in footnotes 2 and 3.

Note that there are different ways by which a street network can be transformed into a mathematical concept of a graph. In the *primal representation*, street intersections are nodes and they are connected by street segments, which are the edges. This representation is used, for example, by Mahfoud et al. (2018), and also in the present research. In the *dual representation*, the roles are reversed so that street segments are the nodes and street intersection are the edges. This representation is more common in criminology and is used, for example, in Frith et al. (2017).

The difference between the dual and primal representations is not relevant in our analysis, however, because both centrality measures are aggregated to the neighborhood level. The neighborhood network measures of centrality are calculated as the average closeness and betweenness centrality measures of all intersections located inside the boundaries of the neighborhood.

There are substantive reasons for not calculating centrality measures on the complete street network, but to limit the calculation to street segments or intersections located within a given distance or travel time (Davies and Johnson 2015; Frith et al. 2017). To capture centrality in terms of both local pedestrian foot travel and short-distance vehicular traffic, we limited the calculation of the closeness and betweenness centrality measures to street segments from which the focal location could be reached within five minutes by motorized vehicles (Mahfoud et al. 2018). Computational tractability is an additional argument, because it is computationally very expensive to calculate exact centrality measures on complete street networks.

Data Analysis

Including dozens of potentially redundant variables in a training model can lead to identification problems. As a consequence, the resulting model may become oversensitive to the training data and be difficult to generalize. Therefore, it is important to limit the set of covariates to be included in the model selection. As an initial step in the analysis, and based on pair plots and on Pearson correlation coefficients (rather than substantive arguments relating to burglary), neighborhood characteristics with a high correlation (> 0.7) were omitted from the analysis. For details, see Mahfoud et al. (2017).

The resulting start set of the analysis consists of 22 covariates, including neighborhood, year, month, police subdistrict, number of youth centers, number of retail stores, number of educational institutions, residential population, percentage of single households, number of persons generating income, average monthly income, total numbers of assaults, street robberies and commercial robberies during the preceding 3 months.

Next, we divided the data set in a training set and a test set. The first 3 years of data were used as a training set and the last year as a test set. We applied the data analysis to the training set while the test set is kept untouched for validation purposes.

First we assessed the training set for outliers and collinearity. The presence of outliers was graphically assessed by the Cleveland dot plot and analytically by the Local Outlier Factor (LOF) with 10 neighbors and a threshold of 1.3 (Breunig et al. 2000). Results of this analysis show that the training data exhibits a percentage of outliers of 7.6. The majority of these occurred in December and January due to the daylight-darkness effect (Coupe and Blake 2006), Christmas and New Year's Eve. Due to the nature of burglary data and the high percentage of outliers, we decided to perform the analysis without removing or

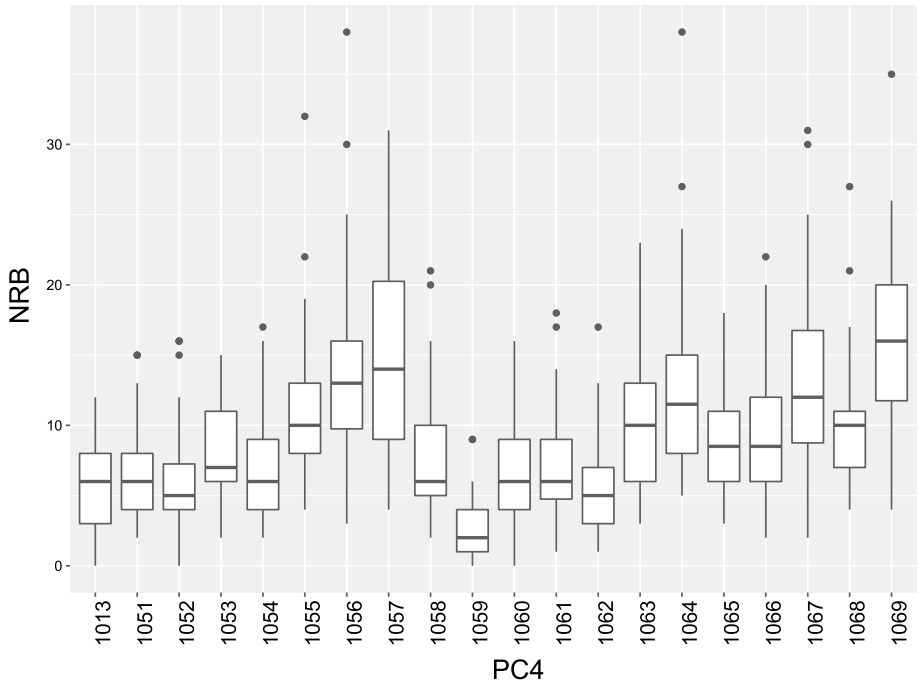


Fig. 3 Conditional boxplot of the number of burglaries by neighborhood (PC4). This plot indicates heterogeneity in the number of burglaries within the different neighborhoods

pre-processing them. And also because the methods that we want to use do not support gaps in the data.

The collinearity was assessed using the variance inflation factor (VIF), which measures the amount by which the variance of a parameter estimator is increased due to collinearity with other covariates. Using all the covariates the VIFs were calculated. Then the covariate with the highest VIF was removed from the data set and the VIFs were calculated again. This was repeated until all VIFs are below the threshold of 3 (Zuur et al. 2007), resulting in a final set of covariates containing the following covariates: the temporal covariate month; the number of educational institutions (sEI), the number of retail stores (sRET), percentage of single-person households (aSH), the number of persons that generate income (sNPI), the total observed street robberies in the neighborhood and adjacent neighborhoods over the last 3 months (sMuGL3M) and finally, the average monthly income (aAMI).

Furthermore, the relationship between residential burglaries and the categorical covariates was assessed using conditional box plots. Results show a temporal monthly effect and a spatial neighborhood effect on the burglaries, as shown by Figs. 3 and 4.

Space-Time Modeling Approach

The first law of geography, introduced by Tobler (1970), affirms that “everything is related to everything else, but near things are more related than distant things” (Tobler 1970). Spatial nearness is usually specified by defining pairs of adjacent neighborhoods. Once

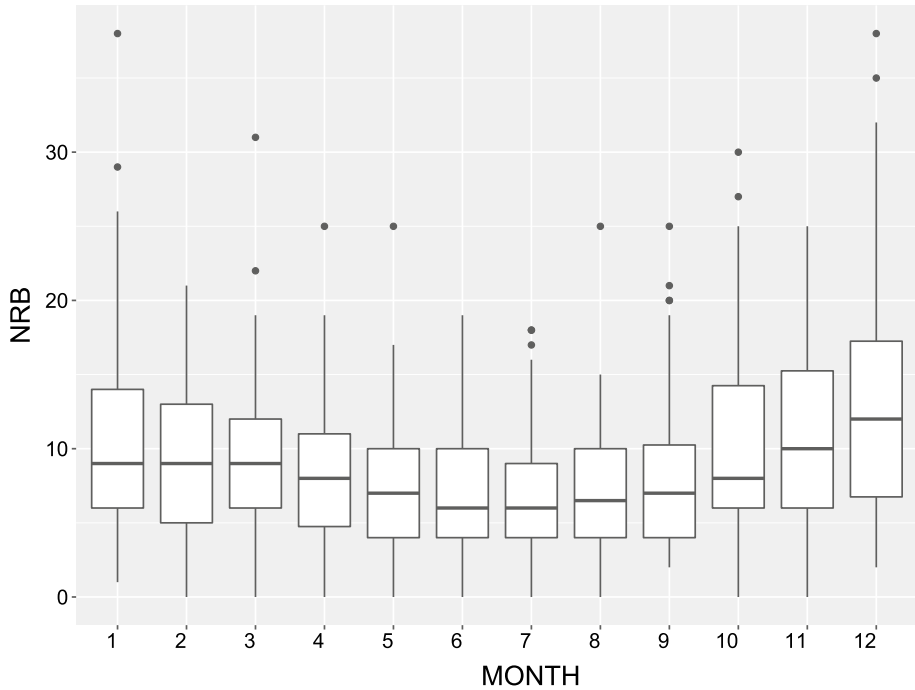


Fig. 4 Conditional boxplot of the number of burglaries by month of year. This plot indicates heterogeneity in the number of burglaries within the different months

adjacency pairs are defined, the map is reduced to a lattice or graph (i.e. network of neighborhoods), with the neighborhoods represented as vertices or nodes (and usually measurements being attached to each node), and edges between nodes representing pairs of adjacent neighborhoods (Hodges 2016).

