Geodetic Distance and Dynamic Outlier Exclusion in EM optimization of Self Exciting Point Process for Homicide Prediction in Chicago

Abstract—**In this paper, we propose a novel algorithm to improve the state-of-the-art results of homicide prediction in Chicago crime dataset. A marked self-exciting point process (M-SEPP) based epidemic type aftershock sequence (ETAS) model was applied to the Chicago crime dataset by Mohler [1] for improving the prediction rate of homicides over and above the traditional chronic hot spot approach. Expectation-maximization (EM)-type optimization has long been employed for the computation of the maximum likelihood estimates of parameters in the ETAS model. However, improvement in crime prediction has been slow due to the technological challenges in modeling the spatio-temporal distribution of criminal behavior, which results from a complex interaction of biological, social, psychological, and other influencing factors. We propose the *GeoDOME* algorithm, which incorporates geodetic distance and dynamic outlier exclusion in the EM-type algorithm of the ETAS model. It guarantees significant improvement in the prediction rate of homicides compared to the best result documented to date. A theoretical basis for the principles used in the modified algorithm is also provided. The increase in the prediction accuracy of homicides has been experimentally validated using the same Chicago crime dataset.**

Index Terms—**Epidemic type aftershock sequence (ETAS), Expectation-maximization (EM), Geodetic distance, Homicide prediction, Marked self-exciting point process, Maximum likelihood estimate (MLE), Outlier exclusion**

I. Introduction

The high density of street crimes in cities across the world has always prompted researchers to explore and innovate new methods of crime reduction. Hot spot policing, problem-oriented policing, and community policing are some of the well-known models of crime prevention. However, resources at the disposal of law enforcement agencies are increasingly becoming inadequate to keep pace with the complex and fast-changing crime scenario and a high expectation for public safety and security. In the last decade, the advances in computational power and the availability of massive data have prompted police authorities to harness the data-driven predictive models for strengthening their crime reduction efforts. A growing body of research has ample evidence that predictive policing not only augments but also exceeds in performance and efficiency over other methods of crime reduction.

In the past, authors have analyzed crime using data mining techniques. A novel multivariate time series (MTS) clustering was proposed by Chandra et al. [2] in combination with the similarity measure of dynamic time warping (DTW) and the Malmquist data envelopment analysis (DEA) model. It is used for grouping similarly performing police administrative units and ranking them based on their crime prevention performance. A comprehensive intelligent police information system (IPS) was designed and developed by Gupta et al. [3] for the National Crime Record Bureau (NCRB), India. The IPS combines innovative data mining techniques such as Malmquist DEA model and parametric Minkowski distance measure with multi-criteria decision making. It functions as a decision support system (DSS) for the law enforcement functionaries at all levels, in extracting crime hot spots, predicting crime trends, and undertaking the assessment of police administrative units on their crime prevention measures.

Researchers have also used classification techniques in machine learning for crime prediction. Kang et al. [4] have found the deep learning approach outperforming the support vector machine (SVM) and kernel density estimate (KDE), in predicting crimes in Chicago, using multimodal data from domains such as law enforcement, demographics, education, and economics. Dash et al. [5] used a network analytic approach for predicting crime counts in Chicago and reported that the support vector regression (SVR) exceeded in performance over auto-regression and polynomial regression both. However, the prediction accuracies of these works are still low.

An unmarked self-exciting point process was implemented using a nonparametric expectation-maximization (EM)-type algorithm by Mohler et al. [6] for predicting burglary in Los Angeles. This model was put into practice by the Los Angeles police department, achieving a significant reduction in burglary by 7.4% [7]. Recently, Kajita et al. [8] have used the Green function instead of the EM-type algorithm to compute the cascading effects in near-repeat crimes, showing improvement in the average prediction rate compared to that of [6], in respect of only high volume crimes.

To predict homicides, Mohler [1] implemented a marked self-exciting point process (M-SEPP) on Chicago crime dataset, using a parametric EM-type algorithm and reported a 17 % improvement in predictive accuracy over the chronic hot spot method. In order to deal with the relatively smaller homicide dataset, this model incorporates every crime type (ie. mark) of violent nature- involving use of handgun- as predcitor of homicides. The utility of this model has been demonstrated on a larger dataset of Pittsburgh, Pennsylvania by Reinhart et al. [9] and, by Zhuang et al. [10] using robbery data of Castollen, Spain.

The prediction rates of homicides reported in [1], [9] and, [10] are not high enough for a widespread adoption of these models in their present forms by the police departments across the world. In the backdrop of technological and practical challenges in achieving higher crime prediction rate, this paper aims to improve the prediction percentage of homicides in [1], using innovative methods of geodetic distance and dynamic outlier exclusion. We propose a modified algorithm- 'Geodetic Dynamic Outlier Exclusion in Parametric EM-type Optimization of Marked Self-Exciting Point Process (*GeoDOME*), which guarantees significant improvement over state of the art. The superiority of the proposed model has been proved theoretically as well as experimentally using the same Chicago crime dataset as used in [1]. Since there is no improvement in prediction rate of homicide in [9] and [10] compared to that of [1], we compare the improved results of *GeoDOME* algorithm with that of the latter only.

