Implementing a Bayesian approach to criminal geographic profiling

Mike O'Leary
Department of Mathematics
Towson University
Towson, MD
moleary@towson.edu

ABSTRACT

The geographic profiling problem is to create an operationally useful estimate of the location of the home base of a serial criminal from the known locations of the offense sites. We have developed and released new software based on Bayesian methods that attempts to solve this problem. In this paper, we discuss some of the geographic and computational challenges in implementing this new method.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Science—*sociology*; G.3 [**Mathematics of Computing**]: Probability and Statistics—*statistical computing*; I.6 [**Computing Methodologies**]: Simulation and Modeling—*applications*

General Terms

Algorithms

Keywords

Criminology, geographic profiling

1. INTRODUCTION

The geographic profiling problem is the problem of estimating the location of the home base of a serial offender from the known locations of the crimes in the series. At its core this is an operational problem for police agencies and it lies at the intersection of geography, criminology, mathematical modeling and computation. We will describe a new mathematical approach that we have developed for the geographic profiling problem [9]; moreover we have developed and released prototype software that implements those new methods. In this paper, we will discuss the details of how that theoretical framework was implemented with particular focus on the geographic and computational questions that arose with that method.

In the geographic profiling problem, an analyst is presented with the locations of a series of linked crimes presumably committed by the same offender. The analyst's goal is to construct an estimate for the likely location of the offender's anchor point. The anchor point may be the offender's residence, but it may also be some other location of importance to the offender- the residence of a friend or relative, a place of employment, or even a favorite bar or hangout.

There are a number of existing methods for the geographic profiling problem. One category are called spatial distribution strategies, following the terminology of [13]. These techniques produce a point estimate of the offender's anchor point through an estimate of the center of the crime series. One common spatial distribution strategy is to estimate the anchor point with the centroid (mean center) of the locations of the crime sites. A second strategy is to use the center of minimum distance. This point, also called the Fermat-Weber point, is chosen so that the sum of the distances from this point to the crime sites are at a minimum. This can be calculated iteratively via for example Weiszfeld's algorithm [3].

A second class of approaches are called probability distance strategies. These are the methods that have been implemented in the commonly used software tools that are currently being used by police agencies. To describe these methods, let us first fix some notation. We assume that the crime series consists of n linked crimes, and that these have taken place at the locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. The offender's anchor point will be denoted by $\mathbf{z}$. To craft a probability distance strategy, we need to make two selections- a distance function d and a decay function f. There are a number of reasonable choices for the distance function d; two of the most common choices are the Euclidean distance and the Manhattan distance. For points $\mathbf{x} = (x^{(1)}, x^{(2)})$ in the plane, Euclidean distance $d_2(\mathbf{x}, \mathbf{y})$ is given by

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{|x^{(1)} - y^{(1)}|^2 + |x^{(2)} - y^{(2)}|^2}$$

while the Manhattan distance $d_1(\mathbf{x}, \mathbf{y})$ is given by

$$d_1(\mathbf{x}, \mathbf{y}) = |x^{(1)} - y^{(1)}| + |x^{(2)} - y^{(2)}|.$$

The decay function f is used to model the relationship between the locations of the crime sites and the location of the offender's anchor point. There are a number of common decay functions in common use, including linear, normal, lognormal, and negative exponential; see [6, Chp. 10] for a discussion.

With a distance function d and a decay function f selected, a probability distance strategy then calculates a hit score $S(\mathbf{y})$ for each location $\mathbf{y}$ by summing

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

In effect, the hit score is found by placing a copy of the function $\mathbf{y} \mapsto f(d(\mathbf{x}_i, \mathbf{y}))$ on each crime site $\mathbf{x}_i$ and then summing the result. Regions with a high hit score are considered to be more likely to contain the offender's anchor point than regions with a low hit score.

Rossmo [11, Chp. 10] recommends the use of a Manhattan distance metric with a particular algebraic distance decay. Canter, Coffey, Huntley and Missen in [1] used a Euclidean distance metric and evaluated a family of distance decay functions modeled after negative exponentials. Ned Levine's CrimeStat program [6] allows for varying distance metrics and varying distance decay functions.