In general, it is assumed that neighborhoods i and j are defined as neighbors, denoted by $i \sim j$ if they have a common border, and the set of adjacent neighborhoods of neighborhood i is denoted as $\mathcal{N}(i)$ and its size by \mathcal{N}_i . One of the most used models that account for spatial correlation is the Besag model, also called the intrinsic Gaussian Markov random field (IGMRF) or the intrinsic conditional autoregressive (iCAR) model (Besag et al. 1991). However, this model takes only the spatially structured effects into account, which results in misleading parameter estimates as the unstructured random error will also be modeled as spatial correlation. In a prior paper (Mahfoud et al. 2019), we used the Besag–York–Mollié (BYM) model that combines the iCAR specification with neighborhood-specific effects modeled as exchangeable to account for the spatial correlation.

Tentatively suggesting that Tobler’s First Law of Geography also applies on the temporal dimension (“things near in time are more related than things distant in time”), in the present paper, we extend the Besag–York–Mollié (BYM) model by a space-time interaction δ_{it} , that would explain the differences in the time trend of burglaries between the different neighborhoods. In this model, the linear predictor η_{it} is modeled by means of an intercept, a structured neighborhood-specific effect (u_i) and an unstructured neighborhood-specific effect (v_i), a temporally structured effect (γ_t) modeled by a Random Walk process and a temporally non-structured effect modeled as exchangeable among the time periods.

Blangiardo and Cameletti (2015) discuss extensively four types of interactions. These interactions concern the interaction between one of the spatial components (structured or unstructured) with one of the temporal components (structured or unstructured). The simplest interaction (US-UT) is the one between the unstructured components of space and time. In this interaction, it is assumed that there is no spatial or temporal structure on the interaction. The second type of interaction (US-ST) concerns the interaction between the unstructured component of space with the structured component of time. In this interaction, it is assumed the parameter vector of the space-time interaction $(\delta_{it}, \dots, \delta_{iT},)$ has an autoregressive structure on the temporal component. This temporal structure is independent between the neighborhoods. The third type of interaction (SS-UT) involves the structured component of space and the unstructured component of time. In this interaction, it is assumed that the parameters at the t -time point $(\delta_1, \dots, \delta_n)$ have a spatial structure defined through the CAR specification. This structure is independent between the other time points. In the fourth type of interaction (SS-ST), it is assumed that the structured component of space interacts with the structured component of time. In this case, the temporal dependence structure of each neighborhood depends on the temporal pattern of the adjacent neighborhoods. All these four types of space-time interactions will be considered and the model can be formulated as:

$$\begin{aligned}
 y_{it} &\sim \text{Poisson}(\mu_{it}); \mu_{it} = E_i \rho_{it} \quad \log(\rho_{it}) = \eta_{it} \\
 \eta_{it} &= b_0 + u_i + \nu_i + \gamma_t + \phi_t + \delta_{it} \\
 \nu_i &\sim N(0, \tau_\nu) \\
 u_i | \{u_j; j \neq i\}, \tau_u &\sim N\left(\frac{1}{N_i} \sum_{j=1}^n a_{ij} u_j, \frac{1}{N_i \tau_u}\right) \\
 \gamma_t | \gamma_{t-1}, \gamma_{t-2} &\sim N\left(2\gamma_{t-1} + \gamma_{t-2}, \frac{1}{\tau_\gamma}\right) \\
 \phi_t &\sim N(0, 1/\tau_\phi)
 \end{aligned} \tag{3}$$

where E_i is the population in neighborhood i , and is used as an offset term. An offset is used as a correction factor in the model specification to change the scale, because the interest lies more in the rates of relative risk on burglary than in the average number of burglaries. b_0 is the average burglary rate in all neighborhoods. $\eta_{it} = \log(\rho_{it})$ is the linear predictor, and the average number of burglaries μ_{it} is $E_i \rho_{it}$. \mathbf{u} is the spatially structured component with precision τ_u , modeled by means of a first-order intrinsic Gaussian Markov random field (Rue and Held 2005). ν and ϕ are zero-mean white noise processes with precisions τ_ν and τ_ϕ , respectively. a_i is an indicator variable that is equal to 1 if neighborhoods i and j are adjacent. Different studies show that often repeat and near repeat victimization tend to occur within days or weeks after the first burglary victimization (Johnson et al. 2007). For this reason, we choose a random walk model with a lag of 2 months to model the temporally structured effect of time.

In the first instance, the models were fitted without covariates. Then the models were extended to account for the effect of different neighborhood characteristics related to demographic, socio-economic and street network characteristics. Finally, the models with and without covariates were compared with each other and with the spatial model and the spatio-temporal model without interaction, as discussed in Mahfoud et al. (2019).

The models were fitted using the Integrated Nested Laplace Approximation (INLA) developed by Rue et al. (2009) and implemented in the R-INLA package. INLA is a deterministic algorithm for Bayesian inference specially designed for the class of Latent Gaussian Models (LGM). It was first introduced by Rue et al. (2009) and its fundamental idea consists of applying the Laplace Approximation (LA) technique to perform accurate and fast Bayesian inference on LGM. To ensure fast and accurate approximation of the posterior densities, INLA makes use of a computationally efficient class of Gaussian processes, namely the Gaussian Markov random field (GMRF) (Opitz 2017; Wang et al. 2018). INLA is thus based on three key components, namely the LGM framework, the GMRF and the principal of the LA. The main ideas of each key component are summarized in the “Appendix”.

For fitting the models within a Bayesian framework, we need to specify prior distributions for the fixed parameters and the hyperparameters. For both, we used the default parameters used by R-INLA. INLA makes use of a normal diffuse priors with zero mean and a precision equal to 0.001. This corresponds to a variance equal to 32^2 . This means that 99%, which corresponds to 3 times the standard deviation, of the values of the fixed parameters are expected to lie between -96 and $+96$. These diffuse priors are used for all fixed parameters except for the intercept where the used precision is equal to zero to allow for large standard deviations (Zuur et al. 2017). For the hyperparameters in R-INLA, we need to specify the prior distribution for the log of the precision. The R-INLA default is the log gamma distribution with shape parameters $a = 1$ and inverse scale parameter $b = 0.00001$. To ensure that the results are not sensitive to the specification of the prior distribution of the hyperparameters, the models were also fitted with other hyperprior distributions. Carroll et al. (2015) advise the use of a $gamma(1, 0.5)$ for the hyperparameters for a Poisson GLMM.

