GEOGRAPHIC PROFILING THROUGH SIX-DIMENSIONAL NONPARAMETRIC DENSITY ESTIMATION

AUSTIN C. ALLEMAN*

ADVISOR: GEORGE O. MOHLER

Abstract. Geographic profiling is the problem of identifying the location of the offender anchor point (offender residence, place of work, etc.) of a linked crime series using the spatial coordinates of the crimes or other information. A standard approach to the problem is 2D kernel density estimation, which relies on the assumption that the anchor point is located in close proximity to the crimes. Recently introduced Bayesian methods allow for a wider range of criminal behaviors, as well as the incorporation of geographic and demographic information. The complexity of these methods, however, make them computationally expensive when implemented. We have developed a nonparametric method for geographic profiling that allows for more complex criminal behaviors than 2D kernel density estimation, but is fast and easy to implement. For this purpose, crime locations and anchor point are considered as one data point in the space of all crime series. Dimension reduction is then used to construct a 6D probability density estimate of offender behavior using historical solved crime series data, from which an anchor point density corresponding to an unsolved series can be computed. We discuss the advantages and disadvantages of the method, as well as possible real-world implementation.

1. Introduction. Methodologies for geographic profiling have changed vastly since the first recognition of mathematical relationships attributed to criminal behavior [1, 2]. The basic aim is the estimation of an anchor point, defined as a place of frequent visitation by a serial offender, which may be a home, a place of employment, the residence of a relation or close friend, etc. Consider the spatial coordinates of the crime series, characterized by distinct events $\mathbf{x}_1$, $\mathbf{x}_2$,..., $\mathbf{x}_n$ and anchor point $\mathbf{z}$, each with x and y coordinates on a grid. One simple heuristic is to calculate the arithmetic mean of the crime series $(\overline{x}, \overline{y})$ as an initial guess for the anchor point. Another possibility is to choose the point which minimizes the sum of the distances to crimes of the series, called the Fermat-Weber point. The first known instance of geographic profiling, according to criminologist David Canter, successfully made use of the former technique as means for locating the hometown of the Yorkshire Ripper in 1980 [3].

In the early 1980s, Brantingham and Brantingham made a substantial contribution to geographic profiling efforts with the observation of distance decay behavior [1, 2] amongst serial criminals. This behavior forms the basis for Kim Rossmo's Criminal Geographic Targeting (CGT) algorithm, which makes use of a scoring function

$$S(\mathbf{y}) = \sum_{i=1}^{n} f(d(\mathbf{x}_i, \mathbf{y})), \tag{1.1}$$

to target high-priority search areas, where $\mathbf{y}$ is a location on the search grid, $d(\mathbf{x}_i, \mathbf{y})$ is the distance from $\mathbf{y}$ to crimes of the series, and f is a decay function simulating the behavior Brantingham and Brantingham observed. To obtain the hit score, the scoring function assigns to each crime location a copy of the function $f(d(\mathbf{x}_i, \mathbf{y}))$ and sums the value of each at every location $\mathbf{y}$ on the search grid. CGT is famously responsible for the successful geographic profile of a serial rapist in Lafayette, LA, which pinpointed the offender's home to within 0.2 miles.

---
*DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, SANTA CLARA UNIVERSITY, SANTA CLARA, CA 95053

Recently, however, new approaches to the problem of locating the anchor have been proposed. The National Institute of Justice issued a statement in 2005 citing geographic profiling as an "anecdotal success", though unsuitable as a model for crime [7]. This assessment has been received largely as a challenge within the crime science community. O'Leary proposes an approach based on Bayesian analysis [8] (see Section 2.2); Mohler and Short recently developed a method that employs models of foraging behavior [6] (see Section 2.3). While these methods have shown promise in terms of their accuracy, they are difficult to implement, requiring sophisticated numerical techniques, and are computationally intensive.

In this paper we introduce a new method for geographic profiling where a particular anchor point and crime series, $\mathbf{u} = (\mathbf{z}, n, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$, is treated as one observed data point in the space of all crime series. We then estimate the joint probability density $P(\mathbf{u})$ using historical solved crime series, dimension reduction, and six-dimensional kernel density estimation. The method is able to detect patterns of movement and identifies attractive destinations for serial criminals directly from the data, rather than making parametric assumptions on offender behavior a priori. The outline of this paper is as follows. In Section 2, we review the literature on geographic profiling methods. In Section 3, we develop our nonparametric model for offender behavior and show how it can be used for geographic profiling. In Section 4, we apply our methodology to historical solved crime series data on residential burglaries in Los Angeles. We compare the effectiveness of the method to the widely used CGT algorithm.