Though helpful, none of these approaches has been truly successful. Indeed, the National Institute of Justice [8]says

> "Though there have been anecdotal successes with geographic profiling, there have also been several instances where geographic profiling has either been wrong on predicting where the offender lives/works or has been inappropriate as a model."

See also Paulsen's analysis [10]. One of the issues with these existing methods is that none is able to truly incorporate geography in a meaningful way. To that end, a number of new approaches have since been developed for the geographic profiling problem, including the Bayesian journey to crime approach of Levine [6] and the technique of Mohler and Short [7] that begins with a kinetic model of offender behavior. In [9] we presented our new method based on Bayesian methods.

2. OUR BAYESIAN MODEL

We begin by modeling offender behavior; in particular we start with the assumption that an offender with anchor point $\mathbf{z}$ will commit a crime in the series at the location $\mathbf{x}$ with probability density $P(\mathbf{x}|\mathbf{z})$. This approach however, does not allow for the possibility that different offenders have different relative mobility. To correct for this, we assume that each offender has an average distance α that they are willing to travel to offend so that the appropriate probability density is $P(\mathbf{x}|\mathbf{z}, \alpha)$. In this approach a young teenager without a driver's license would have a smaller value of the average offense distance α than an older offender who was very familiar with the area.

If we then assume that locations in the series are selected independently, we find that the probability density that the offender selects the locations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ for the series is precisely

$$P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n|\mathbf{z}, \alpha) = \prod_{i=1}^{n} P(\mathbf{x}_i|\mathbf{z}, \alpha).$$

Bayes theorem then gives us the expression for the probability density of the offender's anchor point

$$P(\mathbf{z}) \propto \int_0^\infty \left[\prod_{i=1}^{n} P(\mathbf{x}_i|\mathbf{z}, \alpha)\right] H(\mathbf{z})\pi(\alpha)\, d\alpha \qquad (1)$$

were $\pi(\alpha)$ is the prior distribution for the average offense distance of α throughout the population and $H(\mathbf{z})$ is the prior distribution of anchor points. We have assumed that these priors are independent and we have marginalized over α.

To complete our model, we need to specify our form for $P(\mathbf{x}|\mathbf{z}, \alpha)$. We certainly expect that there is a distance decay component to this probability, so we start with the supposition that

$$P(\mathbf{x}|\mathbf{z}, \alpha) \propto D(d(\mathbf{x}, \mathbf{z}), \alpha) \qquad (2)$$

for some decay function D. However, to incorporate the impact of geography on target selection, we also suppose that that there is a function $G(\mathbf{x})$ that represents the attractiveness of a target at $\mathbf{x}$, and so suppose that

$$P(\mathbf{x}|\mathbf{z}, \alpha) \propto G(\mathbf{x}). \qquad (3)$$

In particular, if $G(\mathbf{x}) = 0$, we are assuming that the offender cannot offend at $\mathbf{x}$; this lets us account for regions where offenses are impossible- *e.g.* you cannot have a residential burglary in a body of water. Larger values of $G(\mathbf{x})$ indicate areas where offenses are more likely.

Combining (2) and (3), we obtain the expression

$$P(\mathbf{x}|\mathbf{z}, \alpha) = D(d(\mathbf{x}, \mathbf{z}), \alpha)G(\mathbf{x})N(\mathbf{z}, \alpha) \qquad (4)$$

where the normalization $N(\mathbf{z}, \alpha)$ satisfies

$$N(\mathbf{z}, \alpha) = \left[\iint D(d(\mathbf{x}, \mathbf{z}), \alpha)G(\mathbf{x})\, dx^{(1)}\, dx^{(2)}\right]^{-1} \qquad (5)$$

and is chosen to ensure that $P(\mathbf{x}|\mathbf{z}, \alpha)$ represents a probability distribution.

Thus (1), (4) and (5) together give us a theoretical expression that can be evaluated to estimate the geographic regions where the offender's anchor point is most likely. To do so, we will need to specify how to measure each of the quantities that appear above. A more thorough discussion of the theoretical background for this model can be found in [9]. In this paper, we want to discuss the geographic and computational issues that arise when trying to convert this theory into a practical tool.