Model Selection and Model Validation

To be able to compare the performance of the models, we used methods based on the deviance as on the predictive distribution. First, we used the deviance information criterion (DIC), proposed by Spiegelhalter et al. (2002). The DIC is defined as:

$$DIC = D(\bar{\theta}) + 2p_D, \quad (4)$$

where $D(\bar{\theta})$ is the deviance using the posterior mean of the parameters and p_D is the effective number of parameters. As the posterior marginal distributions of some hyperparameters, especially the precisions might be highly skewed, INLA evaluates the DIC at the posterior mode of the hyperparameters and for the latent field INLA uses the posterior mean (Blangiardo and Cameletti 2015). We used also the Watanabe-Akaike information criterion (WAIC). This criterion is based on data partition. From the methods based on the predictive distribution, we used the conditional predictive ordinate (CPO) (Pettit 1990). CPO is a leave-one-out cross-validation relying on splitting the data into two exclusive sets, a data set used for fitting the model (\mathbf{y}_f) and a test set (\mathbf{y}_c) used for performing model criticism (Blangiardo and Cameletti 2015; Zuur et al. 2017). In the leave-one-out cross validation, it is assumed that $\mathbf{y}_f = \mathbf{y}_{-it}$ and $\mathbf{y}_c = \mathbf{y}_{it}$ and the CPO is given by:

$$CPO_{it} = p(y_{it}^* | \mathbf{y}_f), \quad (5)$$

where y^* represents a future occurrence. Actually, the CPO gives the probability of an omitted observation given all data. Note that the CPO is calculated for each observation, and

Table 1 DIC, WAIC, and the LCPO of the different models including and excluding covariates

Model	DIC		WAIC		LCPO	
	excl.cov	incl.cov	excl.cov	incl.cov	excl.cov	incl.cov
Spatial	4189	4179	4208	4199	2.922	2.916
Space + time	3995	3992	4022	4019	2.793	2.792
Space × time US-UT	3807	3806	3810	3808	2.718	2.718
Space × time US-ST	3967	3967	4011	4011	2.788	2.788
Space × time SS-UT	3984	3978	4039	4033	2.808	2.803
Space × time SS-ST	3968	3968	4012	4012	2.788	2.788

that a higher CPO value indicates better model fit. In order to compare the models based on CPOs, a summary measure based on the sum of the log of the CPO values, \overline{LCPO} given by Eq. (6) is calculated (Carlin and Louis 2008; Roos and Held 2011). For all three statistics, the DIC, the WAIC, and the \overline{LCPO} , lower values indicates a better model fit. There do not exist significance tests for numerical differences between two values of these statistics; it can only be concluded that the lower value represents a better fit, but not whether the lower value is ‘significantly’ better than the higher value.

$$\overline{LCPO} = -\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \log(CPO_{it}). \quad (6)$$

To assess the predictive performance of the models, the Root Mean Squared Error (RMSE) and the Weighted Absolute Percentage Error (WAPE) are calculated using an out-of-sample data set. If y_{it} denotes the realization in neighborhood i and in month t , and \hat{y}_{it} denotes the forecast in the same neighborhood and in the same month, then the forecast error is given by $e_{it} = y_{it} - \hat{y}_{it}$. The RMSE and the WAPE are given by Eqs. (7) and (8), respectively.

$$RMSE = \sqrt{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{it}^2}, \quad (7)$$

$$WAPE = \frac{\sum_{i=1}^N \sum_{t=1}^T |e_{it}|}{\sum_{i=1}^N \sum_{t=1}^T y_{it}}. \quad (8)$$

Results

In this section, the results of the discussed models will be presented.

Model Fit and Model Validation

In this subsection, we will discuss the results from fitting the models discussed in Sect. 3. First, the models were fitted with and without covariates. The DIC and the WAIC were

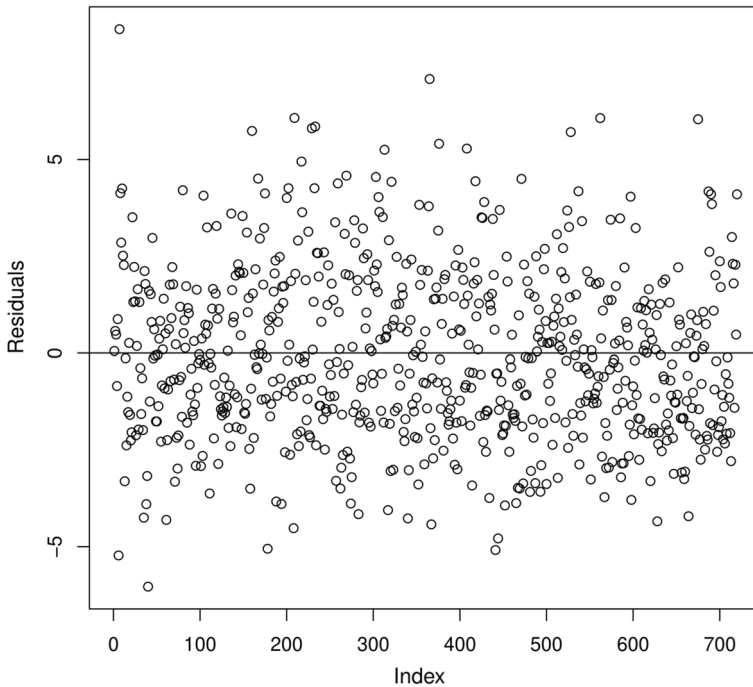


Fig. 5 Residuals of the optimal model including covariates

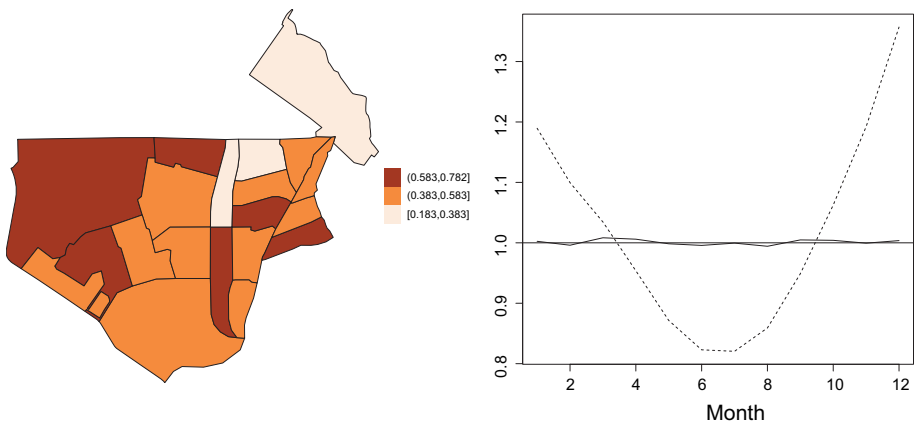
calculated for the different models. As can be seen from Table 1, the spatio-temporal model with an interaction between the unstructured components of space and time (Space time interaction US-UT, also called the unstructured interaction model) results in the lowest DIC, WAIC, and \overline{LCPO} values. This model can be regarded as the best model among the other models. However, there are minor differences in the performance of this model when including or excluding the covariates. Including neighborhood characteristics makes a clear model improvement only for the spatial model and the spatio-temporal model with the third type of interaction (SS-UT).

The adequacy of the unstructured interaction model is assessed graphically by drawing the Pearson residuals of this model. The residual plot is given in Fig. 5. It is clear that the residuals are scattered around zero. Based on this plot, there is no indication of over-dispersion or under-dispersion. Calculating an estimate of the dispersion parameter reveals that the model is slightly under-dispersed, which means that the observed variability is slightly lower than the expected variability.

The estimated parameters of the unstructured interaction model on the logarithmic scale are presented in Table 2. From this table, we can see that the number of educational institutions (sEI), the number of persons generating income (sNPI) and the average betweenness are not important, as the zero lies within their 95% credible intervals. The number of retail stores in the neighborhood, the number of street robberies, and the average closeness have a positive effect on burglaries. The number of households with a single households has a negative effect on burglaries. To assess the exact effect of these covariates on residential burglaries, we converted the posterior distributions from the logarithmic scale to the

Table 2 The posterior mean and 95% credible intervals of the fixed effects on the logarithmic scale of the unstructured interaction model US-UT, including covariates

Estimate	Mean	SD	0.025 quantile	0.975 quantile
Intercept	- 7.519	0.020	- 7.559	- 7.479
Educational institutions	0.009	0.028	- 0.046	0.065
Retail stores	0.081	0.029	0.024	0.139
Single-person household	- 0.267	0.045	- 0.357	- 0.179
Persons generating income	- 0.049	0.038	- 0.124	0.026
Street robberies	0.050	0.021	0.008	0.090
Betweenness	- 0.054	0.060	- 0.168	0.067
Closeness	0.157	0.053	0.052	0.260

**Fig. 6** Posterior mean of the excess risk $p(\exp(u + v) > 1|y)$ and the posterior temporal trend of burglaries of the optimal model including the covariates. The dashed line represents the temporally structured effect ($\exp(\gamma_t)$) and the solid line represents the unstructured effect ($\exp(\phi_t)$)

original scale of the data. Then we calculated the posterior mean and the 95% credible intervals on the original scale. This analysis reveals that an increase in one unit in the number of retail stores, the number of street robberies and in the closeness is associated with an increase of 8.51%, 5.10%, and 17.13%, respectively, in the risk of burglary. Among all covariates, the closeness has the most impact on the risk of residential burglaries. In contrast, an increase in one unit of the average number of households with a single households results in a decrease in the risk of burglaries of 23.32%.