2. Existing geographic profiling methods.

2.1. CGT Algorithm. Given a crime series containing n linked events at locations $\mathbf{x}_1 = (x_1, y_1)$, $\mathbf{x}_2 = (x_2, y_2)$,..., $\mathbf{x}_n = (x_n, y_n)$ over a discretized search area, the CGT algorithm begins by first determining the mean nearest neighbor distance $\bar{d}$. The two most common choices for geographic profiling are Euclidean,

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2},$$

and Manhattan (or street) distance,

$$d(\mathbf{x}_i, \mathbf{x}_j) = |x_i - x_j| + |y_i - y_j|.$$

CGT employs the latter and measures the distance from each crime instance $\mathbf{x}_i$ to the remaining $n - 1$ crimes of the series and records the distance to the nearest neighbor. The mean of these nearest neighbor distances,

$$\bar{d} = \sum_{i=1}^{n} \frac{\min_{j \neq i}(d(\mathbf{x}_i, \mathbf{x}_j))}{|n|}, \tag{2.1}$$

which will be used to determine the *buffer zone radius* in Eq. 2.2. CGT next determines the hit score for points in the search area using the scoring function [13],

$$S(x, y) = k \sum_{i=1}^{n} \left(\frac{\phi}{(|x - x_i| + |y - y_i|)^h} + \frac{(1 - \phi)B^{g-h}}{(2B - |x - x_i| + |y - y_i|)^g} \right) \tag{2.2}$$

where k is a scaling constant; B is the buffer zone radius, equal to half the mean nearest neighbor distance; $h = g = 1.2$ are constants selected from a gravity model

formulation developed to describe migration patterns of criminal fugitives [12, 11]; and ϕ is a value determined by the piecewise function,

$$\phi = \begin{cases} 1, & \text{if } |x - x_i| + |y - y_i| \geq B \\ 0, & \text{if } |x - x_i| + |y - y_i| < B \end{cases} \tag{2.3}$$

The result, as mentioned earlier, is a concentration of high hit scores near the crimes of the series. Based on chosen distance metrics and the number of offenses, the hit score map will vary, but there is a well-defined search area, with peaks around the buffer zones surrounding each crime. In Figure 2.1, we provide an example of the output of the CGT algorithm for a burglary series in Los Angeles, CA.
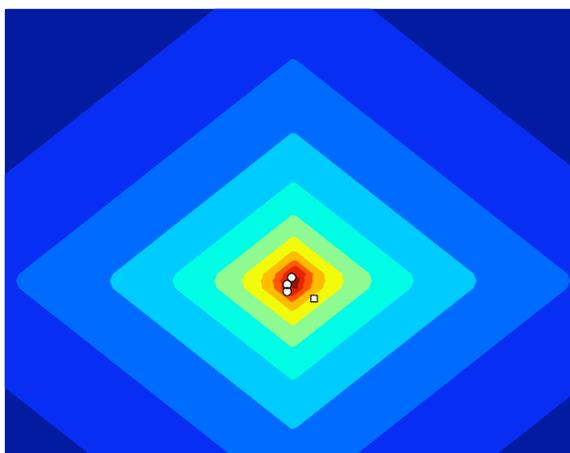

Fig. 2.1: CGT output for a crime series in Los Angeles, CA, logarithmic scale (white square indicates anchor point, white circles indicate crime locations).

Alternative distance decay functions

Canter and Hammond detail a comparison of the efficacy of various decay functions including logarithmic, exponential, quadratic, and linear functions in modeling raw data on serial killers [4] and note that most functions will work without significant variation in the accuracy of results. General forms for some decay functions include:

- Linear: $f(d) = A + Bd$;
- Quadratic: $f(d) = Ad^2 + Bd + C$;
- Negative Exponential: $f(d) = Ae^{-Bd}$;
- Logarithmic: $f(d) = A + B \times ln(d)$;
- Truncated Negative Exponential: $\begin{cases} Bd, & \text{if } d < C; \\ Ae^{-\beta d}, & \text{if } d \geq C; \end{cases}$

Once f has been chosen, hit scores are then calculated according to (1) for all $\mathbf{y}$ in the domain.