3. DISTANCE AND DISTANCE DECAY

To convert this theoretical framework into a practical tool, we must begin by selecting a common framework for place and distance. To do so, we will represent points by their latitude and longitude using the same reference datum as the U.S. Census. Then, to measure distance between points $\mathbf{x} = (x^{(1)}, x^{(2)})$ and $\mathbf{y} = (y^{(1)}, y^{(2)})$ where the first coordinate is longitude and the second is latitude, we can use the great-circle distance

$$d(\mathbf{x}, \mathbf{y}) = 2\sin^{-1}\sqrt{\sin^2\left(\frac{x_2 - y_2}{2}\right) + \cos x_2 \cos y_2 \sin^2\left(\frac{x_1 - y_1}{2}\right)}.$$

Notice that the result of d is an angle; this is the central angle between the two rays from the center of the earth to the two points on the surface. It can be converted to approximate distance in miles by ensuring that d is measured in radians and then multiplying by the radius of the earth in miles. It should be noted that this approximation to the straight line distance does not account for features like elevation changes or the deviation of the surface of the earth from a sphere.

As yet there is no consensus on the best or correct form for the distance decay behavior of serial offenders; see for example [2, 5, 6]. We select a bivariate normal distribution for D, and assume that

$$D(d(\mathbf{x}, \mathbf{z}), \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2}\, d(\mathbf{x}, \mathbf{z})^2\right). \tag{6}$$

We can also examine the probability density for the distances traveled; as a function of α it has the density

$$f(r|\alpha) = 2\pi r D(d(\mathbf{x}, \mathbf{z}), \alpha) = \frac{\pi r}{2\alpha^2} \exp\left(-\frac{\pi r^2}{4\alpha^2}\right) \tag{7}$$

where $r = d(\mathbf{x}, \mathbf{z})$ is the travel distance. Note that $f(r|\alpha)$ is simply the the probability density that one of the sites at a distance r would be chosen namely $D(r, \alpha) = D(d(\mathbf{x}, \mathbf{z}), \alpha)$, multiplied by the number of sites at a distance r namely $2\pi r$, which is the circumference of a circle of radius r. As a consequence, we see that the distribution of distances follows a Rayleigh distribution in the distance r.

With $D(d(\mathbf{x}, \mathbf{z}), \alpha)$ chosen, we need to construct an estimate for $\pi(\alpha)$, which is the prior estimate for the distribution of average offense distances. It is fundamental to note that this is the distribution of the average offense distances *across offenders*. In particular, though this is related to the distribution of offense distances across offenses (obtained for example by examining crime statistics) the distribution here is across people.

To perform this estimation, let us first assume that we know that the distribution of distances from home to offense sites across known offenses is given by the function $A(r)$. Practically, we are unlikely to know the exact form of the distribution $A(r)$, but we can estimate it from crime statistics. Indeed, let us suppose that we have a sample of S solved crimes, and that the distance from offense site j to the corresponding offender anchor point is ρ_j.

Choose a discretization size $\epsilon > 0$, then define $r_j = j\epsilon$ and $r_j^* = (j - \frac{1}{2})\epsilon$, to subdivide the real axis into a sequence of bins $[r_{j-1}, r_j)$ each with center r_j^*. To estimate the value of A in the center of bin $[r_{j-1}, r_j)$, namely $A(r_j^*)$, we let a_j be the number of distances ρ_s in this bin,

$$a_j = \#\{s \mid r_{j-1} \le \rho_s < r_j\}. \tag{8}$$

Then we have the relationship

$$A(r_k^*)\epsilon \approx \frac{a_j}{S} \tag{9}$$

where both sides of approximate the probability that ρ lies in the bin $[r_{j-1}, r_j)$.

Returning to our estimate of $\pi(\alpha)$, we begin with the fundamental relationship

$$A(r) = \int_0^\infty f(r|\alpha)\pi(\alpha)\, d\alpha \tag{10}$$

which states that the number of offenses at the distance r can be found by by multiplying the probability density that an offender with average distance to offend α chooses the offense distance r by the probability density that an offender actually has the offense distance α, and then integrating over all possible values of α. In particular, this accounts for two sources of variation- the variation in offense distances selected by one offender, and the variation in average offense distances across multiple offenders.

