CrossMark

Improving the Robustness and Accuracy of Crime Prediction with the Self-Exciting Point Process Through Isotropic Triggering

Gabriel Rosser [1] · **Tao Cheng** [1]

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract The self-exciting point process (SEPP) is a model of the spread of crime in space and time, incorporating background and triggering processes. It shows promising predictive performance and forms the basis of a popular commercial software package, however few detailed case studies describing the application of the SEPP to crime data exist in the scientific literature. Using open crime data from the City of Chicago, USA, we apply the SEPP to crime prediction of assaults and burglaries in nine distinct geographical regions of the city. The results indicate that the algorithm is not robust to certain features of the data, generating unrealistic triggering functions in various cases. A simulation study is used to demonstrate that this outcome is associated with a reduction in predictive accuracy. Analysing the second-order spatial properties of the data demonstrates that the failures in the algorithm are correlated with anisotropy. A modified version of the SEPP model is developed in which triggering is non-directional. We show that this provides improved robustness, both in terms of the triggering structure and the predictive accuracy.

Introduction

The ability to predict the future location of crime hotspots confers myriad advantages on a police force, including planning effective proactive interventions, reducing the

---

✉ Gabriel Rosser
   gabriel.rosser@gmail.com

1   SpaceTimeLab, Department of Civil, Environmental & Geomatic Engineering, University College London, Chadwick Building, Gower Street, London WC1E 6BT, England

🖄 Springer

level of crime, promoting public safety and improving the efficiency with which resources are allocated. Police forces worldwide have used mapping tools for many decades to visualise and interpret past crime patterns (Groff and La Vigne 2002), however this process is fundamentally retrospective in nature. Predictive approaches, by contrast, involve methods that forecast the future distribution of crime risk based on well-stated models and assumptions. Such methods are invariably computational in nature and forecasts are based upon historic crime data, census data or land use data.

A notable barrier to the widespread operationalisation of a new crime prediction method is its robustness to the data sources to which it is applied. A successful prediction method should function on a wide variety of datasets, irrespective of the way those data have been recorded or the specific spatio-temporal patterns within the data. The self-exciting point process (SEPP) model is a recently-developed crime prediction method whose inputs are the locations and times of historic crimes. It achieves strong predictive performance when applied to a residential burglary dataset from Los Angeles, USA (Mohler et al. 2011). The SEPP model also forms the basis of the commercial crime prediction package PredPol (TM). Despite the fact that this software is currently used in various police forces in the USA, to our knowledge there are only two detailed case studies in the scientific literature describing the application of the SEPP to crime data (Mohler et al. 2011; Mohler 2014). As we discuss below, the implementation details vary between these two studies; we only consider the methodology proposed in the earlier work. Due to the proprietary nature of PredPol, the precise details of the SEPP implementation in this software package are not known.

We begin our study by attempting to apply the SEPP to burglary and assault crime data from the City of Chicago, USA. We divide the data into nine geographical regions, based on an established urban classification. This exercise demonstrates that the SEPP is not robust to certain features of the data, leading to unrealistic results. The characterisation and resolution of this issue form the primary motivations for this study.

We approach the problem by first identifying a local second-order characteristic of the data that is indicative of poor performance of the SEPP. Having gained insight into the underlying cause of the failure, we use a simulation study to demonstrate that the predictive accuracy of the SEPP deteriorates when we reproduce similar conditions in a simulated dataset. We develop a modified SEPP model in which triggering is non-directional to improve performance in such situations. Henceforth we will refer to this as the isotropic SEPP model. We evaluate the existing and isotropic methods to show how the model structure and predictive performance compare. Our results show that the isotropic model matches or improves upon the performance of the original method in most cases.

Our study makes two key contributions to the field of crime prediction. First, our analysis of second-order effects allows us to probe the structure of a crime dataset in order to assess a priori how the SEPP model will perform. Several other crime prediction methods are also based on the concept of the boost hypothesis (for example the ProMap method of Bowers et al. (2004)), therefore this approach is widely applicable beyond the realm of the SEPP. Second, we develop an effective alternative to the regular SEPP that is more robust to the vagaries of real police recorded crime data, which facilitates operational use of the method.

Data and Methods

Crime Data from the City of Chicago

For the purposes of this study, we use open access crime data records from the City of Chicago covering a 1 year period starting on 1st March 2011. We consider two different crime types, assault (21,480) and burglary (26,428 crimes). For the purposes of modelling and analysis, we further divide the city into 9 'sides', as shown in Fig. 1a. For brevity, we abbreviate these to FN (Far North side), and so on. Chicago's sides are defined as a collection of multiple community areas, which are used for urban planning purposes.

We choose sides as our areal units in order to subject the SEPP to a rigorous evaluation process by applying it to varied datasets. While designing this study, we hypothesised that the differences in geography and geodemographics between the sides would lead to variation in the boost effect between them, which would be reflected in the magnitude and spatial extent of the SEPP's triggering component.

Figure 1b, c illustrate two such differences in the spatial arrangement of burglary crimes; the former shows densely clustered crimes arranged along a residential road network in S and the latter illustrates a very sparse crime pattern in FSW. In particular, the arrangement in S appears to suggest that crimes are preferentially aligned with streets that run north-south (n/s). As we demonstrate below, the variation between regions leads to some severe and problematic effects on the SEPP model, which motivates the remainder of the study.

Figure 1a also highlights several geographical features that might feasibly affect the SEPP by introducing 'holes' in the spatial point pattern of crimes. For example, no burglaries or assaults occur in a large southern portion of the FSE that contains a water
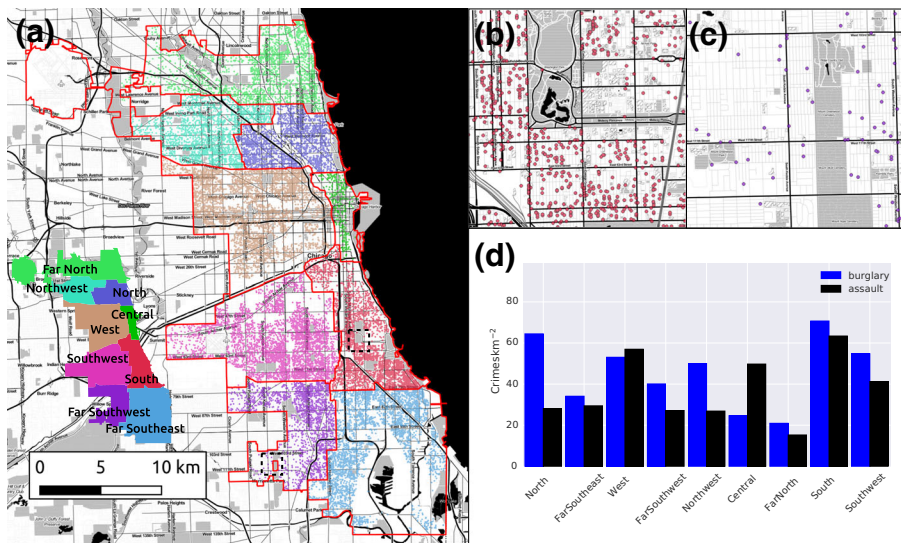

Fig. 1 a The nine sides of the City of Chicago with burglary crimes overlaid. The basemap is © Stamen (stamen.com). b and c Magnified plots of the two regions enclosed by black dashed boxes. d Crime density by crime type and side

reclamation plant, railway sidings, marshland and golf courses. Similarly, large gaps appear throughout FN, SW and W primarily due to the presence of airports and parks.

Figure 1d shows significant variation in the density of the two crime types between the different sides of Chicago. In N and NW, which have large residential areas, burglaries are twice as frequent as assaults; the opposite is true in C, which is the main commercial centre of the city.

The Self-Exciting Point Process

A wide variety of approaches to crime prediction exists in the scientific literature, commercial software packages and bespoke software applications used by police agencies around the world. From an operational stance, all predictive approaches have in common the aim of forecasting which spatial regions require particular police attention, whether this is to pre-empt a predicted increase in criminal activity or because that location is already experiencing a substantial level of crime. The SEPP method uses only historic crime records to generate predictions. Other methods also incorporate geodemographic variables (Johnson et al. 2009), however these are typically static or infrequently updated and we do not consider them further.