The remainder of this paper is organized as follows. The second section deals with the marked self-exciting point process for homicide prediction. In the third section, the modified algorithm *GeoDOME* is described along with its theoretical basis. The improved results of the proposed model is compared with that of [1] in the fourth section. We also present a metric for an optimal range of grid cells for predictive police deployment. The final section presents our conclusions.

II. Marked Self-Exciting Point Process for Prediction of Homicides

The motivation of this paper is to improve the prediction rate of homicide reported in [1], which uses the marked self-exciting point process (SEPP) [11] and the ETAS model [12].

Self-exciting point process (SEPP) is based on the computation of conditional intensity function, which is the space-time varying rate parameter (λ) of the underlying Poisson point process. It has two additive components- background (μ) and triggering (g) intensity functions, representing the background and triggering effects of past events, respectively. Epidemic-type aftershock sequence (ETAS) [12], is a branching structure model of the self-exciting point process (SEPP), frequently used for prediction in epidemiology and seismology.

We now give a brief overview of [6], followed by [1].

Mohler et al. [6] used a non-parametric conditional intensity function for implementing the unmarked self-exciting point process-based ETAS model to predict burglary in Los Angeles. Subsequently, a parametric conditional intensity function was applied by Mohler [1] in the marked self-exciting point process (M-SEPP)-based ETAS model for predicting homicides using Chicago crime dataset. The techniques used in this model is described below.

Let λ denote the conditional intensity function for homicides at a given point in space-time. It measures the expected number of homicide events per unit area, per unit time, at a point in space-time, given the history of crime events till that time. λ is the sum of background and triggering intensity functions denoted by μ and g, respectively. Let m_i denote the mark, i.e., crime-type of i^{th} crime event and (x, y, t) refer to space-time coordinates, consisting of the geographic coordinate system (i.e., latitude, longitude) and 'day' as the unit of time. For the marked self-exciting point process, λ_{ij}, μ_{ij} and g_{ij}, at (x_j, y_j, t_j) of j^{th} homicide, due to i^{th} crime event at $(x_i, y_i, t_i, m_i; t_j > t_i)$, are given as [1]:
Conditional Intensity Function:

$$\lambda_{i,j} = \mu_{i,j} + g_{i,j} \tag{1}$$

Background intensity:

$$\mu_{i,j} = \frac{\alpha_{m_i}}{t_j} \cdot \frac{1}{2\pi\eta^2} \mathrm{e}^{-\frac{(x_j - x_i)^2 + (y_j - y_i)^2}{2\eta^2}} \tag{2}$$

Triggering intensity:

$$g_{i,j} = \theta_{m_i} \cdot \omega \mathrm{e}^{-\omega(t_j - t_i)} \cdot \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\frac{(x_j - x_i)^2 + (y_j - y_i)^2}{2\sigma^2}} \tag{3}$$

where, α_{m_i}, θ_{m_i} are the branching factors, denoting the average number of homicides caused by a crime event of mark m_i (predictor of homicides), through background and triggering effects, respectively. η and σ denote the spatial bandwidths in the Gaussian kernels of μ and g, respectively (Eqn: (2) & (3)). ω is the temporal bandwidth in the exponential function used in g (Eqn: (3)).

Starting with an initial guess for the parameters, the model in [1] was trained using an algorithm inspired by the expectation-maximization (EM)-type implementation of the ETAS model in seismology [13]. Chicago crime dataset of the year 2007-12, comprising homicides and five other types of crimes, involving the use of 'handgun'- robbery, assault, weapon violation, battery, and criminal sexual assault- was used to train the model and to predict homicides. The prediction rate reported in [1], though acclaimed to be the highest to date, is still low and that has motivated us to propose the *GeoDOME* algorithm as described in the next section. The improved prediction rate of the proposed model is compared with that of [1] in the results section.

III. Geodetic Dynamic Outlier Exclusion in Parametric EM-type Optimization of Marked Self Exciting Point Process (*GeoDOME*)

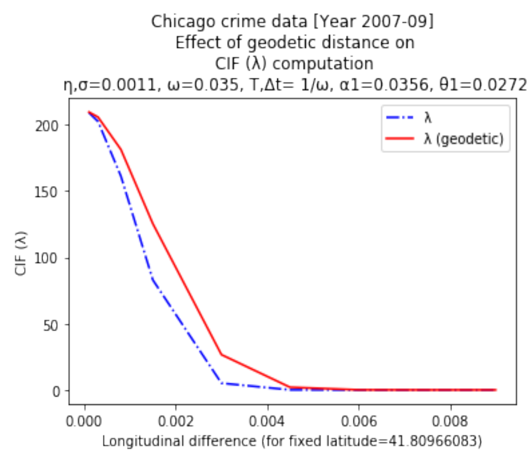
This paper aims to improve the prediction rate for homicides, using innovative methods of detecting and excluding outliers at every iteration of the EM-type algorithm and using geodetic distance instead of planar Euclidean distance. The algorithm is termed as *GeoDOME*. Varying the number of outliers to be excluded, makes a difference in the rate of prediction of homicides.

Subsection III-A gives the advantages of using geodetic distance in the computation of conditional intensity function. In subsection III-B, procedure and justification for dynamic detection and exclusion of homicide outliers, at every iteration of the EM-type algorithm, is provided. Algorithms for incorporating dynamic outlier exclusion and geodetic distance, during model training and prediction, are given in subsection III-C. In subsection III-D, a theoretical explanation for the impact of outlier homicide events on spatial bandwidth (σ) is provided.