We know that offenders do not travel infinite distances to offend, so we choose a number N so large that $A(r) \approx 0$ for $r > \epsilon N$; then we want to choose $\pi(\alpha)$ so that

$$A(r_j^*) = \int_0^\infty f(r_j^*|\alpha)\pi(\alpha)\, d\alpha$$

for $j = 1, 2, \ldots, N$. The assumption $A(r) \approx 0$ for $r > \epsilon N$, also lets us conclude that $\pi(\alpha) \approx 0$ for $\alpha > \epsilon N$. Indeed, $A(r)$ is the measured number of crimes that occur at the distance r from the anchor point, while $\pi(r)$ is the density of offenders whose average offense distance is r.

To evaluate the integrals, define $\alpha_k = k\epsilon$, $\alpha_k^* = (k - \frac{1}{2})\epsilon$, and apply the midpoint rule to the integral

$$\begin{aligned}\int_0^\infty f(r|\alpha)\pi(\alpha)\, d\alpha &\approx \int_0^{\epsilon N} f(r|\alpha)\pi(\alpha)\, d\alpha \\ &\approx \sum_{k=1}^N f(r|\alpha_k^*)\pi(\alpha_k^*)\epsilon + O(\epsilon^2).\end{aligned}$$

Thus, for each $j, k \in \{1, 2, \ldots, N\}$, we have

$$A(r_j^*) \approx \epsilon \sum_{k=1}^N f(r_j^*|\alpha_k^*)\pi(\alpha_k^*)$$

Then applying (7) and (9) we find

$$\begin{aligned}a_j &= S\epsilon^2 \sum_{k=1}^N f(r_j^*|\alpha_k^*)\pi(\alpha_k^*) \\ &= \frac{\pi S\epsilon}{2} \sum_{k=1}^N \frac{(j-\frac{1}{2})}{(k-\frac{1}{2})^2} \exp\left(-\frac{\pi}{4}\frac{(j-\frac{1}{2})^2}{(k-\frac{1}{2})^2}\right)\pi(\alpha_k^*).\end{aligned}$$

Thus, if we define the matrix

$$G = G_{jk} = \frac{\pi S\epsilon}{2} \frac{(j-\frac{1}{2})}{(k-\frac{1}{2})^2} \exp\left(-\frac{\pi}{4}\frac{(j-\frac{1}{2})^2}{(k-\frac{1}{2})^2}\right)$$

and the vectors

$$\mathbf{a} = (a_1, a_2, \ldots, a_N) \tag{11}$$

$$\pi = (\pi(\alpha_1^*), \pi(\alpha_2^*), \ldots, \pi(\alpha_N^*)) \tag{12}$$

then we obtain the discrete linear system

$$\mathbf{a} = G\pi. \tag{13}$$

Unfortunately, this linear system is ill-posed; this is a consequence of the fact that integral equations of the form (10) are

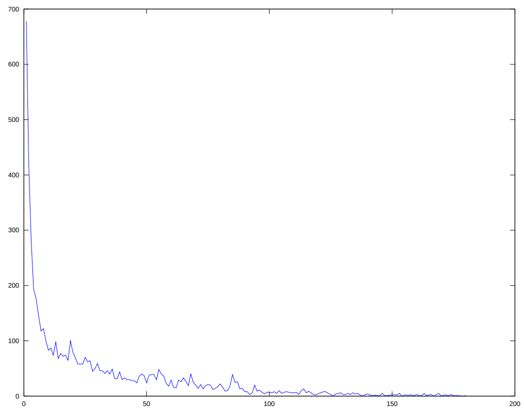

Figure 1: The vector a calculated from residential burglaries in Baltimore County, where $N = 180$ and $\epsilon = 0.002$.

themselves ill-posed. To proceed, we will instead apply Tikhonov regularization [4, 15]. A solution to the equation (13) can be thought of as the vector π that minimizes $\|G\pi - \mathbf{a}\|^2$; the idea in Tikhonov regularization is to instead minimize

$$L_\lambda(\pi) = \|G\pi - \mathbf{a}\|^2 + \lambda\|\pi\|^2. \tag{14}$$

As a consequence, the Tikhonov regularized solution is chosen to balance out the error obtained by fitting π to the (noisy) data (the term $\|G\pi - \mathbf{a}\|^2$) with an estimate of the size of π (the term $\lambda\|\pi\|^2$).