The posterior mean of the excess risk on burglary for each neighborhood after taking account of the covariates is expressed in Fig. 6. As can be seen from this figure, the neighborhoods 1054, 1057, 1062, 1063, 1067, and 1069 have a higher excess risk on burglaries compared to the other neighborhoods. The neighborhoods 1013, 1055, and 1061 have the lower excess risk. The same figure shows the posterior mean of the temporal effect, where lower risk on burglaries is observed between March and September with the lowest risk in June and July. The highest risk is observed in December followed by November and January. Furthermore, we visualize the space-time interaction as given by Fig. 7. As revealed by this figure, more high risk neighborhoods are

Table 3 Proportion of variance explained by each component

Component	Proportion of variance (%)
Structured spatial	16.00
Unstructured spatial	8.33
Structured temporal	4.86
Unstructured temporal	70.72
Space time interaction	0.09

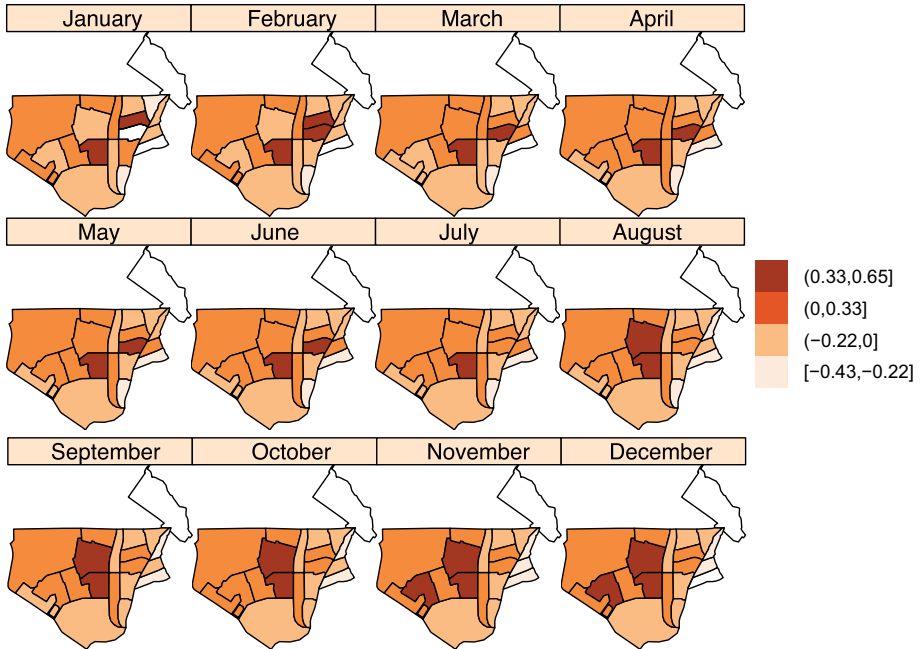


Fig. 7 Posterior mean of the space-time interaction (I) of burglaries (non-spatially or temporally structured). A difference is made between positive and negative ST interaction

observed for November and December, which is in line with the main temporal trend. July has the lowest number of risk neighborhoods. This is also in line with the observed temporal trend. June has a similar temporal trend as July, but has one more neighborhood with high space-time interaction. This space-time interaction is more similar to the one of May. Moreover, November and December show similar space-time interactions. The same holds for September and October. April, May, and June also have a similar space-time interaction pattern. It is also noteworthy to mention that area 1065 has a high space-time interaction for the different months. Area 1064 also has a high space-time interaction between August and December. Considering the areas with high excess risk illustrated in Fig. 6, neighborhood 1057 is the one that is most strongly characterized by a high space-time interaction. This neighborhood exhibits a high space-time interaction between February and June. Surprisingly, this neighborhood has negative space-time interaction in January. Finally, comparing the excess risk of the different neighborhoods with the estimated space-time interactions, we observe that neighborhoods with low

excess risk correspond to neighborhoods with negative space-time interactions. However, neighborhoods with high excess risk do not always correspond to positive space-time interactions.

As discussed earlier, the effect of the repeat and the near-repeat victimization effect is not highly pronounced, but it is still important to evaluate the proportion of variation explained by the spatial component and the temporal component that are directly related to the repeat and the near-repeat effects, respectively. As stated by Blangiardo and Cameletti (2015), the variance of the spatially structured component and the spatially unstructured component are not directly comparable. However, an empirical estimate of the posterior marginal variance can be obtained using a simulation-based approach. For each neighborhood, a large sample is drawn from the corresponding marginal posterior distribution. Then, for each simulation, the empirical variance s_u^2 is calculated as:

$$s_u^2 = \frac{\sum_{i=1}^n (u_i - \bar{u})}{n - 1}. \quad (9)$$

The expected values of the variances of the unstructured component of space, the structured component of time and the unstructured component of time are also calculated, as is the proportion of variance that is explained by each component. The proportion of variance explained by the structured component of space is calculated as given by Eq. (10). As can be seen from Table 3, 16% of the total variability in burglaries can be explained by the spatial structure. Considering only the spatial variance, around 65.73% of the variability could be explained by the spatial structure. In contrast, only around 4.86% of the temporal variance could be explained by the temporal structure. From the total variance only 6.47% could be explained by the temporal structure. The largest part of the variance is captured by the unstructured temporal component.

$$prop_{str.spatial} = \frac{s_u^2}{s_u^2 + \sigma_v^2 + \sigma_\gamma^2 + \sigma_\phi^2 + \sigma_\delta^2}. \quad (10)$$

Performance on an Out of Sample Test

In order to assess the predictive performance of the estimated models, it is important to evaluate the forecast accuracy using genuine forecasts. For this reason, we calculate different accuracy measures using an out of sample test. As we are comparing the performance of the models using the same dataset and sometimes we have low number of realizations, the RMSE, the MAE, and the WAPE are good measures to assess how well the models perform on this new data set. First, we compared the total performance of the models. Results of this analysis show that the unstructured interaction model has the lowest values of the different accuracy measures, and thus the best forecasting performance. The results are consistent with or without covariates (see Table 4). Then we calculate the accuracy measures for each neighborhood and each month separately to check the consistency of the results. As can be seen from Fig. 8, the unstructured interaction model clearly outperforms the other models. The same holds when excluding the covariates. It is noteworthy to mention that the WAPE shows higher peaks in neighborhood code 1059. This can be explained by the fact that the WAPE is sensitive for low realizations and, as can be seen from Fig. 9,

Table 4 Forecast accuracy metrics based on an out-of-sample test

Model	RMSE		MAE		WAPE	
	excl.cov	incl.cov	excl.cov	incl.cov	excl.cov	incl.cov
Spatial	5.585	5.572	4.046	4.072	40.563	40.819
Space + time	4.001	3.964	3.058	3.033	30.658	30.410
Space × time SS-UT	2.264	2.272	1.793	1.798	17.972	18.029
Space × time US-ST	4.350	4.295	3.246	3.215	32.543	32.226
Space × time SS-UT	4.347	4.305	3.216	3.186	32.236	31.941
Space × time SS-ST	4.365	4.313	3.244	3.218	32.517	32.256