A. Incorporating geodetic distance

Conditional intensity function (λ) and its components- background (μ) and triggering (g) intensity functions assume Gaussian forms (Eqn: 1, 2, & 3). It is well known that about 68% of values drawn from a normal distribution are within one standard deviation (σ) away from the mean; about 95% of the values lie within two standard deviations, and about 99.7% are within three standard deviations (3-sigma rule). Hence background (μ) and triggering intensity (g) reduce to near-zero values for spatial distances larger than the equivalent of 3-4 times spatial bandwidths (η, σ). Additionally, the triggering intensity (g) reduces to near-zero values for the time difference larger than the equivalent of 3-4 times temporal bandwidth (inverse of the decay constant ω). It has been noticed that any error in distance computation in the range $(0 - 4\eta)$ has a significant effect on the computed values of μ, g, and λ.

Treating the geographic coordinate system as planar and scaling it by a factor of $\frac{R\pi}{180}$ results in the naïve Euclidean distance, which always overestimates the geodetic distance. Due to different values of geodetic distance for the same differential along a random latitude and longitude (except at the equator), naïve Euclidean distance brings anisotropy in a geospatial model. An estimation of the model parameters based on Euclidean distance is, therefore, bound to be inaccurate, resulting in a deterioration in its predictive performance [14].



lambda

Fig. 1: Variation in λ with longitudinal difference along a fixed latitude

Fig: 1 shows that for a given set of parameters, values of λ, computed using geodetic distance, in Eqn: 1 are significantly higher than that based on planar Euclidean distance used in [1]. Here, the horizontal axis represents the longitudinal difference along a fixed latitude; the 'red solid' and 'blue dashed' lines correspond to the values of λ for geodetic and planar Euclidean distances, respectively. We have used the average of latitude values in the Chicago crime dataset 2007-09, as fixed latitude here. The parameters- α_1, θ_1, η, σ, and ω, used here, are the average of convergence parameters (Table: (II) & (III)), in the EM-type algorithm.

The predictive accuracy of the model in [1] depends on λ values, particularly 'top k' λ values (k=400 used for comparison in this paper). Hence a higher values of λ have the advantage of correctly contributing to the higher prediction performance as compared to that obtained with the lower values. Inspired by this important observation, we modify the distance formula used in the Gaussians of Eqn: 2 & 3 [1] as given below:

$$d = \frac{R\pi}{180}\sqrt{(\phi^2 + \beta^2\psi^2)} \tag{4}$$

where, R is the radius of Earth; $\phi = \phi_2 - \phi_1$ and $\psi = \psi_2 - \psi_1$ are the latitude and longitude differences, respectively between two data points, used in the distance computation.

If parameters η, σ and ω are expressed in the geographic coordinates as followed in [1], then the factor ($\frac{R\pi}{180}$) in Eqn: (4) gets canceled in the Gaussian. Here, β is a parameter characteristic of the formula used for the geodetic distance. In the case of equirectangular formula, β is $\cos\phi_m$, where $\phi_m = \frac{\phi_1+\phi_2}{2}$ i.e., cosine of the arithmetic mean (AM) of the latitudes of two data points, used in the distance computation [15]. For short distances, a better approximation of geodetic distance is obtained by selecting β to be the geometric mean (GM) of the cosines of the two latitudes, i.e. $\sqrt{\cos\phi_1 \cos\phi_2}$. If complexity of the program is an essential consideration, then considering the cosine of either latitude, i.e, $\cos\phi_1$ or $\cos\phi_2$ as β, for short distances, also gives a reasonable approximation of the geodetic distance. Hence, β can be selected in the proposed *GeoDOME* algorithm, depending upon the data size and the efficiency of the implementation program.

As the homicide prediction looks for 'top- k' λ values, the geodetic distance results in a higher prediction accuracy of homicide as compared to the Euclidean distance used in [1]. The improved prediction of homicides is illustrated in the results section. In subsection III-B, we describe how detection and exclusion of outliers at every iteration of the EM-type algorithm also results in a higher prediction rate.

B. Dynamic outlier exclusion

While training the model using the EM-type algorithm, It is observed that, at the space-time coordinates of certain homicide events, the values of conditional intensity function, i.e., λ are either very small or close to zero. At each iteration of the EM-type algorithm, the majority of such homicide events are the same, indicating that there is very small or no contribution from history crime events to such homicide events, in terms of background and triggering effects. We refer to these homicide events (data points) as dynamic outliers.

To get a perspective on λ values, we computed λ values at the space-time coordinates of all the observed homicide events, due to all the history crime events. While training the model using the Chicago crime dataset 2007-09, the average value of λ, was in the range of 74.7 to 79.7 homicides per deg^2, per day, at each iteration of the EM-type algorithm. During implementation of homicide prediction for the year 2010, the space-time average of λ at the midpoint of a grid cell was found to be 10.51. This gives us an idea that a value of λ close to zero or less than 1 is very small compared to that in the high probability space-time regions of homicide occurrence. Hence, we form the following hypothesis:

H: Homicide events marked with very low λ values are those, which either occur in less crime-prone areas or/and occur during the very early part of the period of the training dataset.