If one graphs the value of $\log\|G\pi - \mathbf{a}\|$ versus $\log\|\pi\|$ as λ varies for the minimizer in (14), one obtains a graph that has the general shape of an 'L'. Indeed, for small values of λ the ill-posedness of the problem implies that the size of $\|\pi\|$ grows rapidly, while for large λ the solution is unable to accurately fit the data, and so $\|G\pi - \mathbf{a}\|$ grows rapidly. We choose the value of λ nearest to the vertex of the 'L', *i.e.* the point of maximum curvature; this is called the 'L'-curve method.

To illustrate this, let us apply this method to residential burglaries in Baltimore County. We have a data set of 5863 solved residential burglaries for the county covering the period 1991-2008. If we select $\epsilon = 0.002$ as our bin size, and choose $N = 180$ bins, then the graph of the vector $\mathbf{a}$ appears in Figure 1

Attempting to solve the linear system directly leads to the wildly oscillatory solution seen in Figure 2, while the Tikhonov regularized solution gives the results in Figure 3. The minimization in the Tikhonov regularization (14) was taken over vectors π with all nonnegative entries; because π comes from a probability density, we know *a priori* that the components must be nonnegative.

4. TARGET ATTRACTIVENESS

One potentially reasonable approach to estimating the relative attractiveness of different targets would be to identify geographic and demographic variables that are correlated with the crime type under consideration. Interestingly,

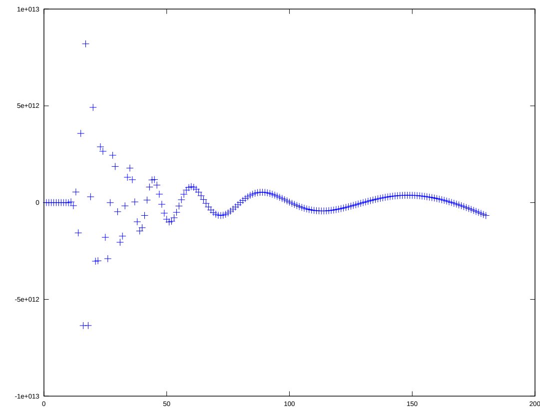

Figure 2: Attempt to directly solve (13) using the pseudo-inverse; the known data a is calculated from residential burglaries in Baltimore County, while $N = 180$ and $\epsilon = 0.002$.

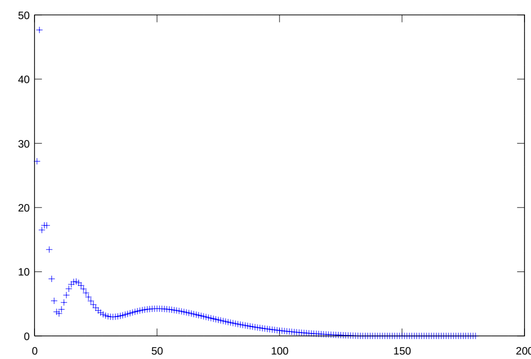

Figure 3: Tikhonov regularized solution with nonnegative entries to estimate π, plotted using data from Baltimore County residential burglaries. Here $N = 180$ and $\epsilon = 0.002$.

Tseloni, Wittebrood, Farrell and Pease [14] compared geographic features that influenced burglary rates across three different countries (Britain, the U.S., and the Netherlands). Though the effect of some variables on crime rates appeared consistent across the different nations, there were some variables that were significant in different nations, but in opposite directions. For example, increased household affluence indicated higher burglary rates in Britain, while it indicated lower burglary rates in the U.S. This suggests that this approach may not prove fruitful without a significant study of the particular crime in the particular jurisdiction where the series is being investigated- a rather high hurdle.