Geographic profiling methods making use of the scoring function (1) have been the target of criticism due to their non-probabilistic nature. While a distance decay function may accurately model criminal target selection, summing the distance functions does not necessarily provide an accurate estimate of the anchor point. In fact, from the perspective of kernel density estimation, a model of the form (1) actually estimates the location of the next crime.

2.2. O'Leary's Method. In response to the recent criticisms of geographic profiling methodology [10, 14, 5], O'Leary recently developed an innovative method based on a Bayesian analysis of crime site locations [8]. The method assumes a probabilistic model of offense location $P(\mathbf{x}|\mathbf{z}, \alpha)$ conditioned on the location of the anchor point and the average offense distance α. The probability density of the location of the anchor point $\mathbf{z}$ and average offense distance α given crime locations $\mathbf{x}_1, ..., \mathbf{x}_n$ is given as,

$$P(\mathbf{z}, \alpha|\mathbf{x}_1, ..., \mathbf{x}_n) = \frac{P(\mathbf{x}_1, ..., \mathbf{x}_n|\mathbf{z}, \alpha)\pi(\mathbf{z}, \alpha)}{P(\mathbf{x}_1, ..., \mathbf{x}_n)}, \tag{2.4}$$

where $P(\mathbf{x}_1, ..., \mathbf{x}_n)$ is the marginal distribution and $\pi(\mathbf{z}, \alpha)$ is the prior distribution, a representation of knowledge about the probability an offender has a particular anchor point and average offense distance before crime series data is included. The simplest case is that $\mathbf{z}$ and α are independent and the prior distribution can then be simplified to

$$\pi(\mathbf{z}, \alpha) = H(\mathbf{z})\pi(\alpha). \tag{2.5}$$

Assuming event independence, the joint probability that crimes were committed at $\mathbf{x}_1, ..., \mathbf{x}_n$ given $\mathbf{z}$ and α can be rewritten as,

$$P(\mathbf{x}_1, ..., \mathbf{x}_n|\mathbf{z}, \alpha) = \prod_{i=1}^{n} P(\mathbf{x}_i|\mathbf{z}, \alpha). \tag{2.6}$$

Then the posterior distribution becomes

$$P(\mathbf{z}, \alpha|\mathbf{x}_1, ..., \mathbf{x}_n) \quad \propto \quad P(\mathbf{x}_1|\mathbf{z}, \alpha)...P(\mathbf{x}_n|\mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha).$$

It now becomes necessary to isolate a probability density for $\mathbf{z}$. This is done by integrating the conditional distribution, which gives,

$$P(\mathbf{z}|\mathbf{x}_1, ..., \mathbf{x}_n) \quad \propto \int P(\mathbf{x}_1|\mathbf{z}, \alpha)...P(\mathbf{x}_n|\mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha)d\alpha. \tag{2.7}$$

All that is left to determine is prior distribution information $H(\mathbf{z})$, $\pi(\alpha)$, and an appropriate behavior model $P(\mathbf{x}|\mathbf{z}, \alpha)$. O'Leary discusses the use of the distance decay function $f(d(\mathbf{x}, \mathbf{z}))$ incorporating target attractiveness $G(\mathbf{x})$, and a normalization factor

$$N(\mathbf{z}) = \frac{1}{\int \int f(d(\mathbf{x}), \alpha)G(\mathbf{x})dx^{(1)}dx^{(2)}}, \tag{2.8}$$

to achieve a general form for a behavior model,

$$P(\mathbf{x}|\mathbf{z}, \alpha) = f(d(\mathbf{x}), \mathbf{z})G(\mathbf{x})N(\mathbf{z}), \tag{2.9}$$

where $f(d(\mathbf{x}, \mathbf{z}))$ is a Rayleigh distribution. As for the prior distribution, O'Leary recommends population density modeled using kernel density estimation on a block-by-block basis for $H(\mathbf{z})$ and direct historical crime data for $\pi(\alpha)$ [8].