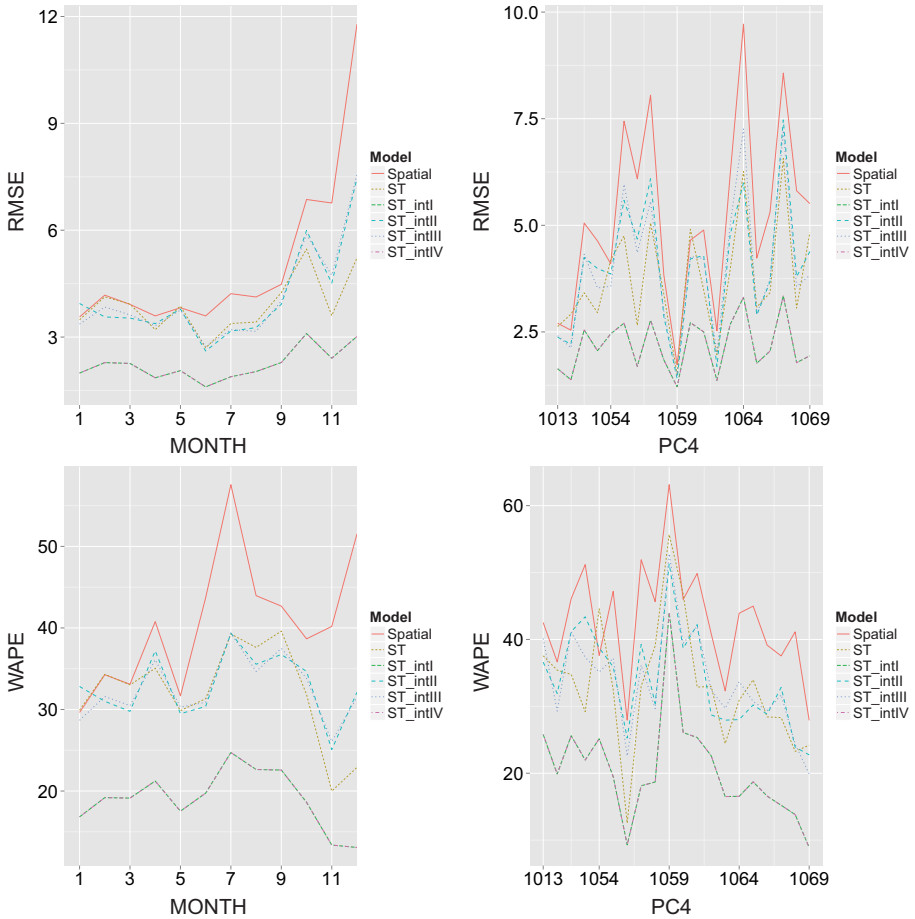


Fig. 8 Performance of the different models per month and per neighborhood using an out-of-sample test

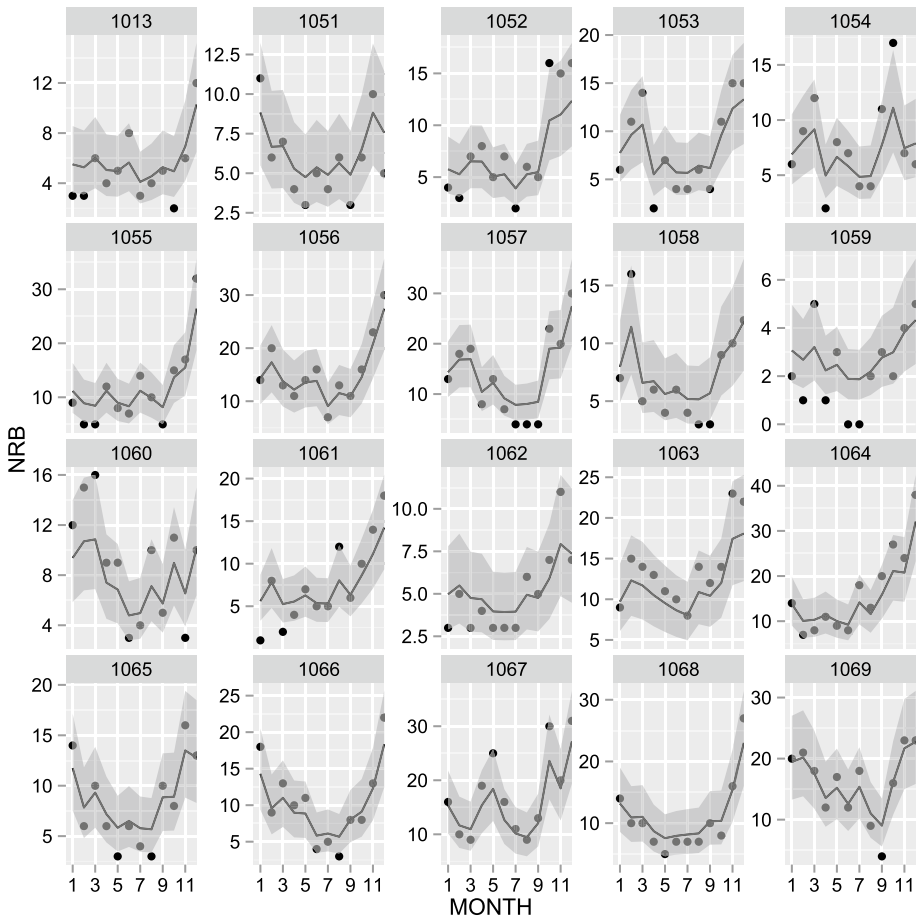


Fig. 9 Forecast versus fit of the optimal model based on an out-of-sample test (including covariates)

this neighborhood has a low number of burglaries. The same reasoning holds for the month of July.

Finally, we assessed graphically the predictions from the unstructured interaction model. Figure 9 shows the realizations including outliers versus the forecast and the credible intervals for the different neighborhoods. As can be seen from this figure, the numbers of burglaries are well-forecasted using this model. Almost all observations are obtained within the 95% credible intervals. Finally, we opt for a backtesting procedure to get more insight in the percentage of realizations that lies within the estimated 95% CI. The total number of realizations that lie within the 95% was calculated and divided by the total number of observations in the test set. This final analysis revealed that 88.33% of the realized burglaries lie within the estimated 95% CI. Deeper analysis reveals that neighborhood 1059 has the highest percentage of exceedances, which was about 14.29%. In the data analysis part, it was shown that this neighborhood has the lowest average number of burglaries among all neighborhoods. Concerning exceedances in the different months, February, August, and

September have the highest percentage of exceedances. The variation in the number of burglaries in these months is quite high.

Conclusions and Future Work

Over the past decades, crime forecasting has gained increasing popularity both in academia and in law enforcement agencies. Many police agencies have started to use predictive policing, the practice of using crime data and algorithms to allocate manpower and other resources to the places and times where and when crime is forecasted to occur.

The purpose of the present study was to contribute to the literature on modeling and forecasting crime by applying the Integrated Laplace Approximation (INLA), a recently developed Bayesian method that is well suited for the spatial, temporal and spatio-temporal modeling and forecasting of crime. Motivated by its potential use in allocating police resources, we evaluated the predictive performance of the models in an empirical study of monthly residential burglary frequencies of twenty neighborhoods in the city of Amsterdam, the Netherlands, from January 2009 to December 2012. Numbers of monthly neighborhood burglaries were modeled as a function of (a) where and when prior residential burglaries and other crimes had taken place, (b) demographic and socio-economic characteristics of the neighborhood, (c) presence of retail business and other facilities, and (d) neighborhood-level street network characteristics. In summarizing the findings we distinguish between substantive and methodological findings.

The substantive findings apply to three elements of the burglary frequency explanation and forecast: *spatio-temporal interaction*, *street network structure*, and *crime generators*.

A major feature of the INLA framework is its ability to model *spatio-temporal interaction*. In modeling monthly neighborhood levels of current crime as a function of prior crime, we compared the performance of six models with spatial, temporal, and spatio-temporal components. The model with only spatial components predicts current crime on the basis of *where* (in which neighborhoods) prior crime occurred, irrespective of when it did. The model with only temporal components predicts current crime as a function of *when* prior crime occurred, irrespective of where it did. The spatio-temporal interaction models include both dimensions simultaneously. Four different types of spatio-temporal interaction can be distinguished based on whether a structure is imposed on the spatial and temporal components: type US-UT (both space and time unstructured), type US-ST (space unstructured, time structured), type SS-UT (space structured, time unstructured) and type SS-ST (both space and time structured). Spatial structure was based on proximity, which assumes that current crime is affected more strongly by crime in nearby neighborhoods than by crime in distant neighborhoods. Temporal structure is based on recency, which assumes that current crime is affected more strongly by recent crime than by earlier crime.