Our approach then, is to move away from techniques that try to predict regions of higher or lower crime likelihood in favor of simply measuring observed behavior. In particular, an officer investigating a series of crimes needs to begin with a representative sample of historical crimes of the same type as the series under consideration; let us suppose that these have occurred at the locations $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M$. Though these locations are just a sample, they are likely to be concentrated in areas that are likely crime sites; thus we want $G(\mathbf{x})$ to be large in regions where crimes are common while $G(\mathbf{x})$ should be small where crimes are uncommon. To convert this sample into a density defined everywhere, we apply kernel density parameter estimation [12]. In particular, we start with the family of quartic kernel functions $K(\mathbf{x}|\lambda)$ with bandwidth λ given by

$$K(\mathbf{x}|\lambda) = \begin{cases} \frac{3}{\pi\lambda^6}(|\mathbf{x}|^2 - \lambda^2)^2 & \text{if } |\mathbf{x}| \leq \lambda, \\ 0 & \text{if } |\mathbf{x}| \geq \lambda. \end{cases}$$

Notice that $\iint K(\mathbf{x}|\lambda)\, dx^{(1)}\, dx^{(2)} = 1$ for all λ, while the radius of the region where K is nonzero is exactly λ. Then we can construct an approximation of our target attractiveness by calculating

$$G(\mathbf{x}) = \sum_{i=1}^{N} K(\mathbf{x} - \mathbf{c}_i|\lambda)$$

for any choice of bandwidth λ. The question of the optimal selection of the bandwidth parameter λ remains open.

5. ANCHOR POINT DENSITY

We also need to construct a method to estimate the prior distribution of offender anchor points before we account for the precise details of the crime series. Our approach is to start with the supposition that the local density of offender anchor points is proportional to the local population density. The advantage of this approach is that block-level population data is available directly from the U.S. Census. Another reasonable approach would be to start with the locations of the anchor points of historically identified offenders and proceed via kernel density parameter estimation as was done target attractiveness.

If we want to use the U.S. Census data, we need to convert the block level data into an appropriate distribution $H(\mathbf{z})$. One approach would be to let $H(\mathbf{z})$ be defined piecewise, with the value on each block equal to the population density of that block. Though this most closely matches the available data, it runs into the modifiable areal unit problem. A related issue is that the results, when presented to the investigating officer need to be presented in the form of a map. If the resolution of the underlying map is comparable to the size of a block, then the piecewise nature of the block level data may result in discretization errors.

As an alternative, we choose a modification of the kernel density parameter estimation process, and define

$$H(\mathbf{z}) = \sum_{i=1}^{N_{\text{blocks}}} = p_i K(\mathbf{z} - \mathbf{q}_i|\sqrt{A_i})$$

where each block has population p_i, center $\mathbf{q}_i$ and for each block we have chosen a different bandwidth equal to the side length of a square with the same area A_i as the block. This has the effect of smoothing out the population density from one block into neighboring points.

6. GEOGRAPHY

The next fundamental problem is determining how we want to represent the underlying geography in our code. Because we are representing points by their longitude and latitude, our geographic region is just a small subset of the plane. We then subdivide this portion of the plane into a family of equilateral triangles with common circumradius and disjoint interiors.

The circumradius of these triangles then determines the smallest spatial scale that our code can resolve. Moreover, because we know that both population density and target attractiveness vary rapidly over scales as small as the diameter of a Census block, we know that this circumradius cannot be significantly larger that this value. On the other hand, this presents a significant computational hurdle, as halving the circumradius will quadruple the number of triangles in the mesh, and hence quadruple the amount of computation.

To address this issue, we instead use two triangular meshes, a coarse grid and then a fine refinement of this mesh. This approach allows us to precompute values like $G(\mathbf{x})$ and $H(\mathbf{z})$ once and only once, and to store these values for each element of the fine mesh. On the other hand, when evaluating the normalization integral (5), we can use only those elements of the fine mesh that are most significant for the computation; less significant elements can be approximated by the corresponding element in the coarse mesh, while the least significant elements can simply be ignored.

7. NORMALIZATION

The normalization function $N(\mathbf{z}, \alpha)$ is defined by (5); for simplicity in exposition, we start by considering the inverse

$$I(\mathbf{z}, \alpha) = \iint D(d(\mathbf{x}, \mathbf{z}), \alpha) G(\mathbf{x})\, dx^{(1)}\, dx^{(2)}. \qquad (15)$$

This integral is a challenge to evaluate numerically; to do so, we begin with our coarse mesh Δ of equilateral triangles with disjoint interiors. Then we rewrite the integral above as

$$\begin{aligned} &\iint D(d(\mathbf{x}, \mathbf{z}), \alpha) G(\mathbf{x})\, dx^{(1)}\, dx^{(2)} \\ &= \sum_{T \in \Delta} \iint_T \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2} d(\mathbf{x}, \mathbf{z})^2\right) G(\mathbf{x})\, dx^{(1)}\, dx^{(2)}. \end{aligned}$$