To validate our hypothesis, the outlier homicide events of Chicago crime dataset 2007-09 were separately found using three different criteria as given below:

- ☐ Criterion A: Homicide events for which 'the nearest neighbor history crime event' is at a distance greater than a threshold, d_{th}.
- ☐ Criterion B: Homicide events, corresponding to the 'the lowest k_1 λ values,' out of the λ values computed at the space-time coordinates of all the homicide events in the dataset, at the convergence point of the EM-type algorithm.
- ☐ Criterion C: Homicide events, corresponding to the 'the lowest k_2 triggering intensity (g) values,' out of the 'g' values computed, at the space-time coordinates of all the homicide events in the dataset, at the convergence point of the EM-type algorithm.

In Table: I, the above three types of homicide outlier events are listed for $d_{th} = 0.0030°$ i.e., 333 m, k_1=50, and k_2=50. Here, n(A)=42, n(B)=50, and n(C)=50 denote the number of outlier homicide events, corresponding to criteria A, B and C respectively. To denote an outlier homicide event uniquely in the list A, B, and C of Table: I, we have used the serial number of such homicide events in the chronological list of all the homicide events of Chicago crime dataset 2007-09. A comparison and analysis of these outlier homicide events in Chicago crime dataset 2007-09 led to the following important observations:

Criterion	Serial number of the outlier homicide events in Chicago crime dataset 2007-09
A	n(A)=42 A=[0, 2, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 27, 31, 32, 35, 37, 42, 62, 67, 95, 98, 100, 105, 112, 114, 127, 130, 133, 142, 182, 218, 278, 281, 298, 321, 353, 375, 413, 737, 987, 1231]
B	n(B)=50 B=[0, 1, 2, 4, 7, 8, 9, 10, 14, 27, 32, 33, 34, 35, 37, 42, 95, 98, 100, 105, 109, 112, 114, 127, 133, 142, 182, 218, 221, 278, 281, 298, 321, 350, 353, 375, 392, 466, 622, 679, 737, 807, 894, 916, 987, 1013, 1139, 1163, 1230, 1231]
C	n(C)=50 C=[0, 1, 2, 3, 4, 7, 8, 9, 14, 27, 32, 33, 34, 35, 37, 42, 95, 98, 100, 109, 112, 127, 133, 142, 192, 281, 298, 321, 333, 353, 392, 496, 520, 546, 645, 646, 679, 737, 894, 895, 974, 987, 1013, 1133, 1163, 1231, 1316, 1320, 1344, 1412]

TABLE I: Outlier homicide events in Chicago crime dataset 2007-09

1) It is observed that almost the majority of the outlier homicides events happen to be among the initial 10% of homicide events in the dataset, which are sparse in spatial dimensions. For instance, as given in Table: I, 30 out of 42 outlier homicide events (i.e., 71%) of criterion A are among the initial 10% of homicide events in the dataset. The same pattern is observed for criteria B and C also, with 52% and 48% of outlier homicide events, respectively, among the initial 10% of homicide events in the dataset
2) Homicide events, with very small λ values at their space-time coordinates (criterion B), belong to the low probability space-time region of homicide occurrence. Such data points have very few background and recent crime events in the vicinity and thus tend to bias the model away from the ground truth. Exclusion of such outliers is a standard statistical approach for reducing the bias and improving the model's prediction rate.
3) Homicide events, having very small triggering intensity (g) but high background intensity (μ) in some cases, i.e., criterion C, represent old crime hot spots. These hotspots are becoming too feeble (in terms of recent criminal activity) to trigger new homicide events. Exclusion of such outliers reduces the bias of the model and takes it towards higher predictability, which is validated experimentally.

Above observations form the basis of considering a combination of three outlier detecting criteria in the proposed *GeoDOME* algorithm, where parameters d_{th}, k_1 and, k_2 are used to tune the model for improved performance. The combined criterion is referred to as $A \cup B \cup C$.

It has been observed by Archambeau [16] that convergence of the EM algorithm is sensitive to outliers. Yang [17] reported a significant reduction in the bias after removing the effects of outliers in the EM algorithm. Hence, exclusion of homicide outliers, at each iteration of the EM-type algorithm of the M-SEPP based ETAS model, nudges its convergence point towards optimal values in parameter space, leading to a better trained model and higher predictabilty of homicides. This has been explained theoretically in subsection III-D and also validated experimentally in the section IV. In the next subsection, we describe the *GeoDOME* algorithm

C. *GeoDOME algorithm*

This subsection presents the *GeoDOME* algorithm (Algorithm: 1) which incorporates geodetic distance and dynamic outlier exclusion in the expectation-maximization (EM)-type algorithm, followed in [1].

Using geodetic distance (Eqn: (4)), the background (μ) and triggering (g) intensity functions (Eqn: (2) , (3)), as used in [1], are modified as below:

Background intensity:

$$\mu_{i,j} = \frac{\alpha_{m_i}}{2\pi\eta^2 t_j} \mathrm{e}^{-\frac{(x_j - x_i)^2 + \boldsymbol{\beta^2}(y_j - y_i)^2}{2\eta^2}} \tag{5}$$

Triggering intensity:

$$g_{i,j} = \frac{\theta_{m_i}\omega \mathrm{e}^{-\omega(t_j - t_i)}}{2\pi\sigma^2} \mathrm{e}^{-\frac{(x_j - x_i)^2 + \boldsymbol{\beta^2}(y_j - y_i)^2}{2\sigma^2}} \tag{6}$$

where, β is the geometric mean (GM) of the cosines of two latitudes i.e., $\sqrt{\cos \mathrm{x}_i \cos \mathrm{x}_j}$, as given in section III-A. x_i and x_j are the latitudes of i^{th} history crime and j^{th} homicide event, respectively.

In the E-step of the EM-type algorithm, the background & triggering probabilities (p^b_{ij}, p^t_{ij}), of influencing j^{th} homicide event by the i^{th} history crime, are computed, using an initial guess for the model parameters- η, σ, ω, α_m, and θ_m, as given below:

Background probability:

$$p^b_{ij} = \frac{\mu_{i,j}}{\sum_{i=1;\ t_i<t_j}^{K} \lambda_{i,j}} \tag{7}$$

Triggering probability:

$$p^t_{ij} = \frac{g_{i,j}}{\sum_{i=1;\ t_i<t_j}^{K} \lambda_{i,j}} \tag{8}$$

where, K= number of crimes of all types (i.e., marks), in the training dataset.

The M-step of the EM-type algorithm is modified as below: To estimate the parameters η, σ, ω, α_m and, θ_m from the training dataset, the probability values obtained from Eqn: (7) & (8) are used in the following modified equations:

Background spatial bandwidth (η):

$$\eta^2 = \frac{\sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^b_{ij} \left((x_j - x_i)^2 + \boldsymbol{\beta^2} (y_j - y_i)^2 \right)}{2 \sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^b_{ij}} \tag{9}$$

Triggering spatial bandwidth (σ):

$$\sigma^2 = \frac{\sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^t_{ij} \left((x_j - x_i)^2 + \boldsymbol{\beta^2} (y_j - y_i)^2 \right)}{2 \sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^t_{ij}} \tag{10}$$

Temporal decay constant:

$$\omega = \frac{\sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^t_{ij}}{\sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^t_{ij} (t_j - t_i)} \tag{11}$$

Background branching factor:

$$\alpha_m = \frac{\sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^b_{ij} i(m_i = m)}{K_m} \tag{12}$$

Triggering branching factor:

$$\theta_m = \frac{\sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^t_{ij} i(m_i = m)}{K_m} \tag{13}$$

where, Ω is the set of outlier homicides as described in subsection III-B. The procedure for creating the outlier set (Ω) is provided in *GeoDOME* algorithm (Algorithm: 1). K_m= number of crimes of mark m, and K_1= number of homicides (m=1), in the training dataset.

We define below the average triggering factor (τ), which is the average of triggering effects of history crimes to cause a homicide event.

Average triggering factor:

$$\tau = \frac{\sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p^t_{ij}}{\sum_{i=1}^{K} \sum_{\substack{j;\ t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} \left(p^b_{ij} + p^t_{ij} \right)} \tag{14}$$

It is experimentally observed that if the value of τ doesn't change by more than 0.1e-3 in two consecutive iterations of the *GeoDOME* algorithm (Algorithm: 1), other parameters- η, σ, ω, α_m, and θ_m, also don't. Hence, τ is used for the stopping criteria (Algorithm: 1).

Algorithm 1 *GeoDOME* algorithm

1: Compute geodetic distance matrix (D), with an element (i, j) representing the distance between the i^{th} history crime and the j^{th} homicide event, in the training dataset.
2: Compute, similarly, the time difference matrix (T), with element (i, j) representing the time difference between the i^{th} history crime and the j^{th} homicide event ($t_j > t_i$).
3: Detect and generate set A of outlier homicides (criterion-A) using a choice of threshold, $d_{th} \in [0.0024^o - 0.0040^o]$.
4: Guess initial parameters.
   1) Choose η, σ and ω using the domain knowledge of criminology. e.g. $\eta, \sigma \in [0.0009^o - 0.0014^o]$; $\frac{1}{\omega} \in [30 - 50]$ days.
   2) Choose extreme values (very high or very low) for hidden variables- α_m,θ_m, to allow a wide range of possible values in successive iterations.
5: $n \leftarrow 1$
6: DO
7: Compute μ, g, and λ using Eqn-(1, 5, 6), matrices: D & T (step-2 & 3), and the last computed values / initial values (for n=1) of parameters- η, σ, ω, α_m, and θ_m.
8: $G \leftarrow Sorted(g); L \leftarrow Sorted(\lambda)$
9: Compute set B and C of outlier homicides, using criterion-B & C. Choose $k_1 \in [50 - 100]$, $k_2 \in [40 - 70]$).
10: $\Omega \leftarrow A \cup B \cup C$
11: Compute background & triggering probabilities (p^b_{ij}, p^t_{ij}), using Eqn: (7, 8).
12: Compute parameters $(\alpha_m, \theta_m, \omega, \eta, \sigma)$, using Eqn: (9, 10, 11, 12, & 13).
13: Compute 'average triggering factor (τ),' using Eqn: (14).
14: $n \leftarrow n + 1$
15: WHILE $\tau > 0.1e - 3$
16: Output model parameters $(\alpha_m, \theta_m, \omega, \eta, \sigma)$ at convergence.

D. Impact of outlier homicides on the computation of spatial bandwidth (σ)

For convergence of the EM-type algorithm, the estimates of background and triggering spatial bandwidth η, and σ are

Algorithm 2 Prediction algorithm

1: Divide the city area into rectangular grid cells of equal size (e.g., 150m X 150 m), indexed by $j \in [1-N]$, where N= total number of grid cells in the city.
2: Compute homicide frequency in each grid cell, for each day of the test dataset.
3: Compute geodetic distance matrix (D), with an element (i, j) representing the distance between the i^{th} history crime in the training dataset and the center of j^{th} grid cell.
4: Compute λ_1 at the center of each grid cell, using D.
5: C $\leftarrow$ 0 (prediction count of homicides).
6: n $\leftarrow$ 1 (1st day of the test dataset).
7: REPEAT
8: Compute dynamic geodetic distance matrix (D_n) with an element (k, j). representing the distance between the k^{th} history crime of the test dataset (until the n^{th} day) and the center of the j^{th} grid cell.
9: Compute λ_2 at the center of each grid cell, using D_n.
10: $\lambda \leftarrow \lambda_1 + \lambda_2$.
11: Sort λ in descending order.
12: $C_n \leftarrow$ aggregate of homicide frequency in each grid cell, marked by 'top k' λ values.
13: $C \leftarrow C + C_n$
14: $n \leftarrow n + 1$
15: UNTIL n= last day of test dataset.
16: Output C as total homicide prediction in test dataset.

forced to be equal in [1]. Therefore, Eqn: (9) and (10) of *GeoDOME* algorithm reduce to:

$$\eta^2 = \sigma^2 = \frac{\sum_{i=1}^{K} \sum_{\substack{j;\, t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p_{ij} \left((x_j - x_i)^2 + \boldsymbol{\beta^2} (y_j - y_i)^2 \right)}{2 \sum_{i=1}^{K} \sum_{\substack{j;\, t_i<t_j \\ \boldsymbol{j \notin \Omega}}}^{K_1} p_{ij}} \tag{15}$$

where, p_{ij}, is the probability of influencing j^{th} homicide event, by the i^{th} history crime. For the case, when outlier homicide events are not excluded from the computation in the EM-type algorithm, the spatial bandwidth (σ_o) can be expressed as:

$$\sigma_o^2 = \frac{2(N-n)\sigma^2 + \sum_{i=1}^{K} \sum_{\substack{j;\, t_i<t_j \\ \boldsymbol{j \in \Omega}}}^{K_1} p_{ij} \left((x_j - x_i)^2 + \boldsymbol{\beta^2} (y_j - y_i)^2 \right)}{2N} \tag{16}$$

where, N= total number of homicide events, and n= number of outlier homicide events, in the training dataset. If distance threshold for outlier set, $\Omega = A$ for criterion A is $d_{th} = k\sigma$, where $k = 2.5$, then Eqn: (16) becomes:

$$\sigma_o^2 > \frac{2(N-n)\sigma^2 + n(k\sigma)^2}{2N} \tag{17}$$

since $\Sigma_i p_{ij} = 1,\ \forall j$.

The lower bound for the increase in spatial bandwidth ($\Delta\sigma = \sigma_o - \sigma$)- due to the non-exclusion of homicide outlier events in the computation of the EM-type algorithm, is obtained as:

$$\Delta\sigma > \frac{(k^2-2)n}{4N}\sigma \tag{18}$$

This shows that the lower bound of the increase in spatial bandwidth ($\Delta\sigma$) is proportional to the number of homicide outliers (n) in the training dataset. In the implementation of *GeoDOME* algorithm (Algorithm: 1), using the combined criterion $A \cup B \cup C$, the value of $\Delta\sigma$ is obtained as $(1.25)\sigma$, which is significant. It proves that if outlier homicide events are not excluded from the computation in the EM-type algorithm, the spatial bandwidth is overestimated. It was also observed that a higher value of spatial bandwidth leads a different convergence point in the EM-type algorithm, which results in lower prediction accuracy of the homicide events. From the criminological perspective also, homicide outliers- being poor indicators of likely occurrence of homicides in the vicinity, tend to bias the model towards lower predictability of homicides. Chainey [18] reports that in the KDE hotspot model for the prediction of crime, the prediction accuracy index (PAI) decreases with increasing spatial bandwidth. For an increase in spatial bandwidth from 100 to 300 m, the prediction accuracy index (PAI) score for 'assault with injury' decreases drastically from 142.8 to 79.4. Hence, σ instead of σ_o is used in the proposed *GeoDOME* algorithm (Algorithm: 1).

In the next section, we describe how the proposed *GeoDOME* algorithm (Algorithm: 1) improve the homicide prediction rate.

IV. RESULTS

This section presents the implementation of the proposed *GeoDOME* algorithm (Algorithm: 1), using the same Chicago crime dataset of year 2007-12 as used in [1]. The dataset is preprocessed to create three training datasets: 2007-2009, 2008-2010 & 2009-2011 and corresponding test datasets: 2010, 2011 & 2012.