Based on three alternative model fit measures, and both for models with and without covariates, we assessed and compared the explanatory power of the six models. Our findings demonstrate that without exception, the type I spatio-temporal interaction model (unstructured time and unstructured space) outperforms the other five models. This conclusion is confirmed by the results of an out-of-sample test, in which the type I model also outperformed the five others on three forecast accuracy measures. The explanatory and predictive superiority of the type I spatio-temporal interaction model speaks against the expectation that a type SS-ST spatio-temporal model (with structured spatial and temporal components) would perform optimally. This expectation was motivated by the extended

(spatial and temporal) interpretation of the First Law of Geography (Tobler 1970), because the type IV model recognizes the spatial and temporal proximity between observations, and uses both proximities in the forecast. Also, and more importantly, the extant literature clearly identifies a compatible underlying mechanism for spatio-temporal interaction, namely the finding that in the wake of a burglary, offenders tend to return to the environment of the prior burglary to re-offend against the same or nearby targets. A potential reason for this lack of confirmation is the spatial and temporal scale of our analysis. For reasons of practical usefulness, we model monthly neighborhood burglary frequencies. However, most empirical findings on repeat and on near repeat burglary find that the level spatio-temporal interaction is particularly strong within the first 2 weeks after the initial burglary and within 200 m from the initial target, and that it decays with time and distance (Johnson et al. 2007). In other words, the majority of returning burglars return to burgle a property in the same neighborhood and in the same month, and therefore the spatially and temporally structured components in the other three spatio-temporal interaction models (types US-ST, SS-UT, and SS-ST) can only detect a very limited subset of these patterns and use it in the forecast. As a result, for medium-range forecasting (monthly neighborhood frequencies), using both time and space remains important but explicitly imposing a structure (spatial proximity and temporal recency) does not add significantly to the accuracy of the forecast. A practical consequence of this finding is that forecasting tools can use relatively simple models. An implied consequence is that forecasting data requirements are less strict. For example, it is not generally necessary to have last month's crime data as input to the forecast of next month's crime data.

The second substantive conclusion concerns the role of street network structure in the explanation and forecast of residential burglary. Street networks structure the daily mobility patterns of the urban population by facilitating and constraining movements. To move from one place to another in the most efficient way, individuals must traverse the street network. Streets at central positions in the network are on the path of more travelers, and their characteristics are therefore better known to larger proportions of the population, including potential burglars. Streets more central in the network are therefore expected to have increased burglary frequencies, and the neighborhoods in which they are situated will likely have higher burglary frequencies than neighborhoods with streets with lower network centrality. We used two measures of network centrality in the explanation and prediction of burglary, *betweenness* and *closeness*. Both measure the extent to which a given street is on the path between many other pairs of streets. Our findings suggest that *closeness* more than *betweenness* is predictive of a neighborhood residential burglary frequencies. The finding that centrality in the street network structure is associated with an increased risk of residential burglary confirms prior work using street network measures (Davies and Johnson 2015; Frith et al. 2017), including also the finding that streets with low centrality (such as cul-de-sacs) have reduced burglary rates (Johnson and Bowers 2010; Wu et al. 2015).

The third set of substantive findings applies to the presence in the neighborhood of facilities and other crimes than burglary. We found that numbers of retail stores affected burglary frequencies positively, probably because retail stores attract people (including burglars) who become aware of potential targets in the neighborhood. It was also found that numbers of recent street robberies was positively associated with burglaries, which might indicate a temporary increase in the number of prospective offenders active in the neighborhood, as many offenders tend to be versatile rather than specialized in terms of the types of crime they commit (Guerette et al. 2005). The number of single-person households was negatively related to burglary.

With respect to future work, taking into account that the main criminological promise of the INLA approach is its capabilities as a modeling and forecasting method for spatio-temporal data, the most likely next step in its application is to assess its value on crime data more fine-grained spatial and temporal resolutions. More specifically, daily or weekly crime data at the level of street segments would allow future users of the method to verify its usefulness for short-term and small-area crime forecasting. This may not only contribute to a further theoretical understanding of the origins of spatio-temporal crime patterns, but it may also become a practical forecasting tool for law enforcement agencies that aim to allocate manpower and other resources on a day-to-day and street-by-street basis (as opposed to a medium range month-by-month and neighborhood-by-neighborhood basis).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

In the following, the main ideas of each key component of INLA will be explained.

Latent Gaussian models cover a large class of statistical models with a unified framework for doing statistical inference (Rue et al. 2009). An LGM within the Bayesian framework is generally defined as a generic three-stage hierarchical model: likelihood, latent Gaussian field, and the hyperparameters. In the first stage, the likelihood is defined assuming conditional independence. This means that each variable in \mathbf{y} is conditionally independent given a latent Gaussian field $\boldsymbol{\theta}$ and hyperparameters $\boldsymbol{\psi}_1$. Additionally, it is assumed that the distribution of the univariate likelihood is from the exponential family. Formally, the three stages can be written as:

- Stage 1: The likelihood

$$\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}_1 \sim \prod_{i=1}^n p(y_i|\eta_i(\boldsymbol{\theta}), \boldsymbol{\psi}_1). \quad (11)$$

- Stage 2: The latent Gaussian field:

$$p(\boldsymbol{\theta}|\boldsymbol{\psi}_2) \sim \mathcal{N}\left(\mu(\boldsymbol{\psi}_2), Q^{-1}(\boldsymbol{\psi}_2)\right). \quad (12)$$

- Stage 3: The hyperparameters

$$\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \sim \boldsymbol{\pi}(\boldsymbol{\psi}), \quad (13)$$

where $Q_{\boldsymbol{\psi}_2}$ depends on the hyperparameters and describes the dependence structure of the data. The resulting joint posterior reads:

$$p(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y}) \propto \boldsymbol{\pi}(\boldsymbol{\psi}) \boldsymbol{\pi}(\boldsymbol{\theta} | \boldsymbol{\psi}_2) \prod_{i=1}^n p(y_i | \boldsymbol{\eta}_i, \boldsymbol{\psi}_1). \quad (14)$$

In order that INLA works efficiently, the latent field should not only be a GF but a GMRF. A GMRF is a Gaussian field with additional conditional independence properties, giving the precision matrix a sparse structure. We say that $\boldsymbol{\theta}$ is a GMRF if it has a multivariate normal density with additional conditional independence, also called the Markov property. This means that the conditional distribution of $\boldsymbol{\theta}$ at some location given all other $\boldsymbol{\theta}$'s depends only on $\boldsymbol{\theta}$ at neighboring locations resulting in a sparse precision matrix. This sparsity of the precision matrix is very tractable from a computational point of view, especially when the model dimension is high. In general, it is hard to quantify the computational benefits from using a GMRF as it depends on the sparsity pattern of the precision matrix. For models with a spatial structure the computational benefit decreases from $\mathcal{O}(n^3)$ for a dense matrix to $\mathcal{O}(n^{3/2})$ when a sparse matrix is used. This computational benefit is also paired with a reduction in the memory requirement that reduces for the same model from (n^2) to $(n \log(n))$ which makes it easier to run larger models (Rue et al. 2009). A detailed explanation of GMRF with different examples can be found in Rue and Held (2005).

The power of INLA is provided by using GMRF in constructing the additive models. Interpreting the likelihood $\prod_{i=1}^n p(y_i | \boldsymbol{\theta}_i; \boldsymbol{\psi}_1)$, so that y_i only depends on its linear predictor $\boldsymbol{\theta}_i$ results in the generalized linear model set up. Interpreting $\boldsymbol{\theta}_i$ as a linear predictor $\boldsymbol{\eta}_i$ which is additive with respect to other effects as given by the following equation:

$$\boldsymbol{\eta}_i = \beta_0 + \sum_{m=1}^M \beta_m x_{mi} + \sum_{l=1}^L f_l(z_{li}), \quad (15)$$

where β_0 is the intercept and the parameter vector $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_M\}$ represents the linear effect of some covariates x_1, \dots, x_M on the response variable. The model components $\mathbf{f} = \{f_1(\cdot), \dots, f_L(\cdot)\}$ are defined on a set of covariates $z = \{z_1, \dots, z_L\}$ and can take different forms. These could be a random intercept, a random slope, non-linear effects of covariates, spatial random effects, temporal random effects, time trends, etc.