Using the midpoint method, we then obtain the approximation

$$\begin{aligned} &\iint D(d(\mathbf{x}, \mathbf{z}), \alpha) G(\mathbf{x})\, dx^{(1)}\, dx^{(2)} \\ &\qquad \approx \frac{3\sqrt{3}}{16} \frac{1}{\alpha^2} \sum_{T \in \Delta} R_T^2 \exp\left(-\frac{\pi}{4\alpha^2} d(\mathbf{x}_T, \mathbf{z})^2\right) G(\mathbf{x}_T) \end{aligned}$$

where R_T is the circumradius and x_T is the centroid of the triangle T.

If the triangle $T \in \Delta$ satisfies $d(\mathbf{x}_T, \mathbf{z}) > 3\alpha$, then

$$\exp\left(-\frac{\pi}{4\alpha^2} d(\mathbf{x}_T, \mathbf{z})^2\right) < e^{-9\pi/4} \approx 0.000851;$$

for this reason we ignore such triangles in our approximation. On the other hand, for triangles $T \in \Delta$ that satisfy $2\alpha <$

$d(\mathbf{x}_T, \mathbf{z}) \le 3\alpha$, we only have

$$\exp\left(-\frac{\pi}{4\alpha^2} d(\mathbf{x}_T, \mathbf{z})^2\right) < e^{-\pi} \approx 0.0432$$

so we will retain all of these triangles. Because the largest contribution to the integral comes from the remaining triangles for which $d(\mathbf{x}_T, \mathbf{z}) \le 2\alpha$, each of these triangles will then be subdivided into their fine triangles before the sum is evaluated.

Although we have outlined how we are able to evaluate the integral $I(\mathbf{z}, \alpha)$, we do not want to do so for every potential value of $\mathbf{z}$. Indeed, because we merely need to generate a map of the result to present to the analyst, it is sufficient to simple evaluate $P(\mathbf{z})$ and hence $I(\mathbf{z}|\alpha)$ for points $\mathbf{z}$ that are the centroids of triangles in the fine mesh. However, this turns out to be impractical, as the computation time to evaluate $I(\mathbf{z}|\alpha)$ even once is significant. Testing has shown that this process, simplified as it was in the previous discussion, is still by far the most computationally expensive portion of the algorithm.

Rather than use this algorithm at the centroid of every fine triangle, we would like to use this method only to calculate the values of $I(\mathbf{z}|\alpha)$ on the vertices of the coarse triangles and interpolate into the fine triangles within. Unfortunately, the approximation of $I(\mathbf{z}, \alpha)$ by linear interpolation within a coarse triangle does not always produce reasonable results; in fact the accuracy of the approximation deteriorates as $\alpha \downarrow 0$.

Recall that α is the average distance the offender is willing to travel and that the dependence on $\mathbf{z}$ of the integrand is through (6). Examining this, we see that if $\mathbf{z}_1$ and $\mathbf{z}_2$ are far apart relative to α, then $d(\mathbf{x}, \mathbf{z}_1)/\alpha$ and $d(\mathbf{x}, \mathbf{z}_2)/\alpha$ are very different, and so $I(\mathbf{z}_1, \alpha)$ and $I(\mathbf{z}_2, \alpha)$ are likely different.

To proceed, we write the integral as

$$I(\mathbf{z}, \alpha) = \frac{1}{4\alpha^2} \iint \exp\left(-\frac{\pi}{4}\frac{|\mathbf{x}-\mathbf{z}|^2}{\alpha^2}\right) G(\mathbf{x}) dx^{(1)} dx^{(2)}$$

Set $\xi = \frac{1}{2\alpha}(\mathbf{x} - \mathbf{z})$, then $d\xi = \frac{1}{4\alpha^2} d\mathbf{x}$ so

$$I(\mathbf{z}, \alpha) = \iint \exp(-\pi|\xi|^2) G(\mathbf{z} + 2\alpha\xi)\, d\xi.$$

From this we can clearly see that $I(\alpha) \to G(\mathbf{z})$ as $\alpha \downarrow 0$.