The proposed *GeoDOME* algorithm (Algorithm: 1) was applied on all the three training datasets. The convergence of the EM-type algorithm is slow, if the initial choice of parameters is very far from the ground truth [19]. Hence, a reasonable initial guess for the parameters- σ, η, and ω was made using the criminological theory of diminishing influence of crime, beyond 500 m in space and 100 days in time dimensions [20]. Furthermore, as the hidden parameters- α_m, and θ_m are unobservable beforehand, so their extreme values (very high or very low) were taken as the initial guess, to allow a wide range of their values in successive iterations of the EM-type algorithm [19]. On convergence of the proposed *GeoDOME* algorithm (Algorithm: 1) applied on each training dataset, the model parameters- σ, η, ω, α_m, and θ_m were obtained. The values of the background and triggering spatial bandwidths ($\sigma = \eta$) and the decay constant (ω) are given in Table: II. The values of background and triggering branching factors for each of the six crime types (α_1, to α_6 and β_1 to β_6) are given in Table: II & III, respectively.

Training Set	σ	ω	α_1	α_2	α_3	α_4	α_5	α_6
2007-09	0.0012	0.029	0.0294	0.0040	0.0402	0.0326	0.0337	0.0761
2008-10	0.0011	0.038	0.0241	0.0150	0.0333	0.0279	0.0349	0.0000
2009-11	0.0009	0.039	0.0535	0.0173	0.0205	0.0400	0.0306	0.0000

TABLE II: Model parameters $(\sigma, \omega, \alpha_m)$

Training Set	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
2007-09	0.0355	0.0061	0.0050	0.0122	0.0281	0.0000
2008-10	0.0302	0.0025	0.0022	0.0095	0.0253	0.0000
2009-11	0.0158	0.0008	0.0061	0.0023	0.0227	0.0000

TABLE III: Model parameters (θ_m)

For prediction, the city of Chicago was divided into grid cells of size 150m X 150 m. The convergence parameters of each training set was used in prediction algorithm (Algorithm: 2) to predict homicide events in the corresponding test dataset. For each day of the prediction period in test dataset, conditional intensity function (λ) was computed at the center of each grid cell, using up to date history data and that was sorted in descending order. Among these grid cells, four hundred grid cells having the highest λ values were selected. The total number of homicides, that occurred in the selected grid cells, gives the prediction count of homicides for each day of the prediction period.

Mohler [1] quotes the baseline prediction of 153 homicides (11% of the total during 2010-12) in 400 grid cells flagged per day, using the Chronic hotspot approach and improves it to 180 (12.9% of the total), an increase of 17.6% over the baseline. Using novel methods of geodetic distance and dynamic outlier exclusion, the *GeoDOME* algorithm (Algorithm: 1) further improves the homicide prediction to 210 (15.1% of the total), with an increase of 37.3% over the same baseline and 16.7% over the results of [1]. The comparative performance of three prediction methods is given in Table IV.

Prediction method	Number of homicides predicted	Prediction Accuracy (%)	Improvement over the baseline (%)
Chronic Hotspot[a]	153	11.0	
[Mohler, 2014] [1]	180	12.9	17.6
GeoDOME	210	15.1	37.3

[a]baseline.

TABLE IV: Comparative results of *GeoDOME* and [Mohler, 2014] [1] with respect to same baseline

Fig 2 shows homicide prediction percentage plotted against the number of grid cells flagged (f) each day for each of the test set 2010, 2011, and 2012, as represented by the solid blue, dashed green, and dotted red lines, respectively.

To find an efficent and effective range of grid cells to be flagged by the police departments, we define the 'marginal rate of homicide prediction' (ξ), as the ratio of homicide prediction percentage to the number of grid cells flagged (f). ξ can be obtained from the predictive accuracy index (PAI) [18], by multiplying the latter by $\frac{100}{F}$, where F= total number of grid cells, covering the entire city area. Fig 3 shows the variation of ξ with respect to f, for each of the test datasets- 2010, 2011, and 2012, represented by the solid blue, dashed green, and dotted red lines, respectively. It shows that for f in the range of [1-1000] grid cells, there is a comparatively higher value of ξ and nearly constant ξ beyond this range.

The law of crime concentration states that 50 % of crimes occur in 5% of the city area [21]. An optimal number of grid cells for predictive police deployment in a city, comparable to Chicago, in area, lies in [1-1000] range. However, it requires further validation, using crime data of other cities. The selection of f within [1-1000] would depend upon the scale of patrol resources available with the police departments, on day to day basis. It is a tradeoff between competing demands of resources for multifarious police functions and the imperative of predictive deployment for targetted crime reduction.

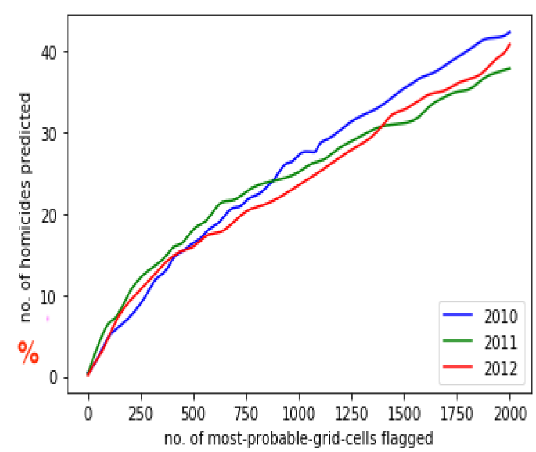


Fig. 2: Percentage homicide prediction

V. Conclusions

In this paper, a modified EM-type algorithm- *GeoDOME*, incorporating dynamic outlier exclusion and geodetic distance has been proposed for improved homicide prediction. The theoretical basis of the model has been provided. It has been experimentally demonstrated that the *GeoDOME* algorithm leads to much higher prediction of homicides as compared to the state-of-the-art results [1] in Chicago crime dataset.

A higher homicide prediction rate of *GeoDOME* algorithm makes it suitable for its widespread use in predictive policing.

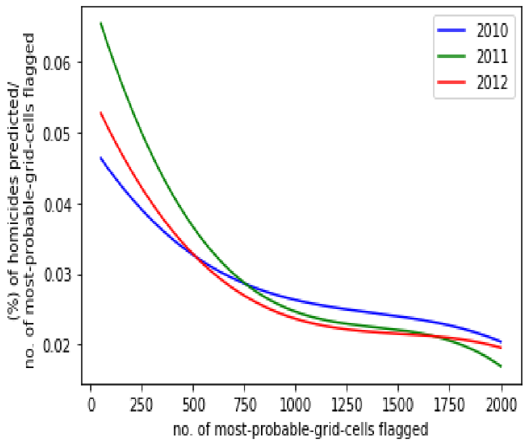


Fig. 3: Marginal rate of homicide prediction

Another offshoot of this research is the use of dynamic conditional intensity values for city wide daily update of crime prone pockets, which can be used by citizens and city planners both.

References

[1] G. Mohler, "Marked point process hotspot maps for homicide and gun crime prediction in Chicago," *International Journal of Forecasting*, vol. 30, no. 3, pp. 491–497, 2014.

[2] B. Chandra and M. Gupta, "Novel multivariate time series clustering approach for e-governance of crime data," in *Sixth International Conference on Developments in eSystems Engineering*. IEEE, 2013, pp. 311–316.

[3] M. Gupta, B. Chandra, and M. Gupta, "A framework of intelligent decision support system for Indian police," *Journal of Enterprise Information Management*, vol. 27, no. 5, p. 512, 2014.

[4] H.-W. Kang and H.-B. Kang, "Prediction of crime occurrence from multi-modal data using deep learning," *PloS one*, vol. 12, no. 4, 2017.

[5] S. K. Dash, I. Safro, and R. S. Srinivasamurthy, "Spatio-temporal prediction of crimes using network analytic approach," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1912–1917.

[6] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, 2011.

[7] G. O. Mohler, M. B. Short, S. Malinowski, M. Johnson, G. E. Tita, A. L. Bertozzi, and P. J. Brantingham, "Randomized controlled field trials of predictive policing," *Journal of the American statistical association*, vol. 110, no. 512, pp. 1399–1411, 2015.

[8] M. Kajita and S. Kajita, "Crime prediction by data-driven Green's function method," *International Journal of Forecasting*, vol. 36, no. 2, pp. 480 – 488, 2020.

[9] A. Reinhart and J. Greenhouse, "Self-exciting point processes with spatial covariates: modelling the dynamics of crime," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 67, no. 5, pp. 1305–1329, 2018.

[10] J. Zhuang and J. Mateu, "A semiparametric spatiotemporal Hawkes-type point process model with periodic background for crime data," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 182, no. 3, pp. 919–942, 2019.

[11] D. J. Daley and D. V. Jones, *An introduction to the theory of point processes. Vol. I Elementary theory and methods. 2nd ed.* Springer, 2003.

[12] Y. Ogata, "Space-time point-process models for earthquake occurrences," *Annals of the Institute of Statistical Mathematics*, vol. 50, no. 2, pp. 379–402, 1998.

[13] A. Veen and F. P. Schoenberg, "Estimation of space–time branching process models in seismology using an EM-type algorithm," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 614–624, 2008.

[14] S. Banerjee, "On geodetic distance computations in spatial modeling," *Biometrics*, vol. 61, no. 2, pp. 617–625, 2005.

[15] Ö. G. Esenbuğa, A. Akoğuz, E. Çolak, B. Varol, and B. Erol, "Comparison of principal geodetic distance calculation methods for automated province assignment in Turkey," *International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management*, vol. 2, pp. 141–148, 2016.

[16] C. Archambeau, J. A. Lee, and M. Verleysen, "On convergence problems of the EM algorithm for finite Gaussian mixtures." in *11th European Symposium on Artificial Neural Networks*, vol. 3, 2003, pp. 99–106.

[17] Z. R. Yang and S. Chen, "Robust maximum likelihood training of heteroscedastic probabilistic neural networks," *Neural Networks*, vol. 11, no. 4, pp. 739–747, 1998.

[18] S. P. Chainey, "Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime," *Bulletin of the Geographical Society of Liege*, vol. 60, pp. 7–19, 2013. [Online]. Available: https://popups.uliege.be:443/0770-7576/index.php?id=422

[19] G. McLachlan, T. Krishnan, and S. Ng, "The EM algorithm," *Papers/Humboldt-Universtat. Berlin: Center for Applied Statistics and Economics*, 2004.

[20] G. Rosser, T. Davies, K. J. Bowers, S. D. Johnson, and T. Cheng, "Predictive crime mapping: Arbitrary grids or street networks?" *Journal of Quantitative Criminology*, vol. 33, no. 3, pp. 569–594, 2017.

[21] D. Weisburd, "The law of crime concentration and the criminology of place," *Criminology*, vol. 53, no. 2, pp. 133–157, 2015.