In the INLA setting, a difference is made between fixed parameters and hyperparameters (variance parameters). Hyperparameters could be residual noise, the over-dispersion parameter of the negative binomial distribution, the shape parameters of the gamma distribution, the variance parameter of a random effects in a mixed effect model, the autoregressive parameter from the autoregression correlation in time series, or the parameters from the spatial correlation, etc. Let us denote the set of fixed parameters by $\boldsymbol{\theta}$ such that $\boldsymbol{\theta} = \{\beta_0, \beta_1, \dots, \beta_M, f_1(\cdot), \dots, f_L(\cdot)\}$ and the set of hyperparameters by $\boldsymbol{\psi} = \{\psi_1, \dots, \psi_K\}$ for K hyperparameters.

The formulation given by Eq. (15) differs from the Generalized Linear Model (GLM) formulation in the model components which are used to represent specific Gaussian processes (Zuur et al. 2017). Assuming Gaussian priors for each fixed effect and each model component and independence from each other, and assuming that the model components are a-priori independent yields that the joint distribution of $\boldsymbol{\theta} = (\boldsymbol{\eta}, \beta_0, \beta_1, \dots, \beta_M, f_1(\cdot), \dots, f_L(\cdot))$ is also a GMRF and its precision matrix consists of the sum of the precision matrices of the covariates and the model components. As the joint distribution depends on the hyperparameters $\boldsymbol{\psi}$, this means that it should be formed many times in INLA. One of the key reasons that the INLA algorithm is so efficient is due to the

fact that the joint distribution can be treated as a GMRF with a precision matrix that is easy to compute.

The third component that we need to explain is the Laplace Approximation. Before the MCMC times, the LA was regarded as a key tool for doing high-dimensional integration. Using the LA, we are typically interested in evaluating the integral $I_n = \int \exp(nf(x))dx$ as $n \rightarrow \infty$.

Representing $f(x)$ by mean of Taylor series expansion evaluated in its mode x^* , with some calculus it can be shown that I_n can be approximated by $\exp\left(nf(x^*)\right)\sqrt{\frac{2\pi}{-nf''(x^*)}}$ which represents the kernel of the Normal distribution with mean x^* and variance $\sigma^{2*} = \frac{-1}{nf''(x^*)}$ (Blangiardo and Cameletti 2015).

In practice, n usually represents the number of i.i.d. replications, each of which has density $f(x)$. Interpreting the $nf(x)$ as the sum of log-likelihoods and x the unknown parameter, if the central limit theorem holds, the Gaussian approximation will be exact as $n \rightarrow \infty$. This means that the LA is more accurate for posteriors that are near-Gaussian compared to densities that are far from Gaussian. As mentioned earlier, the main interest in Bayesian inference lies in the marginal posterior distributions of each element of the parameter vector $p(\theta_i|\mathbf{y})$ and of the hyperparameters $p(\boldsymbol{\psi}_k|\mathbf{y})$. As the posterior density is usually complex (product of Gaussian and non-Gaussian densities), the LA can not be directly used. INLA reformulates the problem as a series of sub-problems and applies the LA only to densities that are almost-Gaussian. The method can be divided into three main tasks. First, an approximation of the joint posterior $p(\boldsymbol{\psi}|\mathbf{y})$ will be proposed. Second, an approximation to the marginals of the conditional distribution of θ_i ; $p(\theta_i|\boldsymbol{\psi}, \mathbf{y})$; given the data and the hyperparameters is proposed (Wang et al. 2018). And finally, exploring the approximated joint posterior distribution of $p(\boldsymbol{\psi}|\mathbf{y})$ and use it for numerical integration. The approximated posterior marginals in INLA can be given by:

$$\tilde{p}(\theta_i|\mathbf{y}) = \sum_k \tilde{p}(\theta_i|\boldsymbol{\psi}^{(k)}, \mathbf{y})\tilde{p}(\boldsymbol{\psi}^{(k)}|\mathbf{y})\Delta\theta^{(k)}, \tag{16}$$

$$\tilde{p}(\boldsymbol{\psi}_i|\mathbf{y}) = \int \tilde{p}(\boldsymbol{\psi}_i|\mathbf{y})d\boldsymbol{\psi}_{-j}, \tag{17}$$

where $\tilde{p}(\boldsymbol{\psi}^{(k)}|\mathbf{y})$ are the density values computed during on $\tilde{p}(\boldsymbol{\psi}|\mathbf{y})$ and $\boldsymbol{\psi}_{-j}$ stands for $\boldsymbol{\psi}$ excluding the j th element. The technical details for each task are described in Rue et al. (2017) and Wang et al. (2018).

References

Andresen MA, Malleon N (2013) Crime seasonality and its variations across space. *Appl Geogr* 43:25–35
 Bavelas A (1950) Communication patterns in task-oriented groups. *J Acoust Soc Am* 22(6):725–730
 Bernasco W (2008) Them again? Same-offender involvement in repeat and near repeat burglaries. *Eur J Criminol* 5(4):411–431
 Bernasco W (2010) Modeling micro-level crime location choice: application of the discrete choice framework to crime at places. *J Quant Criminol* 26(1):113–138
 Bernasco W, Johnson SD, Ruiter S (2015) Learning where to offend: effects of past on future burglary locations. *Appl Geogr* 60:120–129
 Bernasco W, Ruiter S, Block R (2017) Do street robbery location choices vary over time of day or day of week? A test in Chicago. *J Res Crime Delinq* 54(2):244–275

- Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43(1):1–20
- Blangiardo M, Cameletti M (2015) *Spatial and spatio-temporal Bayesian models with R-INLA*. Wiley, New York
- Blokdijk D, Beijersbergen K (2020) Achtergronden en recidive onder daders van high impact crimes veroordeeld in 2002–2016. Technical report, Ministry of Justice and Security
- Bowers KJ, Johnson SD (2004) Who commits near repeats? A test of the boost explanation. *West Criminol Rev* 5(3):12–24
- Breetzke GD, Cohn EG (2012) Seasonal assault and neighborhood deprivation in South Africa: some preliminary findings. *Environ Behav* 44(5):641–667
- Breunig MM, Kriegel H-P, Ng RT, Sander J (2000) Lof: identifying density-based local outliers. In: Dunham M, Naughton JF, Chen W, Koudas N (eds) *Proceedings of the 2000 ACM SIGMOD conference of management of data*, vol 29. ACM, New York, pp 93–104
- Brunsdon C, Corcoran J, Higgs G, Ware A (2009) The influence of weather on local geographical patterns of police calls for service. *Environ Plan B Plan Des* 36(5):906–926
- Camacho-Collados M, Liberatore F (2015) A decision support system for predictive police patrolling. *Decis Support Syst* 75:25–37
- Carlin BP, Louis TA (2008) *Bayesian methods for data analysis*. CRC Press, New York
- Carroll R, Lawson A, Faes C, Kirby R, Aregay M, Watjou K (2015) Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spat Spatio-Temporal Epidemiol* 14:45–54
- Ceccato V (2005) Homicide in Sao Paulo, Brazil: assessing spatial-temporal and weather variations. *J Environ Psychol* 25(3):307–321
- Coupe T, Blake L (2006) Daylight and darkness targeting strategies and the risks of being seen at residential burglaries. *Criminology* 44(2):431–464
- Davies T, Johnson SD (2015) Examining the relationship between road structure and burglary risk via quantitative network analysis. *J Quant Criminol* 31(3):481–507
- Eck J, Chainey S, Cameron J, Wilson R (2005) *Mapping crime: understanding hotspots*. National Institute of Justice, Washington DC
- Erkin D (2017) *Delivery routing challenge for a European retailer master thesis*. Master's thesis
- Everson S, Pease K (2001) Crime against the same person and place: detection opportunity and offender targeting. *Crime Prev Stud* 12:199–220
- Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41
- Frith MJ, Johnson SD, Fry HM (2017) Role of the street network in burglars' spatial decision-making. *Criminology* 55(2):344–376
- Groff ER, La Vigne NG (2002) Forecasting the future of predictive crime mapping. *Crime Prev Stud* 13:29–58
- Grubestic TH, Mack EA (2008) Spatio-temporal interaction of urban crime. *J Quant Criminol* 24(3):285–306
- Guerette RT, Stenius VM, McGloin JM (2005) Understanding offense specialization and versatility: a reap-lication of the rational choice perspective. *J Crim Justice* 33(1):77–87
- Guerry A-M (1833) *Essai sur la statistique morale de la France*. Crochard, Paris
- Haberman CP, Ratcliffe JH (2015) Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology* 53(3):457–483
- Hägerstrand T (1970) What about people in regional science? *Pap Reg Sci Assoc* 24(1):7–21
- Hipp JR, Kim Y-A (2019) Explaining the temporal and spatial dimensions of robbery: differences across measures of the physical and social environment. *J Crim Just* 60:1–12
- Hodges JS (2016) *Richly parameterized linear models: additive, time series, and spatial models using random effects*. CRC Press, New York
- Hu T, Zhu X, Duan L, Guo W (2018) Urban crime prediction based on spatio-temporal Bayesian model. *PLoS One* 13(10):1–18
- Hvistendahl M (2016) Crime forecasters. *Science* 353(6307):1484–1487
- Johnson SD, Bowers KJ (2004) The burglary as clue to the future: the beginnings of prospective hot-spotting. *Eur J Criminol* 1(2):237–255
- Johnson SD, Bowers KJ (2010) Permeability and burglary risk: Are cul-de-sacs safer? *J Quant Criminol* 26(1):89–111
- Johnson SD, Bernasco W, Bowers KJ, Elffers H, Ratcliffe J, Rengert G, Townsley M (2007) Space-time patterns of risk: a cross national assessment of residential burglary victimization. *J Quant Criminol* 23(3):201–219
- Johnson SD, Summers L, Pease K (2009) Offender as forager? A direct test of the boost account of victimization. *J Quant Criminol* 25(2):181–200

- Lammers M, Menting B, Ruiter S, Bernasco W (2015) Biting once, twice: the influence of prior on subsequent crime location choice. *Criminology* 53(3):309–329
- Lantz B, Ruback RB (1997) A networked boost: burglary co-offending and repeat victimization using a network approach. *Crime Delinq* 63(9):1066–1090
- Law J, Quick M, Chan P (2014) Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level. *J Quant Criminol* 30(1):57–78
- Law J, Quick M, Chan PW (2015) Analyzing hotspots of crime using a Bayesian spatiotemporal modeling approach: a case study of violent crime in the greater Toronto area. *Geogr Anal* 47(1):1–19
- Luan H, Quick M, Law J (2016) Analyzing local spatio-temporal patterns of police calls-for-service using Bayesian integrated nested Laplace approximation. *ISPRS Int J Geo-Inf* 5(9):162
- Mahfoud M, Bhulai S, van der Mei RD (2017) Spatio-temporal modeling for residential burglary. In: Bhulai S, Kardaras D (eds) *Proceedings of the 6th international conference on data analytics. IARIA, Barcelona*, pp 59–64
- Mahfoud M, Bhulai S, van der Mei R, Erkin D, Dugundji E (2018) Forecasting burglary risk in small areas via network analysis of city streets. In: *7th International conference on data analytics. IARIA*, p 109
- Mahfoud M, Bhulai S, van der Mei R, Erkin D, Dugundji E (2019) Network analysis of city streets: forecasting burglary risk in small areas. To appear in *International Journal On Advances in Security*
- Miller HJ (2005) A measurement theory for time geography. *Geogr Anal* 37(1):17–45
- Mohler GO, Short MB, Brantingham PJ, Schoenberg FP, Tita GE (2011) Self-exciting point process modeling of crime. *J Am Stat Assoc* 106(493):100–108
- Opitz T (2017) Latent Gaussian modeling and INLA: a review with focus on space-time applications. arXiv preprint [arXiv:1708.02723](https://arxiv.org/abs/1708.02723)
- Pease K et al (1998) *Repeat victimisation: taking stock*, vol 90. Home Office Police Research Group, London
- Perry WL, McInnis B, Price CC, Smith SC, Hollywood JS (2013) *Predictive policing: the role of crime forecasting in law enforcement operations*. Rand Corporation, Santa Monica, CA
- Pettit L (1990) The conditional predictive ordinate for the normal distribution. *J R Stat Soc Ser B (Methodol)* 5:175–184
- Ratcliffe JH (2006) A temporal constraint theory to explain opportunity-based spatial offending patterns. *J Res Crime Delinq* 43(3):261–291
- Ratcliffe JH (2008) *Intelligence-led policing*. Willan Publishing, Cullompton, UK
- Ratcliffe JH (2010) Crime mapping: spatial and temporal challenges. In: *Handbook of quantitative criminology*. Springer, Berlin, pp 5–24
- Roos M, Held L et al (2011) Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Anal* 6(2):259–278
- Rue H, Held L (2005) *Gaussian Markov random fields: theory and applications*. CRC Press, Boca Raton
- Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B (Stat Methodol)* 71(2):319–392
- Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK (2017) Bayesian computing with INLA: a review. *Annu Rev Stat Appl* 4:395–421
- Shaw CR, McKay HD (1942) *Juvenile delinquency and urban areas*. University of Chicago Press, Chicago
- Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. *Nat Phys* 6(10):818
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser B (Stat Methodol)* 64(4):583–639
- Tobler WR (1970) A computer movie simulating urban growth in the detroit region. *Econ Geogr* 46(sup1):234–240
- Todovic D, Makajic-Nikolic D, Kostic-Stankovic M, Martic M (2015) Police officer scheduling using goal programming. *Polic Int J* 38(2):295–313
- Townsley M, Homel R, Chaseling J (2003) Infectious burglaries. A test of the near repeat hypothesis. *Br J Criminol* 43(3):615–633
- Wang X, Ryan YY, Faraway JJ (2018) *Bayesian regression modeling with INLA*. CRC Press, Boca Raton
- Weisburd D (2015) The law of crime concentration and the criminology of place. *Criminology* 53(2):133–157
- Weisburd D, Lum C (2005) The diffusion of computerized crime mapping in policing: linking research and practice. *Police Pract Res* 6(5):419–434
- Weisburd D, Bushway S, Lum C, Yang S-M (2004) Trajectories of crime at places: a longitudinal study of street segments in the city of seattle. *Criminology* 42(2):283–322
- Wu L, Liu X, Ye X, Leipnik M, Lee J, Zhu X (2015) Permeability, space syntax, and the patterning of residential burglaries in urban China. *Appl Geogr* 60:261–265

Zuur AF, Ieno EN, Smith GM (2007) *Analysing ecological data*, vol 680. Springer, Berlin

Zuur AF, Ieno EN, Saveliev AA (2017) *Beginner's guide to spatial, temporal, and spatial-temporal ecological data analysis with R-INLA*. Highland Statistics Ltd., Newburgh, UK

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Maria Mahfoud¹ · Wim Bernasco^{2,3}  · Sandjai Bhulai⁴ · Rob van der Mei^{1,4}

Maria Mahfoud
M.Mahfoud@cwi.nl

Sandjai Bhulai
S.Bhulai@vu.nl

Rob van der Mei
R.D.van.der.Mei@cwi.nl

¹ CWI National Research Institute for Mathematics and Computer Science, Amsterdam, The Netherlands

² Department of Spatial Economics, School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

³ Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), P.O. Box 71304, 1008 BH Amsterdam, The Netherlands

⁴ Department of Mathematics, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands