A self-exciting point process model for predictive policing: implementation and evaluation

G. Rosser[*1] and T. Cheng[†1]

[1] SpaceTimeLab, Department of Civil, Environmental & Geomatic Engineering, University College London, London WC1E 6BT

November, 2014

Summary

The self-exciting point process (SEPP) model has recently been shown to perform well in predicting spatiotemporal crime patterns. However, this model has not been widely applied to crime data and many open questions remain about how best to implement it in a real setting. In this work, we consider a range of practical implementation details relating to the application of SEPP models to real crime data. We propose a robust protocol that optimises the performance of the method, and suggest guidelines for parameter selection.

KEYWORDS: predictive policing, self-exciting point process, kernel density estimate, machine learning

1. Introduction

The criminological theory of near repeat victimisation states that the occurrence of certain crimes increases the risk of further crimes within the local neighbourhood for some ensuing time period (Johnson and Bowers, 2004; Youstin et al., 2011). As a result of this process, crime events tend to cluster in space and time. Predictive policing is concerned with identifying emerging crime 'hotspots' using forecasting methods. This has been the target of much recent research effort (S. Chainey et al., 2002; Bowers et al., 2004) as such a method would be of great utility to police forces worldwide. Existing methods commonly apply statistical analysis or heuristic algorithms to crime data with the aim of identifying hotspots and localising them in time and space (see, for example, Bowers et al., 2004). Such approaches are valuable and have shown reasonable predictive power, however they are better suited to retrospective analysis than forecasting since the underlying methods are not based on well-stated models. Furthermore, such analyses give little insight into the underlying method of generation of crime patterns.

The subject of this work is the application of a self-exciting point process (SEPP) model to crime data. The point process framework is well suited for time and geolocation tagged data, such as records of crime. Methods based on point processes have previously been developed to detect space-time clustering (Diggle et al., 1995), which is useful for retrospective analysis. The SEPP model has been used in the field of seismology to predict earthquake sequences for several decades. As we describe in detail below, this model describes a dynamic point process in space and time in which events may trigger further events within their spatial and temporal neighbourhood (self-excitation). In a promising advance in the field of criminology, Mohler et al. noted the similarity between this model and the criminological theory of near repeat victimisation, and applied the SEPP model to the predictive modelling of crime data (Mohler et al., 2011). Their method outperforms a kernel-based hotspot

[*] g.rosser@ucl.ac.uk
[†] t.cheng@ucl.ac.uk

detection approach in terms of predictions made on real crime data.

Despite the apparent advantages of the SEPP framework for predictive crime modelling, there are several open questions and issues preventing the widespread adoption of the method. Most notably, the process of training the model (i.e. inferring parameters) on data involves the use of kernel density estimates (KDE), whose underlying kernel functions and bandwidths may have a significant effect on the predictive performance of the model (Mohler et al., 2011), or prevent the training algorithm from terminating successfully. In addition, the method is computationally intensive due to the necessity of repeatedly evaluating the KDE at a large number of data points (typically millions to tens of millions per iteration). Finally, there is no available open source implementation of the SEPP model for crime data, which hampers further research and development of the methods discussed.

The subject of this abstract is the development of a robust computational tool to apply the SEPP to crime data. We consider the following real-world implementation issues: (a) the effect of imposing upper thresholds on the temporal and spatial maximum triggering extents; (b) the effect of changing the temporal or areal domain upon which the model is trained; (c) the effect of modifying the kernel functions in the KDE.

We assess the predictive performance of our method using appropriate validation methods, such as the measure of search efficiency rate. We apply our method to open crime data provided by the city of Chicago, USA, to demonstrate its effectiveness.

2. Materials and Methods

2.1. Self-exciting point process

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. The conditional intensity may be described as the sum of background and triggered events:

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{\{k: t_k < t\}} g(t - t_k, x - x_k, y - y_k), \tag{1}$$

where μ denotes the background occurrence rate and g denotes the triggering function. Thus all crimes that have occurred prior to a given time may theoretically contribute some additional expectation of the current crime activity, though in practice this may vanish 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. In practice, this entails declustering the data (Zhuang et al., 2002) to identify those events arising from background activity and those triggered by previous events. Common approaches in the seismology literature involve maximum likelihood estimates based on assumed forms of μ and g (Daley and Vere-Jones, 2003). An alternative approach, employed by Mohler et al. (2011), employs KDEs to avoid this necessity (Zhuang, 2006).