O'Leary has packaged and released software which makes use of his method, which is available for download on his website. The main issue is one of implementation (and not mathematical underpinning); the development of a geographic profile using the present software requires significant computational resources and time. Additionally, O'Leary suggests in his report to the NIJ that the offender behavior model $P(\mathbf{x}|\mathbf{z}, \alpha)$ is simple, and that there is possible improvement to be made [9].

2.3. Mohler and Short's Method. Mohler and Short recently devised a method for geographic profiling [6], based upon a foraging model for offender activity, that allows for the incorporation of geographic features into criminal target selection. The model is characterized by a criminal's random walk from the anchor point $\mathbf{z}$, culminating in an offense committed at location $\mathbf{x}$ and time t. Assuming the offender's position is given by $\mathbf{y}(t)$ at time t, his or her movement is determined by a realization of the stochastic differential equation,

$$\frac{d\mathbf{y}}{dt} = \mu(\mathbf{y}) + \sqrt{2D}\mathbf{R}_t \tag{2.10}$$

where D is the diffusion parameter, $\mathbf{R}_t \in \mathbb{R}^2$ is white noise and $\mu(\mathbf{y})$ is a drift term, used to influence the direction of the walk, possibly based on some environmental factor. Mohler and Short suggest the use of a spatial target attractiveness field $A(\mathbf{y}|\mathbf{z})$ to determine the termination of the walk.

The transition probability density $\rho(\mathbf{x}, t|\mathbf{z})$ of the location of the criminal at time t then solves the Fokker-Planck equation,

$$\frac{d\rho}{dt} = \nabla_{\mathbf{x}} \cdot (D\nabla_{\mathbf{x}}\rho) - \nabla_{\mathbf{x}} \cdot ((\mu(\mathbf{x}))\rho) - A(\mathbf{x}|\mathbf{z})\rho, \tag{2.11}$$

with initial condition given by,

$$\rho_0 = \delta(\mathbf{x} - \mathbf{z}). \tag{2.12}$$

Integration with respect to time yields the target selection probability density,

$$P(\mathbf{x}|\mathbf{z}) = A(\mathbf{x}|\mathbf{z})\rho(\mathbf{x}|\mathbf{z}),$$

where $\rho(\mathbf{x}|\mathbf{z})$ solves the differential equation,

$$-\nabla_{\mathbf{x}} \cdot (D\nabla_{\mathbf{x}}\rho) + \nabla_{\mathbf{x}} \cdot ((\mu(\mathbf{x}))\rho) + A(\mathbf{x}|\mathbf{z})\rho = \delta(\mathbf{x} - \mathbf{z}). \tag{2.13}$$

Mohler and Short detail several particular solutions [6], including cases with no drift, with constant drift, with attractiveness field $A(\mathbf{x}|\mathbf{z})$ taken to be housing density, and with both constant drift and an attractiveness field. The geographic profile is then found in a similar fashion to O'Leary's method (2.7),

$$P(\mathbf{z}|\mathbf{x}_1, ..., \mathbf{x}_n) \quad \propto \int P(\mathbf{x}_1|\mathbf{z}, \alpha)...P(\mathbf{x}_n|\mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha)d\alpha, \tag{2.14}$$

where the conditional densities $P(\mathbf{x}_i|\mathbf{z}, \alpha)$ are computed by using either the adjoint equation to (2.13) or a parametric approximation to the solution. Here housing density and historic crime series data are used to estimate the prior densities. As with O'Leary's approach, Mohler and Short's method is associated with sizeable computational costs.

3. 6D kernel density estimation and geographic profiling. We begin by considering the problem of locating the anchor point as a matter of historical significance. Assuming offenders in a given city share common traits in target selection, it should be the case that certain trends such as travel patterns, local geography, and target attractiveness are represented in historical case data. We note that if past and future offenders do not share common traits, geographic profiling is a futile effort.

Let $\mathbf{u}$ be a variable representing a crime series and anchor point,

$$\mathbf{u} = (\mathbf{z}, n, \mathbf{x}_1, ..., \mathbf{x}_n), \tag{3.1}$$

where $\mathbf{z}$ is the location of the anchor point and n is the number of crimes of the series with locations $\mathbf{x}_1, ..., \mathbf{x}_n$. We begin by modeling the joint probability density $P(\mathbf{u})$. This probability density, assumed to exist, can be viewed as a density over the space of observable criminal behaviors. In practice $\mathbf{u}$ will be of finite but variant dimension, so we'll first reduce the dimension to six in order to construct a kernel density estimate $\hat{P}$. This is done by introducing the vector

$$\mathbf{w} = \quad < z_x, z_y, \overline{x}, \overline{y}, \sigma_x, \sigma_y >, \tag{3.2}$$

where z_x and z_y are the x and y coordinates of the anchor point, $\overline{x}$ and $\overline{y}$ are the x and y coordinates of the arithmetic mean of the n crime locations,

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}, \qquad \overline{y} = \sum_{i=1}^{n} \frac{y_i}{n}, \tag{3.3}$$

and σ_x and σ_y are the standard deviations in the x and y directions of the n crime locations,

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}, \qquad \sigma_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n}}. \tag{3.4}$$

To employ kernel density estimation, a distance metric must be assigned to the 6D space and for simplicity we use Euclidean distance.

Given a set of N crime series (distinguished from n, the number of crimes in a particular series) vectorized as in (3.2), the next step is to scale the data to ensure that the sample variances are equal over all six dimensions, which will allow for the use of a single smoothing parameter [15]. The standard deviation $\beta_1, .., \beta_6$ with respect to each component $w_1, ..., w_6$ is taken over the N series,

$$\beta_m = \sqrt{\frac{\sum_{j=1}^{N} \left(w_m^j - \sum_{l=1}^{N} \frac{w_m^l}{N} \right)^2}{N}} \tag{3.5}$$

and the components of $\mathbf{w}^j$ for each crime series are scaled by the standard deviations yielding the scaled data,

$$\hat{\mathbf{w}}^j = \quad < \frac{z_x^j}{\beta_1}, \frac{z_y^j}{\beta_2}, \frac{\overline{x}^j}{\beta_3}, \frac{\overline{y}^j}{\beta_4}, \frac{\sigma_x^j}{\beta_5}, \frac{\sigma_y^j}{\beta_6} > .$$

Using the scaled data, we then determine the k-th nearest neighbor distance, d^j, for each crime series j in terms of the components of $\hat{\mathbf{w}}$ and the Euclidean distance metric,

$$d(\hat{\mathbf{w}}^i, \hat{\mathbf{w}}^j) = \sqrt{(\hat{\mathbf{w}}^i - \hat{\mathbf{w}}^j) \cdot (\hat{\mathbf{w}}^{i^T} - \hat{\mathbf{w}}^{j^T})}. \tag{3.6}$$

The hyperparameter k is the only parameter of the model, and we explore the effect of varying k in the next section.

Kernel density estimation of P then takes the form,

$$\hat{P}(\mathbf{w}) = \frac{1}{N|\mathbf{H}|^{1/2}} \sum_{j=1}^{N} K\left(\mathbf{H}^{-1/2}(\mathbf{w} - \mathbf{w}^j)\right), \tag{3.7}$$

where K is a kernel function and $\mathbf{H}$ is a symmetric positive definite matrix, the analog of a smoothing parameter in one dimension [15]. If this matrix is allowed to change according to a function $\mathbf{H}(\mathbf{w}^j)$, we arrive at variable-bandwidth kernel density estimation [16]. For the kernel function K we use a Gaussian kernel and for the smoothing matrix, $\mathbf{H}(\mathbf{w}^j)$, we use the diagonal matrix with entries $h_{11} = (\beta_1 d^j)^2,...,$ $h_{66} = (\beta_6 d^j)^2$ along the diagonal. Now (3.7) gives the joint probability $\hat{P}(w_1 = z_x, w_2 = z_y, w_3 = \overline{x}, w_4 = \overline{y}, w_5 = \sigma_x, w_6 = \sigma_y)$ given a set of N solved crime series as

$$\hat{P}(w_1, ..., w_6) \quad \propto \quad \sum_{j=1}^{N} \left[\frac{1}{(2\pi)^3 \prod_{l=1}^{6}(\beta_l d^j)} \prod_{l=1}^{6} e^{-\left(\frac{(\mathbf{w}_l - \mathbf{w}_l^j)^2}{2(\beta_l d^j)^2}\right)} \right]. \tag{3.8}$$

We are interested in the case where locations of crimes of a new series are known, but the anchor point is not. For potential anchor point (z_x, z_y), the geographic profile can be computed for such unsolved crime series,

$$< w_3^*, w_4^*, w_5^*, w_6^* >,$$

by evaluating (3.8) at the new series,

$$\hat{P}(z_x, z_y | w_3^*, w_4^*, w_5^*, w_6^*) \propto \tag{3.9}$$

$$\sum_{j=1}^{N} \left[\frac{1}{(2\pi)^3 \prod_{l=1}^{6}(\beta_l d^j)} e^{-\left(\frac{(z_x - w_1^j)^2}{2(\beta_1 d^j)^2} + \frac{(z_y - w_2^j)^2}{2(\beta_2 d^j)^2}\right)} \prod_{l=3}^{6} e^{-\left(\frac{(\mathbf{w}_l^* - \mathbf{w}_l^j)^2}{2(\beta_l d^j)^2}\right)} \right].$$

4. Results. Anchor points and crime locations for 221 solved crime series with $n \geq 3$ instances (with at least two occurring in separate locations), recorded by the LAPD between 2003 - 2008 are used to test our proposed KDE algorithm against the widely used CGT algorithm. For the KDE algorithm the leave-one-out method [6] is used so that a profile for series j is not estimated using series j. In order to quantitatively assess the methods, as well as for visualization of the profiles, the spatial domain containing the crime series is first divided into a grid of 128×128 cells, each of size $(\frac{140}{128})^2$ km^2. The method of assessment we employ then determines the probability mass (or in the case of CGT, the hit score) of the cell containing the anchor point for a crime series, then orders the masses of all the cells on the grid. The overall ranking of the mass of the anchor point is determined. In the case when cells have identical probability mass, the best possible ranking is assumed. The rank number is divided by $(128)^2$ to determine the percentile in which the probability mass lies. Figure 4.1 compares the number of anchor points correctly flagged by the CGT and KDE algorithms as a function of the percentage of the city flagged. Several different nearest neighbor k-values were tested to determine the sensitivity of KDE on the choice of k. The best value was determined to be $k = 1$, due to the high
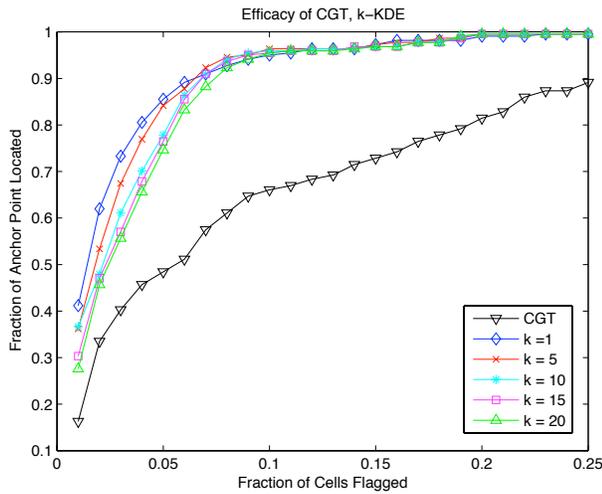
Fig. 4.1: A comparison of the efficacy of various k-values of KDE, CGT

dimensionality and low sample size, but KDE performs well in comparison to CGT over a wide variety of k values. The good performance of KDE for small search areas is particularly promising, since police can realistically only search a few percent of a city the size of Los Angeles.
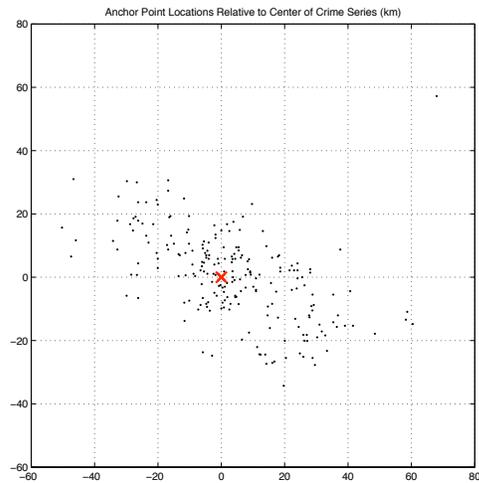

Fig. 4.2: Scatter plot of 221 anchor points relative to the center of the crime series.

For this dataset, KDE tended to stretch along the northwest-southeast axis compared to the radial expansion of CGT. The representation of the data as a collection of anchor points, arranged by their distance and direction to the center of the crime

series as in Figure 4.2, provides a statistical foundation for this behavior, revealing clustering patterns in the dataset. Figure 4.3 provides specific illustration of KDE elongation and displacement. A possible explanation for such biased NW-SE movement is the alignment of major thoroughfares in this direction including I-5, I-405, and Hwy 101 [6].
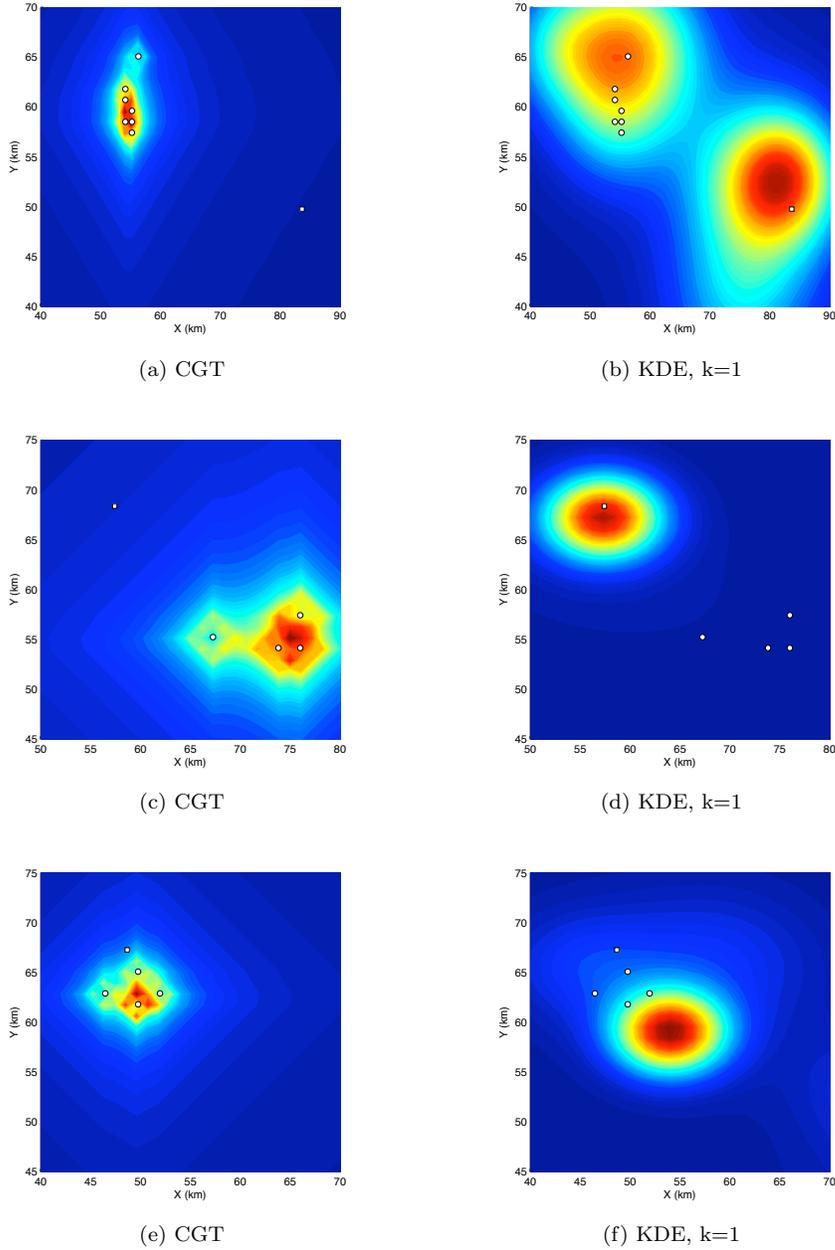

(a) CGT

(b) KDE, k=1

(c) CGT

(d) KDE, k=1

(e) CGT

(f) KDE, k=1

Fig. 4.3: Geoprofiles for different burglary series in Los Angeles, CA.