The SEPP is an approach based on a theoretical model that was originally developed in the context of interpreting seismic activity (Musmeci and Vere-Jones 1992), but was recently applied to crime prediction by Mohler et al. (Mohler et al. 2011). Drawing on the criminological theories of risk heterogeneity (Sparks 1981) and the boost hypothesis (Pease 1998), the SEPP models crimes as arising from a spatially and temporally heterogeneous risk background and a local triggering effect. The combination of two important underlying effects results in a predictive method that outperforms other leading methods in tests of predictive accuracy.

Despite its promising performance in the original study (Mohler et al. 2011), the SEPP has received relatively little attention from police analysts and academics, and has only been applied to a single crime dataset to date. A related study describes an extension to the SEPP that models the interplay between different crime types (Mohler 2014), however for the purposes of the present study we are concerned with the original implementation.

At the core of the SEPP model of crime is the conditional intensity, $\lambda(t, x, y)$, which gives the density of the expected rate of occurrence of crimes in a small neighbourhood around the region (x, y) at time t, conditional upon the history of all occurrences up to that time. Each crime in the dataset is assumed to take the form of a point (t_i, x_i, y_i), where i is the crime index. The conditional intensity may be described as the sum of background and triggered events:

$$\lambda(t_i, x_i, y_i) = \mu(t_i, x_i, y_i) + \sum_{j:t_j < t_i} g\left(t_i{-}t_j, x_i{-}x_j, y_i{-}y_j\right) \qquad (1)$$

where μ denotes the background occurrence rate and g denotes the triggering function. The summation is over all crimes that occurred prior to time t. Therefore all preceding crimes may theoretically contribute some additional expectation of the current crime activity, though in practice this contribution vanishes over some period of time and/or distance.

In order to apply this theory to real data we must estimate the functional forms of μ and g. This entails declustering the data (Zhuang et al. 2002) to identify those events arising from background activity and those triggered by previous events. Two general approaches to this problem have been demonstrated in the literature. The first assumes a parametric form for both functions and optimises the parameters involved using maximum likelihood methods. A recent example is the study by Mohler (2014), which is conceptually similar to the approach taken here but assumes a Gaussian triggering function.

A second distinct approach involves the use of non-parametric functions to estimate μ and g. Mohler et al. (2011) use an expectation-maximisation (EM) algorithm and kernel density estimates (KDEs) to achieve this task. This approach is theoretically more flexible, since it does not require any prior assumptions on the functional form of the background and triggering functions. However, in practice, it introduces additional complications at the optimisation stage. Whilst a detailed comparison of these two approaches would be interesting, it is beyond the scope of the present work. For the remainder of this study, we therefore focus on the non-parametric implementation of the SEPP.

The complete optimisation algorithm is illustrated in Fig. 2, following (Mohler et al. 2011). We now discuss the model and optimisation steps in detail. Let p_{ij} denote the probability that event j was triggered by event i. By convention, p_{ii} denotes the probability that event i arose from the background. Furthermore, $p_{ij} = 0$ if $t_i \geq t_j$, so that all of the probabilities may be encoded in an upper triangular matrix, P. These probabilities are given by

$$p_{ii} = \frac{\mu(t_i, x_i, y_i)}{\lambda(t_i, x_i, y_i)} \tag{2}$$

$$p_{ij} = \frac{g\left(\Delta t_{ij}, \Delta x_{ij}, \Delta y_{ij}\right)}{\lambda(t_i, x_i, y_i)} \tag{3}$$

where $\Delta t_{ij} = t_j - t_i$.

Computing allowed links is necessary to reduce computation time. Two threshold parameters, $\Delta t_{\max}$ and $\Delta d_{\max}$, define the maximum permissible time period and distance, respectively, over which triggering may occur. As Fig. 2 shows, this is equivalent to defining a space-time cylinder around every crime. Imposing this threshold enforces zero values in P, which in turn reduces the number of evaluations of the KDE for the triggering function g in the update stage. The two parameters should be set sufficiently large that there is no persistent boost effect at longer temporal or spatial distances. Throughout this study, we take $\Delta t_{\max} = 120$ days and $\Delta d_{\max} = 500$ metres, which are both greater than empirically-obtained estimates in dense urban locations (Johnson et al. 2007).

Initialisation of P is performed by assuming a simple triggering form that is exponentially decaying in time and bivariate normal in space:

$$p_{ij} = \exp\left(-\alpha\left(t_j - t_i\right)\right)\exp\left(\frac{-\left(x_j - x_i\right)^2 - \left(y_j - y_i\right)^2}{2\beta^2}\right) \tag{4}$$
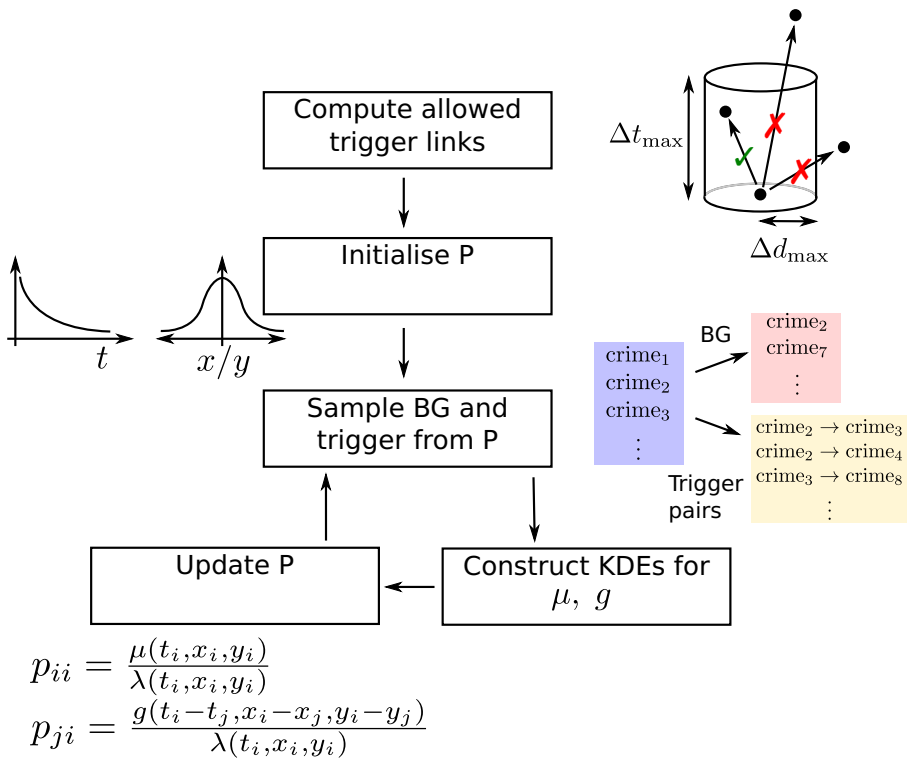
Fig. 2 Flow diagram illustrating the SEPP optimisation algorithm. Full details are given in the main text

where the parameters α and β denote the initial estimate of the time decay constant and spatial bandwidth, respectively. The background probabilities p_{ii} are set equal to one. Finally, the columns of the matrix P are normalised so that they all sum to 1. In all the real datasets tested in this study, we found that choosing $\alpha = 0.1$ day^{-1} and $\beta = 50$ m produced consistent results.

The next step, random sampling, divides the dataset into two groups, background and parent-offspring triggering pairs. This is achieved using an efficient algorithm proposed by Efraimidis and Spirakis (2006). In effect, this step selects one outcome for each datum: either it is placed in the background group, or a parent datum is selected and the pair are placed in the triggering group. The selection is weighted by the non-zero entries in the relevant column of P.

Two KDEs are next constructed, based on the randomly sampled data. The background KDE is separable in time and space, $\mu(t, x, y) = \nu(t)\eta(x, y)$, and is constructed directly from the data points in the background sample. The triggering function is inseparable and is constructed using the difference data, $(\Delta t_{ij}, \Delta x_{ij}, \Delta y_{ij})$, for all parent-trigger pairs (i, j) in the triggering group. We follow the approach of Mohler et al. (2011), who use a variable bandwidth KDE with a Gaussian kernel function with diagonal covariance. This has the form

$$g(\Delta t, \Delta x, \Delta y) = \sum k_T\left(\Delta t; \Delta t_{ij}, \sigma_{\Delta t_{ij}}\right) k_X\left(\Delta x; \Delta x_{ij}, \sigma_{\Delta x_{ij}}\right) k_Y\left(\Delta y; \Delta y_{ij}, \sigma_{\Delta y_{ij}}\right) \quad (5)$$

where the summation is over all difference data and $k_T\left(\cdot; \Delta t_{ij}, \sigma_{\Delta t_{ij}}\right)$ denotes a Gaussian function with mean Δt_{ij} and standard deviation $\sigma_{\Delta t_{ij}}$. The variable bandwidths σ in equation (1) are computed using nearest neighbour (NN) distances from each datum. 100 NNs are used in one dimension (for the background time component), while 15 NNs are used in two or more dimensions (for the spatial background component and the trigger KDE).

The approach taken here differs from that of Mohler et al. (2011) in one important respect. In the original study, the temporal and spatial components of the Gaussian kernel used in the KDE for the triggering function were treated as equivalent, i.e. $k_T \equiv k_X \equiv k_Y$. However, this is incompatible with the requirement that $\Delta t_{ij} > 0$, because some of the density of k_T may lie in the negative time difference region. We therefore enforce this constraint by using a small modification to the KDE in which the temporal component k_T is reflected about $\Delta t_{ij} = 0$. No further normalisation is needed as this transformation preserves density. Further technical details are given in the Electronic Supplementary Material (ESM).

In the final step, the entries of P are updated using equations (2)-(3). The function g must be evaluated once for every pair of crimes (i, j) for which $p_{ij} \neq 0$, i.e. for which triggering is permissible. The number of evaluations is reduced through the application of threshold parameters, as described above.

This algorithm is carried out in successive iterations, with the aim of converging to an estimate of P. We find that between 50 and 100 iterations are sufficient to obtain convergence in all the examples considered here.

In order to verify the implementation of the SEPP, we first tested it using simulated data, following Mohler et al. (2011). In this process, data are simulated with known triggering and background risk functions, against which we may compare the estimates of μ and g to determine the accuracy of the SEPP. We show only the temporal and a single spatial component of g here (see Fig. 3); in both cases, we also include the result obtained without the reflected time component for comparison. This demonstrates a good agreement with those in the published study, with an improvement in the accuracy of the temporal triggering component due to our modification.

Applying the regular SEPP model to the two crime types in Chicago gives the spatial triggering profiles shown in Fig. 4 (red lines). There is significant bias towards n/s triggering in the S and SW burglary data, and in the FN, S, SW and FSE sides in the case of assaults. In all of these cases, the result suggests that crimes are triggered in the n/s axis over a distance an order of magnitude greater than in the east-west (e/w) axis. We define the bias in such cases as 'extreme'. An opposite bias of similar magnitude is apparent in W assaults. Whilst these results could feasibly represent the situation at the microscale (e.g. along a single road), the triggering function in the SEPP is global, in that g applies across the full spatial and temporal extent of each dataset. It is therefore entirely unrealistic that the observed biases can reflect a general mechanism underlying the datasets.

Second-Order Effects in the Crime Data

We now investigate what aspects of the various crime datasets may be contributing to the failure of the SEPP witnessed in the previous section. There are myriad methods of characterising a spatio-temporal point pattern such as a crime dataset. Since the triggering component of the SEPP model considers the links between pairs of crimes, we focus on the second-order properties of the crime datasets. Specifically, we seek to
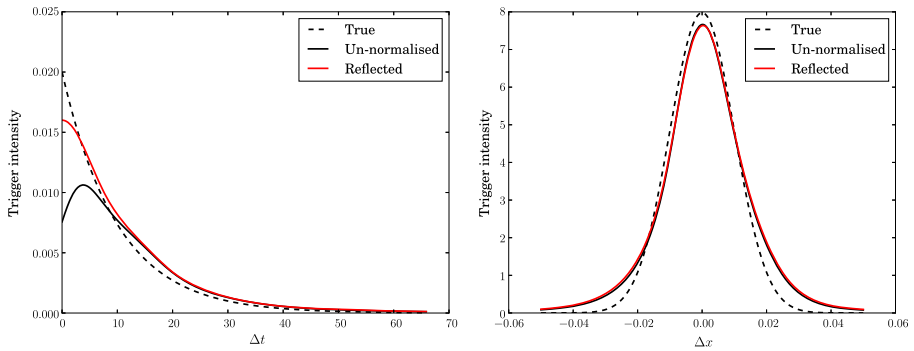
Fig. 3 Triggering structure in time and the first spatial dimension obtained by applying the SEPP to simulated data. The un-normalised results are very similar to those of Mohler et al. (2011)

determine how the point patterns differ between the nine different sides of Chicago, in order to relate this to any variation in the performance of the SEPP.

To quantify the second-order properties of the observed point processes, we use a modified version of Ripley's K function, in which directionality is considered in addition to distance (Dale 2000). We henceforth refer to this metric as the anisotropic Ripley's K function. We begin by computing a list of spatial differences between pairs of crime points, $(\Delta x_{ij}, \Delta y_{ij})$, where i and j refer to the indices of the two crimes in question. These are converted into distance and angular differences, Δd_{ij} and $\Delta \theta_{ij}$, with the latter computed as the angle relative to the positive x-axis made by the straight line connecting crime i to crime j. The estimator for the anisotropic Ripley's K function is given by

$$K_{\Theta}(u) = \frac{A}{N^2} \sum_{i=1}^{N} \sum_{j \neq i} \frac{1}{w_{ij}} I\left(\Delta d_{ij} < u; \Delta \theta_{ij} \in \Theta\right) \qquad (6)$$
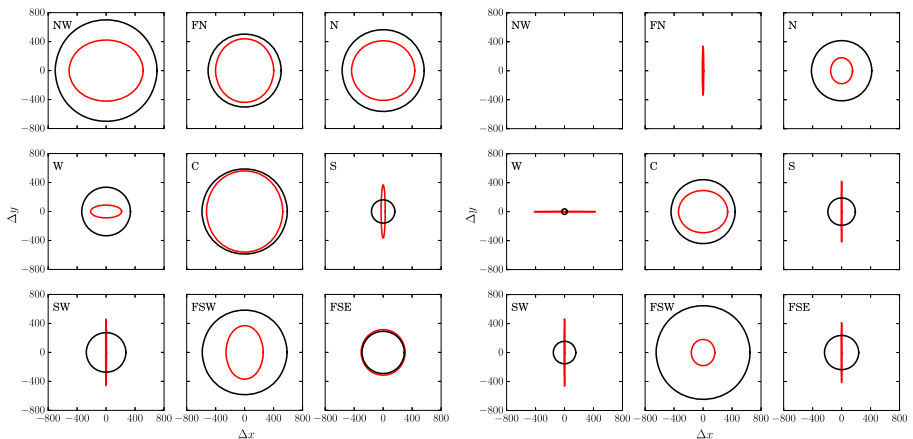

Fig. 4 The spatial triggering structure inferred by the regular SEPP (*red*) and isotropic SEPP (*black*) models for burglary (left) and assault (right) crimes in the different sides of Chicago. The shape plotted in each case contains 90 % of the total triggering density. Absence of lines indicates that no triggering effect was detected

where A is the area of the domain, N gives the number of crimes in the dataset, w_{ij} is the usual edge correction term (Gabriel et al. 2013) and Θ is the angular segment of interest. We define eight segments and pair them based on inversion about the axis (see top right inset in Fig. 5). Pairing the segments increases the number of data points available for the calculation in Equation (5). For comparison, we also compute this value for 100 repeats of a CSR model, in which the same number of crimes are deposited uniform randomly across the domain.

The anisotropic Ripley's K function is a purely spatial measure of second-order anisotropy. In the ESM, we also present analyses of the second-order space-time properties. The results are in excellent agreement with those presented here, suggesting that spatial effects are more relevant to the current discussion than temporal effects. As discussed in the introduction (see Fig. 1 and the related discussion), this is to be expected since there are many underlying spatial constraints on the crime data.

Applying the anisotropic Ripley's K function to the Chicago burglary dataset gives the results shown in Fig. 5. The different sides are best distinguished on the scale $u \leq$ 100m; a plot showing the region $u \leq 500$m is given in the ESM. The K values for all four angular segments lie above the CSR result, indicating significant aggregation at the 1 % level, in all cases but NW, where the diagonal aggregation is not significant. Aggregation is strongest in the n/s axis in 6 of the sides; in W, NW and N, the level of aggregation is similar in the n/s and e/w axes.
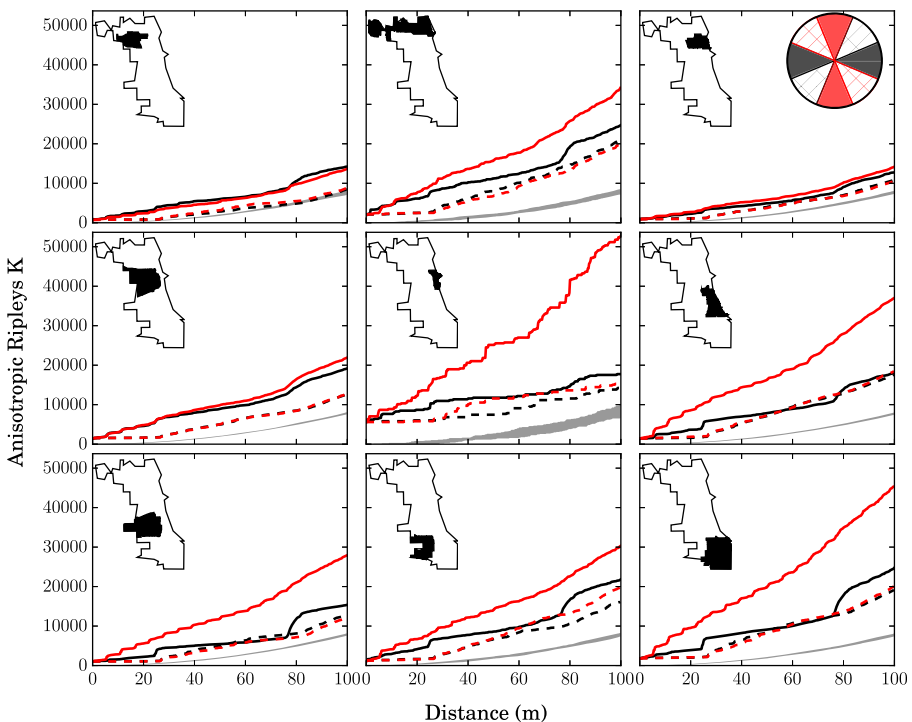

Fig. 5 Variation of the anisotropic Ripley's K function with distance for burglary crime in Chicago. The lower grey shaded region in each plot shows the expected result under the assumption of CSR (100 repeats, uniform random distribution of points). The top right inset shows how to interpret the remaining lines, with black denoting the east/west segments, etc. The domains are in the same order as in Fig. 4

The analogous plot for assault data is shown in Fig. 6 and shows a similar n/s bias to burglaries in S, SW, FSW and FSE. However, no such bias is observable in C and FN. The bias is reversed in C and W assault data, though in C this bias is only evident over the length scales around 30–70 m. The W assault data provides the only example of consistent e/w bias across all length scales.

Comparing the second-order properties discussed above with the SEPP spatial triggering structures in Fig. 4, we see that bias in the SEPP trigger is associated with strong second-order anisotropy in all cases. For example, the strong second-order n/s bias in S and SW is associated with a similar effect in the model for both assault and burglary data. Furthermore, W assaults are the only case where e/w aggregation is greater than n/s; this is associated with a major bias in the SEPP model. However, on the basis of the second-order structure, we would also expect that the SEPP triggering would be biased in the case of burglaries in FN and FSE. Despite this inconsistency, we conclude that the second-order structure is a reasonable, if not infallible, indicator of failure of the SEPP algorithm.

Isotropic Triggering

Having demonstrated significant levels of second-order anisotropy in several of the crime datasets and shown that this is positively associated with the failure of the SEPP algorithm, we now consider a variant of the SEPP in which the spatial component of the triggering function is assumed to be independent of angle, $g \equiv g(\Delta t, \Delta d)$, where Δd denotes the
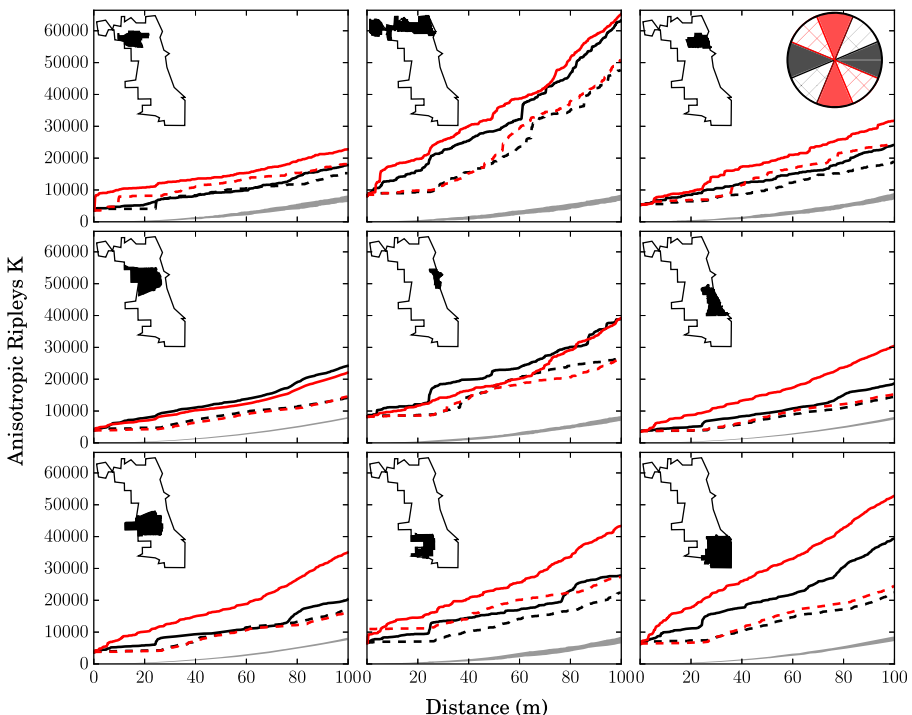

Fig. 6 Variation of the anisotropic Ripley's K function with distance for assault crime in Chicago. All details are as in Fig. 5

Euclidean distance from a crime event. In this form, triggering is isotropic, as it depends solely on the distance from the parent event and is independent of the direction.

To justify this model, we remark that the triggering component in the SEPP model is global: g is assumed to be valid over the whole region of interest. In reality, triggering varies at the street level, depending on the road network and various urban boundaries such as parks and railway lines. The triggering function therefore represents an aggregate of all local triggering functions in the dataset. The isotropic model assumes that the directionality in these local triggering functions will disappear in the aggregated (global) form. The isotropic model is therefore a trade-off: we replace a highly specific, directional, triggering function that is appropriate to a proportion of locations with a general, non-directional, function that is more appropriate on average.

Under the assumption of isotropy, Equation (1) becomes

$$g(\Delta t, \Delta d) = \sum k_T(\Delta t; \Delta t_{ij}, \sigma_{\Delta t_{ij}}) k_D(\Delta d; \Delta d_{ij}, \sigma_{\Delta d_{ij}}) \tag{7}$$

where the time component k_T is unchanged and $\Delta d_{ij} = \sqrt{(\Delta x_{ij})^2 + (\Delta y_{ij})^2}$. As for the standard SEPP, we use a Gaussian kernel function for k_D, however we must ensure that the kernel function k_D is normalised over the whole plane. The spatial normalisation constant is given by

$$K = \int_{-\pi}^{\pi} \int_{0}^{\infty} k_D(u; \Delta d_{ij}, \sigma_{\Delta d_{ij}}) du = 2\pi\sigma_{\Delta d_{ij}}^2 + \sigma_{\Delta d_{ij}}\Delta d_{ij}\sqrt{2\pi^3}\left[1 + \mathrm{erf}\left(\frac{\Delta d_{ij}}{\sigma_{\Delta d_{ij}}\sqrt{2}}\right)\right] \tag{8}$$

A derivation is provided in the ESM. Finally, we define the spatial bandwidths $\sigma_{\Delta d_{ij}}$. These are computed in an analogous process to the standard planar case. The input data $(\Delta t_{ij}, \Delta x_{ij}, \Delta y_{ij})$ are first converted to a radial form, $(\Delta t_{ij}, \Delta d_{ij})$, then the same variable bandwidth NN algorithm is then applied to calculate $\sigma_{\Delta t_{ij}}$ and $\sigma_{\Delta d_{ij}}$ (Mohler et al. 2011).

Having defined the normalisation constants and bandwidths, we may now substitute the isotropic KDE into the optimisation process described in "The Self-Exciting Point Process" Section with no further modifications. Figure 7 illustrates the difference between the regular and isotropic spatial KDEs. In this example, four data points are included. This may be considered a representation of the triggering structure, in which the parent crime is located at the origin. In the isotropic case, the triggering density is assumed to act in all directions, resulting in a ring-like triggering form.

Measuring Predictive Accuracy

We use the hit rate as a measure of predictive accuracy, as described in several preceding studies (Mohler et al. 2011; Bowers et al. 2004). Briefly, the hit rate is defined as the proportion of crimes within a pre-specified prediction time window falling in one or more predicted regions, termed 'hot spots'. We consider prediction time windows of 1 day throughout this study, as this is a typical time period for short-term police patrol planning. The overall coverage of the hot spots may be varied according to the resources available to
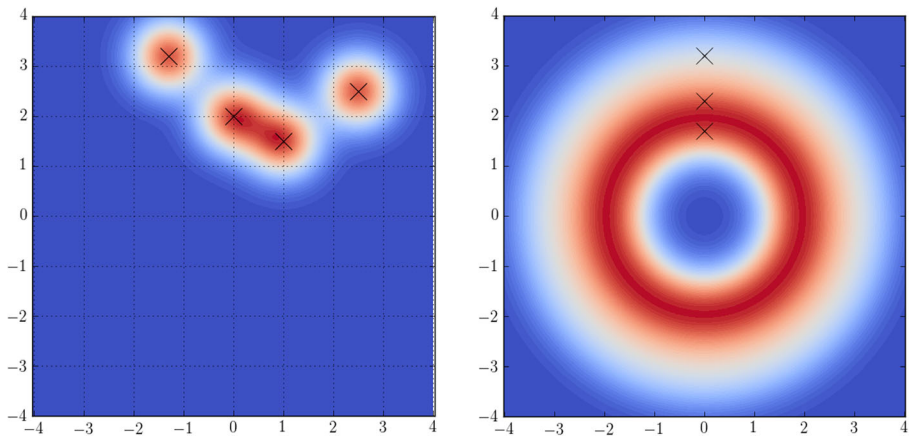
Fig. 7 Illustration of the difference between the standard anisotropic KDE (left) and the isotropic KDE (right) for four hypothetical data points. Blue shading indicates lower density and red higher. Black crosses indicate the centres of the constituent data points; on the right these are arbitrarily placed on the y-axis. Two of the points elide when converted to Euclidean distance

the local police force. We generate predictive hot spots by overlaying a 250 m × 250 m grid on the domain of interest and ranking the squares based on the mean value of the conditional intensity function (see Equation (1)). Grid squares are selected in descending intensity order to generate a series of coverage values. We consider coverage levels up to 20 %, since values above this are of little relevance to operational policing, where resources are too limited to provide high levels of police presence over a large area. Since the hit rate is dependent upon the distribution of crimes in a given prediction time window, this process is repeated over 100 consecutive time windows to reduce the effect of daily variation.

In order to determine the mean intensity in a given grid square, we evaluate the function $\lambda(t, x, y)$, where t is the start of the 1 day prediction time window, at 30 uniform random spatial locations within the square and take the mean value. This process is essentially identical to Monte Carlo integration.

When comparing the hit rate of two different prediction methods, previous studies have considered the mean and standard deviation of the hit rate, aggregated over all prediction time windows (Mohler et al. 2011; Chainey et al. 2008). This is a valid approach, but it does not take into account the fact that each prediction window has a pair of hit rates associated with it. To improve upon this approach, we make the assumption that the difference in hit rate between the two methods is independent of the daily number and distribution of crimes. We then treat the hit rates for the two methods, measured over 100 prediction time windows, as paired samples and apply the Wilcoxon Signed Rank (WSR) test to determine whether the observed results differ significantly (Adepeju et al. 2016).

Simulation Study

We now consider the effect of the previously observed second-order spatial anisotropy and extremely biased triggering functions on the predictive performance of the SEPP. We shall measure this directly for the Chicago crime data in "Predictive Accuracy" Section, however we first seek to demonstrate the link between predictive performance

and triggering bias using a systematic approach. As there is no common baseline for comparison in the real crime datasets, we develop a simulation method to analyse the variation of predictive accuracy with triggering bias.

In order to recreate some of the essential features that lead to second-order spatial anisotropy in the crime data, we simulate crimes occurring on a regular grid arrangement of roads in a 5 km by 5 km region, as shown in Fig. 8. The latitudinal spacing between roads is held constant at 100 m; the longitudinal spacing is varied, taking the values 100 m, 200 m, 400 m and 800 m. This creates a simulated street network that is increasingly dominated by roads that lie on the e/w axis. This is intended to represent a very simplistic model of an urban arrangement that is similarly dominated by roads with roughly equal orientations, which could arise as a result of geographical features such as those described earlier, including railway lines, waterways and terrain. Hence our simulation incorporates some of the possible geographical variation between different regions in Chicago.

Crimes are simulated using the SEPP model (Equation (1)) incorporating a spatially uniform background distribution on the network with a mean intensity of 5 events per day and a triggering function that is exponentially decaying in time and Gaussian in distance:

$$g(\Delta t, \Delta x, \Delta y) = I_t \frac{\alpha}{2\pi\beta^2} \exp(-\alpha\Delta t)\exp\left(\frac{-\Delta d^2}{2\beta^2}\right) \qquad (9)$$

where the distance Δd is measured along the edges of the network, $\alpha = 0.1$ day^{-1} is the temporal triggering decay, $\beta = 50$m is the spatial triggering bandwidth and $I_t = 0.2$ event^{-1} is the triggering intensity. Triggering is only permitted to occur along the spatial network. The simulation of triggering events proceeds as follows. Iterating over every simulated crime point, we simulate a non-stationary Poisson process with an exponentially decaying intensity to generate the number and time of triggered crimes. The locations of any triggered crimes are then generated by drawing a distance from a
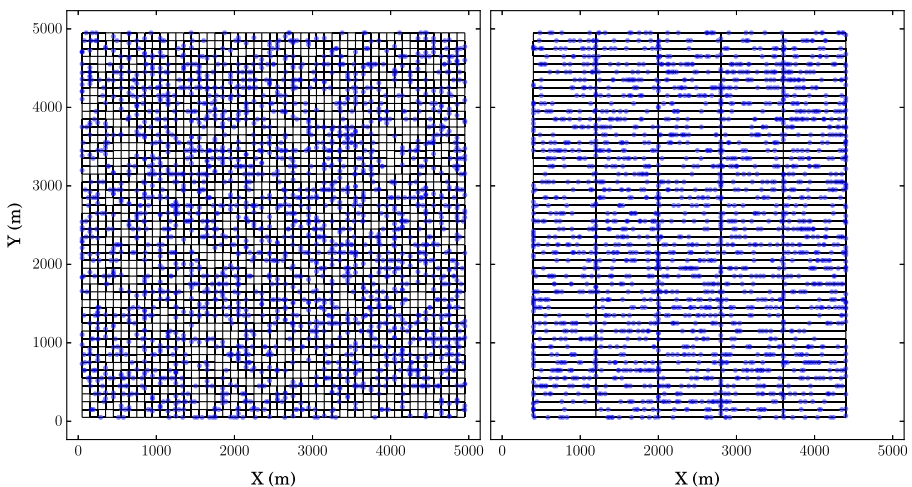

Fig. 8 The spatial arrangement of the simulated data for horizontal grid spacing values of 100 m (left) and 800 m (right). Black lines denote the underlying network, blue dots show the location of a simulated crime

normal distribution with zero mean and standard deviation β and performing a random walk on the network for that distance, starting at the parent point.

We proceed to train the regular SEPP model on the simulated data. The model and algorithm are configured as described in "The Self-Exciting Point Process" Section with one modification. The optimisation process began to generate numerical errors with a horizontal spacing greater than 200 m, since the NN bandwidth selection algorithm would generate a spatial bandwidth equal to zero whenever the randomly sampled trigger pairs consisted exclusively of pairs of points that lie in the same row (see Fig. 2 for reference). This would ordinarily be prevented in real datasets due to the imperfect alignment of the street network. We corrected this error by enforcing a minimum bandwidth of 5 m in the NN selection process. Other viable approaches include rotating the simulated grid network through a small angle and introducing a random component in the simulated roads.

The SEPP model inferred the spatial triggering structures and intensities illustrated in Fig. 9. The spatial triggering functions are similar and mildly skewed towards e/w triggering in the case of 100 m and 200 m horizontal spacing, and extremely biased towards e/w triggering in the case of 400 m and 800 m spacing. This is purely an artefact of the optimisation process: the generative triggering function in the simulated datasets is constant. The intensity of the triggering process is also substantially overestimated in the latter two cases. Both of these outcomes suggest that the SEPP optimisation algorithm converged to a state that inaccurately represents the data.

Finally, we evaluated the predictive accuracy of the SEPP in the case of 100 m and 800 m horizontal spacing. Due to the scale of the simulated networks, we found that the results showed greater sensitivity to variations in the triggering structure when the size of the evaluation grid was set to 50 m × 50 m grid. All details of the evaluation process were otherwise the same as described previously. The results are shown in Fig. 10. The SEPP trained on 100 m spaced data performs significantly better than the extremely biased SEPP trained on 800 m spaced data for coverage levels below 10 %. Above this, the mean hit rate
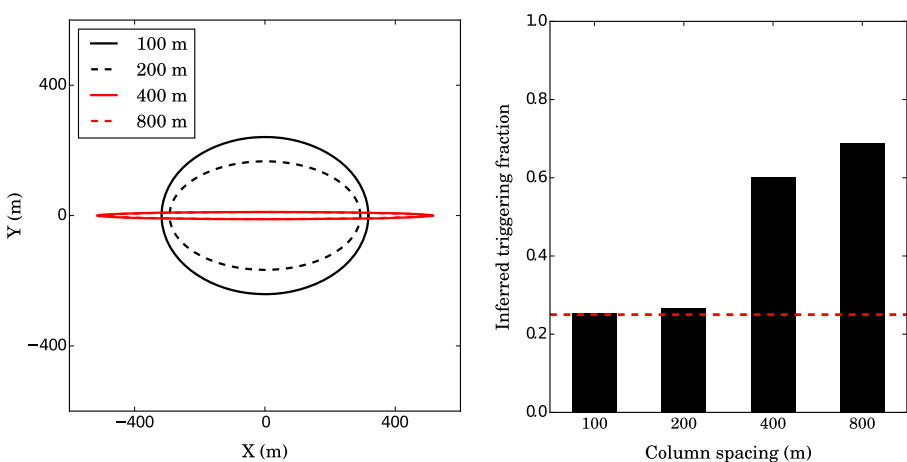

Fig. 9 Inferred triggering structure obtained by applying the regular SEPP to simulated data. (Left) contours enclosing 95 % of the spatial triggering density for the different horizontal grid spacings. The results for 400 m and 800 m are not visually distinguishable. (Right) the inferred strength of triggering. The red dashed line shows the true value used in the simulation
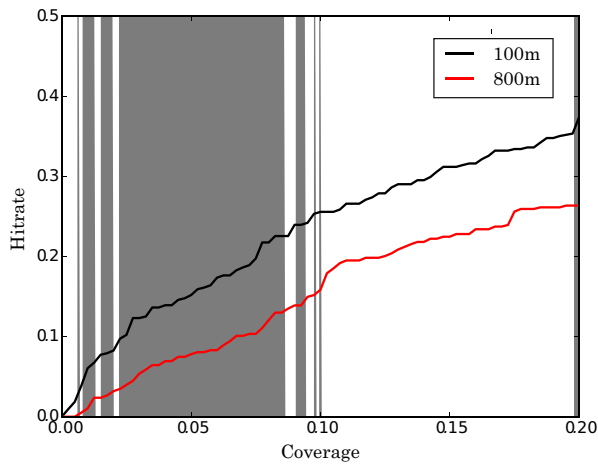
Fig. 10 Plot of mean hit rate against coverage level for simulated data with varying horizontal spacing. Shaded regions show the coverage levels for which there is a significant difference between the hit rate at the 5 % significance level (determined using the WSR test) and the difference in mean hit rate is ≥5%

remains higher by around an absolute difference of 10 % but the WSR test indicates that this is not statistically significant due to the day-to-day variation in the hit rates.

As we discuss below, this simulation study does not indicate a direct causal relationship between triggering bias and reduced predictive accuracy. However, by establishing a strong correlation between the two it provides further motivation for the proposed isotropic variant of the SEPP.

Results

Triggering Component in the Isotropic SEPP

Applying the isotropic variant of the SEPP to the same datasets, we obtain the spatial triggering structure shown in Fig. 4 (black lines), which is explicitly circular in all cases. In the case of burglary, the isotropic result differs noticeably from the original result in the W, S, SW and FSW sides. Three of these (W, S and SW) exhibit moderate to high levels of directional bias when treated with the regular model and were therefore expected to show a different outcome when paired with the isotropic variant. The situation is less clear in the case of assaults, with marked differences in the spatial structure in all cases. The triggering effect disappears entirely in the FN region when modelled as an isotropic SEPP, and becomes highly spatially constrained in W.

In addition to the spatial extents of the triggering structure, it is also straightforward to determine the inferred proportion of crimes occurring due to triggering by summing the off-diagonal entries of the matrix P. This measures the strength of the triggering process in each dataset. The results are shown in Fig. 11. Focusing on the anisotropic results first (black bars), we see that the triggering proportion varies a great deal between regions. The FN, NW, FSW and N sides have the lowest proportion of triggering – less than 25 % - for burglaries and assaults. Indeed, the SEPP algorithm detects no triggering process at all in the NW side assault data. The isotropic SEPP
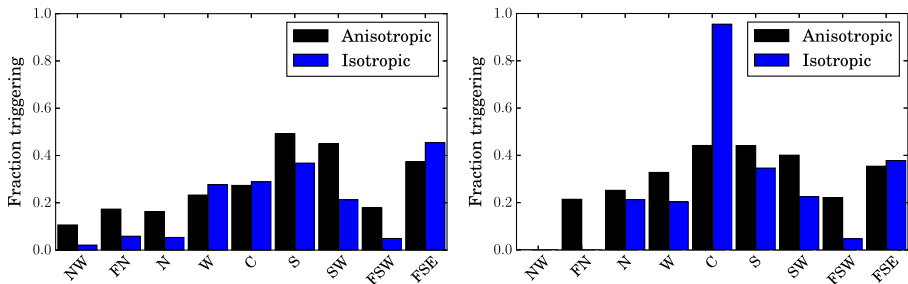
Fig. 11 Proportion of triggered crimes inferred by the SEPP model for burglary (left) and assault (right) crimes in the different sides of Chicago

approximately mirrors the crude trends between sides, but infers an even lower level of triggering in the NW, FN, N and FSW sides, such that triggering plays a very minor role (no role at all in the case of NW and FN assaults).

Predictive Accuracy

Whilst the triggering structure of the trained SEPP model is an important indicator, the ultimate goal of a crime prediction method is to support proactive policing strategies. We therefore compute the predictive accuracy of the trained models, which are plotted in Fig. 12 and Fig. 13. We also include results from applying a simple space-time KDE (STKDE, blue dashed lines) for reference purposes. The STKDE is implemented as a variable bandwidth KDE with bandwidths computed in the same way as they are for the triggering function.

The SEPP models (both regular and isotropic) match or improve upon the performance of the STKDE in all cases, with the exception of burglaries in the N side, where the STKDE has the best performance by a small margin. This results agrees with the findings of Mohler et al. (2011), who reported that the regular SEPP outperforms a conceptually similar method to the STKDE (Bowers et al. 2004).

Comparing the regular SEPP with the new isotropic variant, we observe no significant difference between the results in NW and FN. This is expected as Fig. 11 shows that both SEPP algorithms infer a low level of triggering in both of these sides and therefore altering the form of this component has little effect. The isotropic SEPP outperforms the regular variant by 5 % or more in SW for both crime types at coverage levels from 5 to 20 % in the case of burglary and more scattered coverage levels in the case of assault. A similar margin of improvement is observed in assaults in S and FSE. Referring to Fig. 4, we remark that the greatest improvements in predictive accuracy are found in the cases where the regular SEPP exhibits extreme triggering bias. This result demonstrates the benefit of the new isotropic method when applied to data with high levels of second-order anisotropy; not only are the resulting triggering structures more realistic, but the predictive performance is also increased.

In contrast, the regular algorithm shows greater performance in C burglary and a small coverage range of the FSW and N assault datasets. A comparison with Fig. 4 indicates that the isotropic model performs more poorly when the regular SEPP is unbiased in the case of assault crimes (FSE, C and N) , but that this relationship does not hold in the case of burglaries (for example, compare with FSE). Further work is
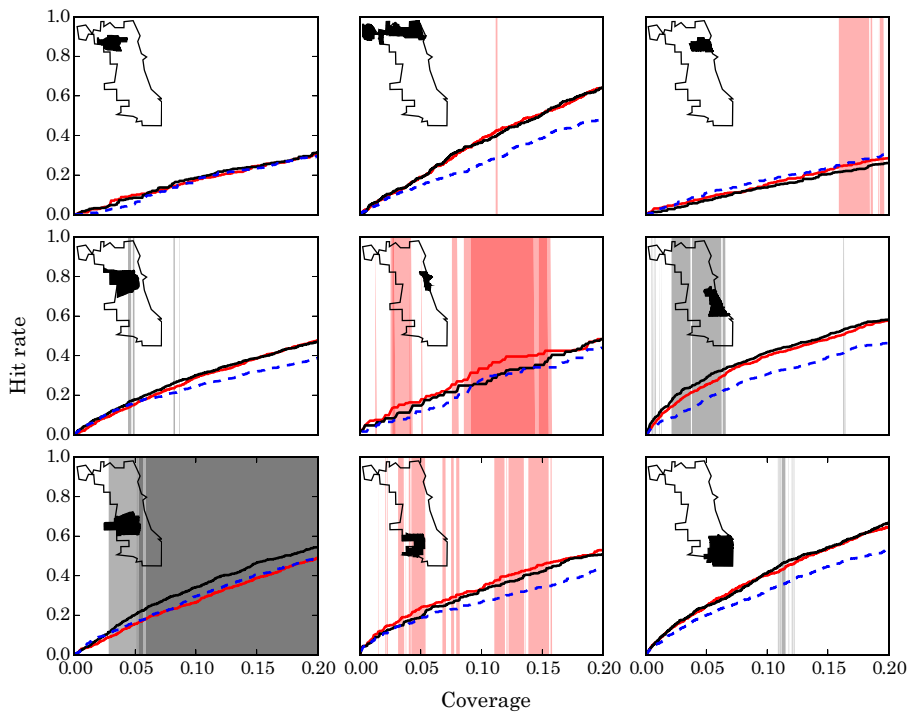
Fig. 12 Plot of mean hit rate against coverage level for burglary (left) and assault (right) crimes in the nine Chicago sides for the regular (anisotropic) SEPP (*red line*), the isotropic SEPP (*black line*) and STKDE (*blue dashed line*). Shaded regions show the coverage levels for which there is a significant difference between the regular and isotropic SEPP results at the 5 % significance level, as determined using the WSR test. The shading colour indicates that the isotropic (*black*) or regular (*red*) method has the stronger performance; lighter shading indicates regions where the difference in mean hit rate is ≥2.5%, for darker regions the difference is ≥5%. The domains are in the same order as in Fig. 4

required to determine the cause of the discrepancy. We note that the C burglary dataset contains relatively few records due to the small area and low residential property density. This may account for the high level of noise in the mean hit rate for this dataset, evident from the more 'jagged' appearance of the curve.

Discussion and Conclusions

Our original aim for this study was to apply the regular SEPP algorithm to the nine sides of Chicago, in order to test the hypothesis that the triggering structure would differ between crime types and by location. In particular, we opted to use the non-parametric form of the model, as proposed by Mohler et al. (2011). In practice, this approach highlighted a lack of robustness in the model, as the algorithm produced unrealistically biased triggering structures in several situations. As a result, we focused on achieving two main aims, namely identifying any characteristics in the crime data that are associated with the failure, and modifying the non-parametric SEPP algorithm to improve its robustness.
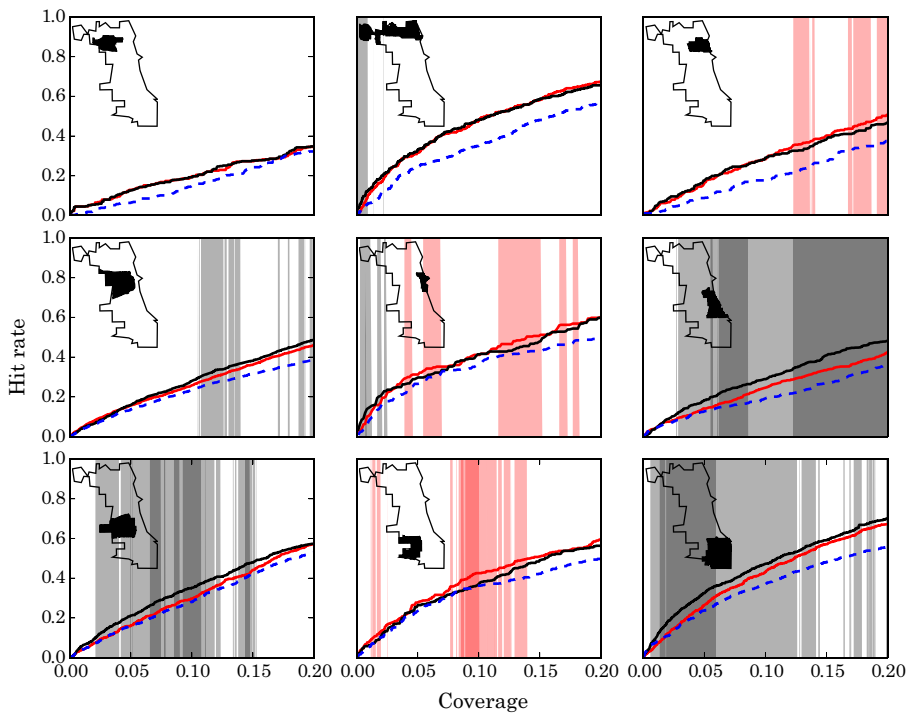
Fig. 13 Plot of mean hit rate against coverage level for assault crimes in the nine Chicago sides. All labelling follows same conventions as Fig. 12

We achieved the first aim by characterising the crime data in terms of second-order spatial properties. This was measured using a modification of Ripley's K function that varies by angle as well as distance. An additional approach that also includes second-order temporal properties is included in the ESM, however this is in almost complete agreement with the anisotropic Ripley's K results and is therefore omitted from the main text. In all cases of extremely biased inferred triggering, we identified significant second-order anisotropy in the data. However, in a further two cases the SEPP triggering structure remains unbiased despite the presence of anisotropy in the data. We conclude that this approach is a reasonable means of identifying problematic datasets in advance of applying the SEPP algorithm. Other methods of characterising the data may provide further insight into the issues faced by the model.

Having determined an association between second-order anisotropy and biased triggering structures, we presented a new isotropic variant of the SEPP, in which triggering is non-directional. The isotropic assumption is justified because the triggering function in the SEPP is global, although it is applied locally. We assume that a generalised triggering function is preferable to a more specific model that is inappropriate in a substantial proportion of locations. Although the isotropic process is conceptually a simplification, the technical details are not straightforward as the triggering KDE must be normalised correctly. This is in direct contrast with the parametric variant of the SEPP (see, for example, (Mohler 2014)), in which the isotropic form of the equations reduces the number of free parameters and simplifies the equations.

The isotropic SEPP is further motivated by the simulation study presented in "Simulation Study" Section. Using a simulation of an SEPP occurring on a street network, we showed that by varying the layout of the street network we are able to reproduce the inference of extremely biased triggering by the regular SEPP and correspondingly poor predictive performance. In the simulation, the generative triggering structure used was isotropic on the street network, therefore we conclude that the observed bias is purely due to the inability of the SEPP optimisation routine to deal with certain types of heterogeneity in the spatial point patterns of the data. This simulation approach constitutes an important systematic demonstration of the correlation between extreme bias in the inferred triggering functions and reduced predictive accuracy. However, the use of a regular network in our simulation study is clearly an oversimplification of the complex arrangements observed in real cities. Further work and more sophisticated simulation scenarios are required to gain deeper insight into the interplay between the SEPP optimisation process and the point patterns in the input data.

Applying the isotropic SEPP to the Chicago crime datasets, we observe more realistic spatial triggering structures in most cases. One possible exception is FN assaults, where the biased triggering component in the regular SEPP disappears completely when the isotropic variant is applied. Nonetheless, we observe no loss of predictive performance in this case.

Finally, we assessed the predictive accuracy of the two methods using the hit rate. The results indicate that the new method generally matches the performance of the regular SEPP, and generates a statistically significant improvement in predictive accuracy in the presence of second-order anisotropy. This improvement amounts to an additional 5 % of daily crime being correctly predicted at certain coverage levels in the range 0–20 %.

The predictive accuracy results suggest a simple rule of thumb for a police practitioner, in order to guide them in algorithm selection. If time and computing resources permit, then first attempting to train the regular SEPP model will determine whether the inferred triggering shows extreme bias (as is the case for SW burglaries, for example). In these cases, our results indicate that the isotropic variant is likely to perform significantly better. If this is not practical, opting for the isotropic model is advisable, since the potential improvements are large and significant, while possible reductions in predictive performance are more minor. Furthermore, we would caution against using the output of the Ripley's anisotropic K function as a sole determinant of algorithm choice, since the results, whilst strongly correlated with extreme triggering bias, are not infallible.

We conclude that our isotropic SEPP algorithm is an improvement over the regular algorithm, as it is both more robust and more accurate. This is a significant contribution to the field of predictive policing, as it facilitates the operational application of the SEPP without the need for commercial software, which has previously been hampered by the lack of detailed information about the algorithm, in addition to its variable performance due to low robustness.

We opted to focus on the non-parametric form of the SEPP for this study, following Mohler et al. (2011). Since the isotropic parametric variant has already been applied elsewhere (Mohler 2014), this study paves the way for a comparison of the two approaches to determine whether they differ in the inferred triggering structures or predictive accuracy. This is beyond the scope of the present study, but warrants further work.

In terms of crime prevention policy, we note that the concept of spatial coverage as an independent variable in our prediction accuracy measurement, whilst commonplace

in the crime prediction literature, is too simplistic to permit direct translation into policing practice. This limitation is ubiquitous in the field of crime analysis and additional work is much needed to bridge the gap between the theoretical hit rate and practical outcomes. Numerous studies have considered the effect of increased police presence on crime rates based on controlled trials (Sherman and Weisburd 1995; Weisburd and Eck 2004) or anomalous events (Di Tella and Schargrodsky 2004), however these studies all involved substantial increases in police numbers in key areas. It is less clear whether more modest changes have any effect (Bradford 2011). Furthermore, there is to our knowledge no evidence on the effects of changing police patrolling behaviour by using a new prediction method.

With these points in mind, we are unable to comment on the range of coverage levels that are of interest to a given police force. Instead, we provide the hit rate measured over a range of coverage values and ensure that the upper level (20 %) is sufficiently large that few, if any, urban police forces can feasibly achieve it. A rescaled version of Fig. 12 and Fig. 13 would be required for a police force concerned with coverage rates below 2 %. Nonetheless, the methodology presented here can be applied without modification in such cases, though the conclusions may differ. Finally, we remark that many metrics other than hit rate may be applied to characterise a crime prediction method (Adepeju et al. 2016), however computing such quantities is beyond the scope of the present study.

The data used in this study are all taken from a single year in the City of Chicago. We divided the dataset based on pre-existing administrative boundaries in order to identify how the SEPP algorithm performs differently. This was highly instructive, however we note that the dataset is still derived from a single urban region. For reasons of space, it was not possible to include further datasets in this study, however further work is required to determine whether our approach remains optimal when applied to data from other urban police departments worldwide.

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*. Published online. Available at https://doi.org/10.1080/13658816.2016.1159684.

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.

Bradford, B. (2011). *Police numbers and crime rates – a rapid evidence review.* London: United Kingdom.

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.

Dale, M. R. T. (2000). *Spatial pattern analysis in plant ecology.* Cambridge University Press.

Efraimidis, P. S., & Spirakis, P. G. (2006). Weighted random sampling with a reservoir. *Information Processing Letters, 97*(5), 181–185.

Gabriel, E., Rowlingson, B. & Diggle, P., 2013. stpp: an R package for plotting, simulating and analyzing Spatio-Temporal Point Patterns. *Journal of Statistical Software*, 53(2).

Groff, E., & La Vigne, N. (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies, 13*, 29–57.

Johnson, S. D., et al. (2009). Predictive mapping of crime by ProMap: accuracy, units of analysis, and the environmental backcloth. In D. Weisburd, W. Bernasco, & G. J. N. Bruinsma (Eds.), *Putting crime in its place* (Ed.). Springer New York: New York, NY.

Johnson, S. D., et al. (2007). Space–time patterns of risk: a cross National Assessment of residential burglary victimization. *Journal of Quantitative Criminology, 23*(3), 201–219.

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., et al. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association, 106*(493), 100–108.

Musmeci, F., & Vere-Jones, D. (1992). A space-time clustering model for historical earthquakes. *Annals of the Institute of Statistical Mathematics, 44*(1), 1–11.

Pease, K., 1998. Repeat victimisation: taking stock, Home Office.

Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": a randomized, controlled trial. *Justice Quarterly, 12*(4), 625–648.

Sparks, R.F., 1981. Multiple Victimization: Evidence, Theory, and Future Research. *The Journal of Criminal Law and Criminology*, 72(2), p.762.

Di Tella, R., & Schargrodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review, 94*(1), 115–133.

Weisburd, D., & Eck, J. E. (2004). What can police do to reduce crime, disorder, and fear? *The Annals of the American Academy of Political and Social Science, 593*(1), 42–65.

Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association, 97*(458), 369–380.