Let p_{ji} denote the probability that event i was triggered by event j. By convention, p_{ii} denotes the probability that event i is a background event. Furthermore, $p_{ji} = 0$ if $t_i < t_j$, so that all of the probabilities may be encoded in an upper triangular matrix P. Under the assumptions of equation (1), these probabilities are given by

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

$$p_{ji} = \frac{g\left(t_i - t_j, x_i - x_j, y_i - y_j\right)}{\lambda(t_i, x_i, y_i)}. \tag{3}$$

In (Mohler et al., 2011), an optimisation routine is proposed in which the background and parent/child events are sampled randomly from the data using the probabilities in P. From these samples, a KDE is computed and P is updated following equations (2) and (3). This algorithm has been validated using simulated data.

2.1.1. Triggering thresholds

In order to calculate p_{ji} (equations (2) and (3)), the KDE g must be evaluated $N(N-1)/2$ times, where N is the number of data points in the dataset. This step becomes prohibitively computationally expensive with increasing datasets. To limit the number of evaluations, it is necessary to impose maximum temporal and spatial extents, Δt_{max} and Δd_{max}, respectively, above which triggering is discounted. The matrix P thus becomes sparse, with non-zero entries only where pairs of data points lie within the threshold.

2.1.2. Kernel function

Both μ and g are inferred using a KDE. In the case of g, triggering is only realistic when the time difference, $\Delta t = t_i - t_j$, is positive. However, a three-dimensional multivariate Gaussian kernel function is used in (Mohler et al., 2011), which permits density at negative Δt. We consider the effect of changing the kernel function used for the temporal component of the triggering KDE, shown in Figure 1 The spatial kernel functions are unchanged.

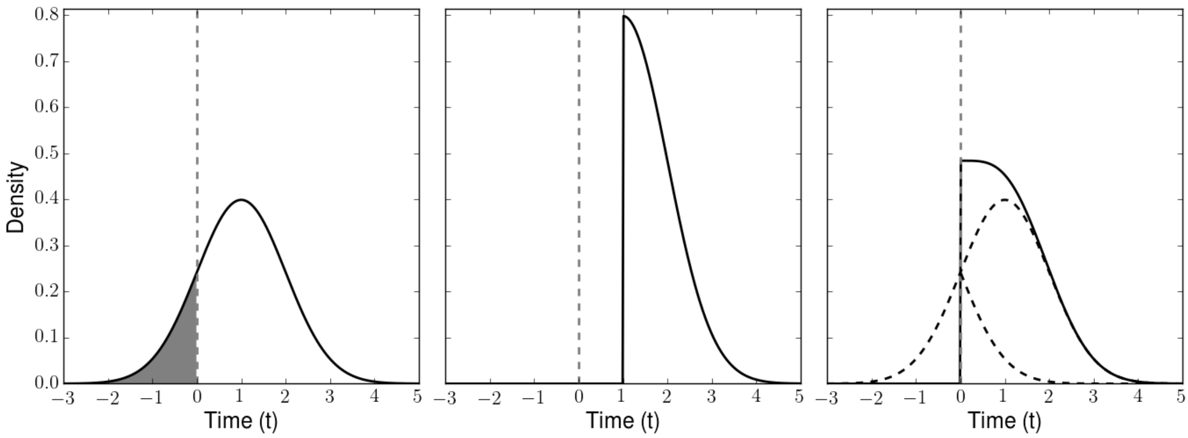

Figure 1 Three kernel functions considered for the temporal component of the triggering KDE. The function plotted is the marginal pdf in the temporal dimension, with a mean of 1. (Left) standard Gaussian, shaded region indicates density at negative time differences; (centre) one-sided Gaussian; (right) Gaussian reflected at $\Delta t = 0$, dashed lines illustrate the reflected portion.

2.1.3. Effect of temporal and spatial domain

Varying the spatial or temporal domain bounding the training data affects the performance of the SEPP model (Mohler et al., 2011) in an unknown manner. We are currently assessing the effect of spatial and temporal domain translation and enlargement, as indicated in **Figure 2**. Results will be presented at the GISRUK 2015 conference.

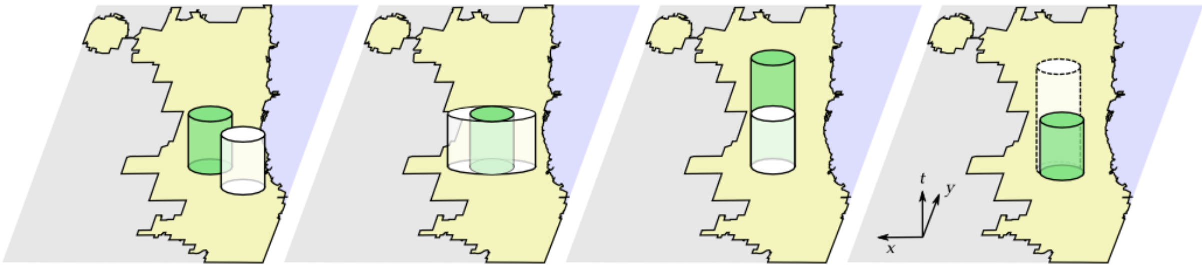

Figure 2 Illustration of the spatial and temporal domains within the city of Chicago used to train the SEPP model. (From left): spatial translation; spatial enlargement; temporal translation; temporal enlargement.

3. Results

3.1. Applying SEPP to Chicago crime data

For this study we are using crime data available on the City of Chicago's online data portal at https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2. Georeferenced, timestamped data are available from 2001 to the present.

Figure 3 shows the crime density heatmap computed using the SEPP for burglaries in February 2001.

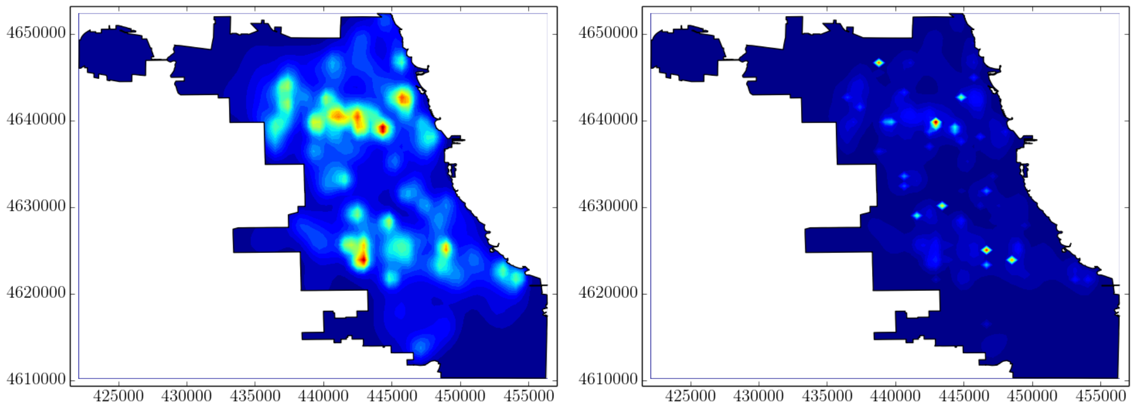

Figure 3 Density maps computed using the SEPP for burglaries in the Chicago region in February 2001. (Left) background density; (right) combined background / trigger density.

3.2. Varying triggering thresholds

As Table 1 indicates, the number of permissible triggering pairs in the SEPP model changes significantly with the spatial and temporal threshold limits. Even with unrestrictive values of these limits, the number of links only reaches 10% of the maximum. Work is ongoing to assess the effect of these thresholds on the predictive performance of the SEPP model.

Table 1 Variation of the number of triggering pairs in the SEPP model with trigger thresholds. The number of data points included was 11902, giving a maximum of 70822851 pairs.

Δt_{max} (days)

Δd_{max} (metres)	10	20	30	40	50	60	90
20	528	773	948	1068	1202	1304	1567
50	842	1336	1713	2047	2352	2621	3257
100	1545	2652	3589	4439	5231	5936	7587
200	4478	8162	11362	14409	17159	19750	25830
300	8732	16248	22878	29206	34886	40207	53078
500	20942	39913	56860	72646	87090	100485	133398
1000	70184	134824	194304	248863	299809	346543	464228
5000	1144490	2206618	3192283	4113140	4978984	5781365	7800521

3.3. Effect of the choice of kernel function

Figure 4 shows the effect of the choice of temporal kernel function on the inferred form of the triggering function g. The kernel functions used are those shown in Figure 1. In all cases, the triggering intensity is greatest immediately following a burglary, decreasing over the course of 3 days. However, the kernels lead to different estimates for g, with the one-sided variant being less smooth and the reflected variant having greater density within the first day following a burglary.

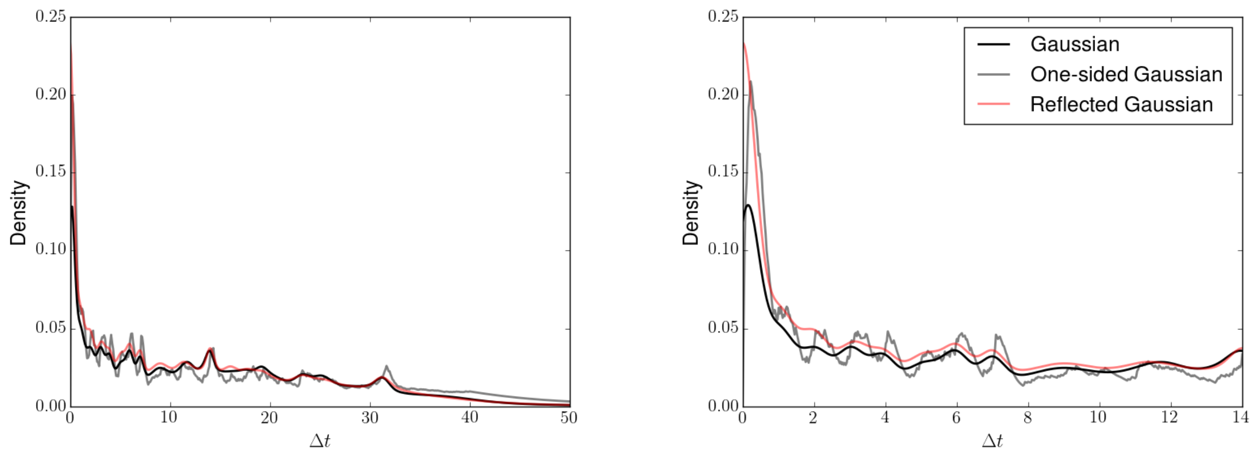

Figure 4 The effect of kernel function selection on the temporal component of the triggering intensity g, computed using Chicago burglary data from 1/1/2010 to 1/7/2010. The two plots show the same data on different scales.

4. Acknowledgements

This work is part of the project Crime, Policing and Citizenship (CPC): Space-Time Interactions of Dynamic Networks (www.ucl.ac.uk/cpc), supported by the UK Engineering and Physical Sciences Research Council (EP/J004197/1).

5. Biography

Tao Cheng is a Professor in GeoInformatics and Director of the SpaceTimeLab for Big Data Analytics (http://www.ucl.ac.uk/spacetimelab) at University College London. Her research interests span network complexity, geocomputation, integrated spatio-temporal analytics and big data mining (modelling, prediction, clustering, visualisation and simulation), with applications in transport, crime, health, social media, and environmental monitoring.

Gabriel Rosser is a research associate in the SpaceTimeLab working on the Crime, Policing and Citizenship (CPC) project. His research interests include GIScience, probabilistic modelling, predictive algorithms and machine learning.

References

Bowers KJ, Johnson SD and Pease K, 2004. Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology*, 44:641-658.

Chainey S, Reid S and Stuart N, 2002. When is a hotspot a hotspot? In *Socio-Economic Applications of Geographic Information Science*, CRC Press.

Daley D and Vere-Jones D, 2003. An Introduction to the Theory of Point Processes (2nd ed.), Springer, New York.

Diggle PJ, Chetwynd AG, Haggkvist R and Morris SE, 1995. Second-order analysis of space-time clustering. *Statistical Methods in Medical Research,* 4:124.

Johnson SD and Bowers KJ, 2004. The burglary as clue to the future: The beginnings of prospective hot-spotting. *European Journal of Criminology,* 1:237.

Mohler GO, Short MB, Brantingham PJ, Schoenberg FP and Tita GE, 2011. Self-exciting point process modelling of crime. *Journal of the American Statistical Association*, 106(493):100-108.

Youstin TJ, Nobles MR, Ward JT and Cook CL, 2011. Assessing the generalizability of the near repeat phenomenon. *Criminal Justice and Behaviour,* 38:1042.

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

Zhuang J, 2006. Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society B,* 68(4):635-653.