5. Discussion. Kernel density estimation has been used extensively within the geoprofiling community to create score functions in 2 dimensions. These methods rely on criminals committing crimes close to the anchor point and cannot handle more complex criminal behaviors. By using KDE to model historical crime series, rather than the crimes in a particular series of interest, the accuracy of geographic profiles is improved. The method is relatively easy to implement and requires only basic computational resources to calculate a geographic profile (generation of a profile for the largest crime series, containing 32 separate instances, takes a matter of seconds).

One possible advantage of the methodology is that it makes few assumptions about the nature of criminal behavior. While Bayesian methods are powerful when an accurate model for criminal target selection is readily available, such a model may be hard to construct, implement, or may not apply to all criminals in a data set. One disadvantage of 6D KDE is that it relies entirely on both the quantity and quality of the data available to the user. In the case that there is not enough data, KDE may not produce an accurate or useful geoprofile. For the method to be applicable it is necessary for police departments to keep meticulous records on all solved crime series.

While KDE adapts to the general direction and distance of travel, another disadvantage is that it has no way of explicitly incorporating geographic features of the environment, for example where there is no possibility of an anchor point (such as a body of water) [6]. There is also the question of data metrics. The use of the inputs $\overline{x}, \overline{y}, \sigma_x,$ and σ_y does not necessarily produce the most accurate results; there is research to be done inasmuch as the number of dimensions for KDE as well as the form of the dimension reduction. Inclusion of the covariance σ_{xy} as a seventh dimension is sensible from a modeling perspective, given the importance of directional elements to the problem of geographic profiling. However, tests performed over varying $k = 1, 5, 10, 15,$ and 20 indicate unaltered or diminished accuracy compared to the six-dimensional estimate. One suggestion for improvement might be the use of the Fermat-Weber point, or the point of minimized thoroughfare travel, instead of the arithmetic mean of the crime series.

REFERENCES

[1] P.J. Brantingham and P.L. Brantingham. *Environmental Criminology*. Sage, Oakland, CA, 1981.

[2] P.J. Brantingham and P.L. Brantingham. *Patterns in Crime*. Macmillan, New York, 1984.

[3] D. Canter, T. Coffey, M. Huntley, and C. Missen. Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, 16(4):457–478, 2000.

[4] D. Canter and L. Hammond. A comparison of the efficacy of different decay functions in geographic profiling for a sample of us serial killers. *Journal of Investigative Psychology and Offender Profiling*, 3:91–103, 2006.

[5] N. Levine. The evaluation of geographic profiling software: Response to kim rossmo's critique of the nij methodology. http://www.nedlevine.com/Response to Kim Rossmo Critique of the GP Evaluation Methodology.May 8 2005.doc, 2005. Accessed July 2011.

[6] G. O. Mohler and M. B. Short. Geographic profiling from kinetic models of criminal behavior. *Preprint, submitted*, 2009.

[7] National Institute of Justice. Geographic profiling. http://www.ojp.usdoj.gov/nij/maps/gp.htm. Accessed Aug 2011.

[8] M. O'Leary. The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, 6:253–265, 2009.

[9] M. O'Leary. A new mathematical approach to geographic profiling. Technical report, Towson University, Dept. of Mathematics, 2009.

[10] T. Rich and M. Shively. A methodology for evaluating geographic profiling software. Technical report, Abt Associates, Inc, Cambridge, MA, 2004.

[11] D. K. Rossmo. Fugitive migration patterns. 1987. Unpublished master's thesis, Simon Fraser University.

[12] D. K. Rossmo. Geographic profiling: Target patterns of serial murderers. 1995. Unpublished doctoral dissertation, Simon Fraser University.

[13] D.K. Rossmo. *Geographic Profiling*. CRC Press, New York, 1999.

[14] D.K. Rossmo. An evaluation of nij's evaluation methodology for geographic profiling software. http://www.txstate.edu/gii/documents/Responseto NIJ GP Evaluation Methodology.doc, 2005. Accessed July 2011.

[15] S. R. Sain. Multivariate locally adaptive density estimation. *Journal of the American Statistical Association*, 91:1525–1534, 1999.

[16] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, 1986.