Continuing our analysis, we can replace G by its Taylor series; then

$$I(\mathbf{z}, \alpha) = \iint \exp(-\pi|\xi|^2)\Big\{G(\mathbf{z}) + 2\alpha\, DG(\mathbf{z}) \cdot \xi + 4\alpha^2 \xi^\top D^2 G(\mathbf{z})\xi + O(\alpha^3)\Big\}\, d\xi.$$

Note that the second term vanishes; indeed

$$\iint \exp(-\pi|\xi|^2) DG(\mathbf{z}) \cdot \xi\, d\xi = \int_0^{2\pi} \int_0^\infty e^{-\pi r^2} |DG(\mathbf{z})| r \cos\theta \cdot r\, dr\, d\theta = 0$$

where θ is the angle measured from the direction $\frac{DG(\mathbf{z})}{|DG(\mathbf{z})|}$. Thus, we can write

$$I(\mathbf{z}, \alpha) \approx G(\mathbf{z}) + c_2(\mathbf{z})\alpha^2 + O(\alpha^3)$$

for small α.

With this in mind, we wish to approximate $I(\mathbf{z}, \alpha)$ for α near 0 by interpolation. We suppose that we already have approximations for $I(\mathbf{z}, \alpha)$ for $\alpha = \alpha_1 < \alpha_2 < \cdots < \alpha_N$, but that the accuracy of the approximations diminishes as $\alpha \downarrow 0$.

We begin by choosing K so large that our approximations are reasonable at α_k for $k \ge K$. In particular, we choose K so large that $\alpha_K^* \ge 2R$, where R is the circumradius of our coarse triangle. For each fixed $\mathbf{z}$ we then use the Hermite approximation $I(\mathbf{z}, \alpha) \approx \psi(\alpha)$ where

$$\psi(\alpha) = \omega_0 + \omega_1(\alpha - \alpha_K) + \omega_2(\alpha - \alpha_K)^2 + \omega_3\alpha(\alpha - \alpha_K)^2$$

so that

$$\psi'(\alpha) = \omega_1 + 2\omega_2(\alpha - \alpha_K) + \omega_3(\alpha - \alpha_K)^2 + 2\omega_3\alpha(\alpha - \alpha_K).$$

The coefficients ω_0, ω_1, ω_2 and ω_3 are chosen so that

$$\psi(0) = I(\mathbf{z}, 0) \qquad \psi(\alpha_K) = I(\mathbf{z}, \alpha_K)$$
$$\psi'(0) = \frac{\partial I}{\partial \alpha}(\mathbf{z}, 0) \qquad \psi'(\alpha_K) = \frac{\partial I}{\partial \alpha}(\mathbf{z}, \alpha_K)$$

From our Taylor series approximations near $\alpha = 0$, we concluded

$$I(\mathbf{z}, 0) = G(\mathbf{z}) \qquad \frac{\partial I}{\partial \alpha}(\mathbf{z}, 0) = 0.$$

The value of $I(\mathbf{z}, \alpha_K)$ we have from our approximations, while we use

$$\frac{\partial I}{\partial \alpha}(\mathbf{z}, \alpha_K) \approx \frac{I(\mathbf{z}, \alpha_{K+1}) - I(\mathbf{z}, \alpha_K)}{\alpha_{K+1} - \alpha_K}.$$

8. THE COMPLETED PROGRAM

We have completed and released prototype software that implements these methods; the tool is currently undergoing efficacy testing.

To illustrate the use of the program, we have applied it to a sequence of convenience store robberies that occurred in Baltimore County. There were six elements in the series; but the first, fourth and sixth crime in the series all occurred at the same location. We have presented the result of our algorithm in Figure 5; the crime location that is multiply labeled is the location of the multiple offenses.

In contrast to the spatial distribution strategies, our approach does not produce a single point estimate for the offender. We also notice that the regions that the algorithm predicts are most likely to contain the offender are not centered around the geographic center of the crime locations. Rather, it is skewed towards the east- that is it is skewed towards Baltimore City and away from the less populated suburbs.

Another reason that the search area is skewed towards Baltimore City is that the data in our series contains only the known crime locations from the Baltimore County police. Baltimore County and Baltimore City have separate police departments, and we only have the data from the County.

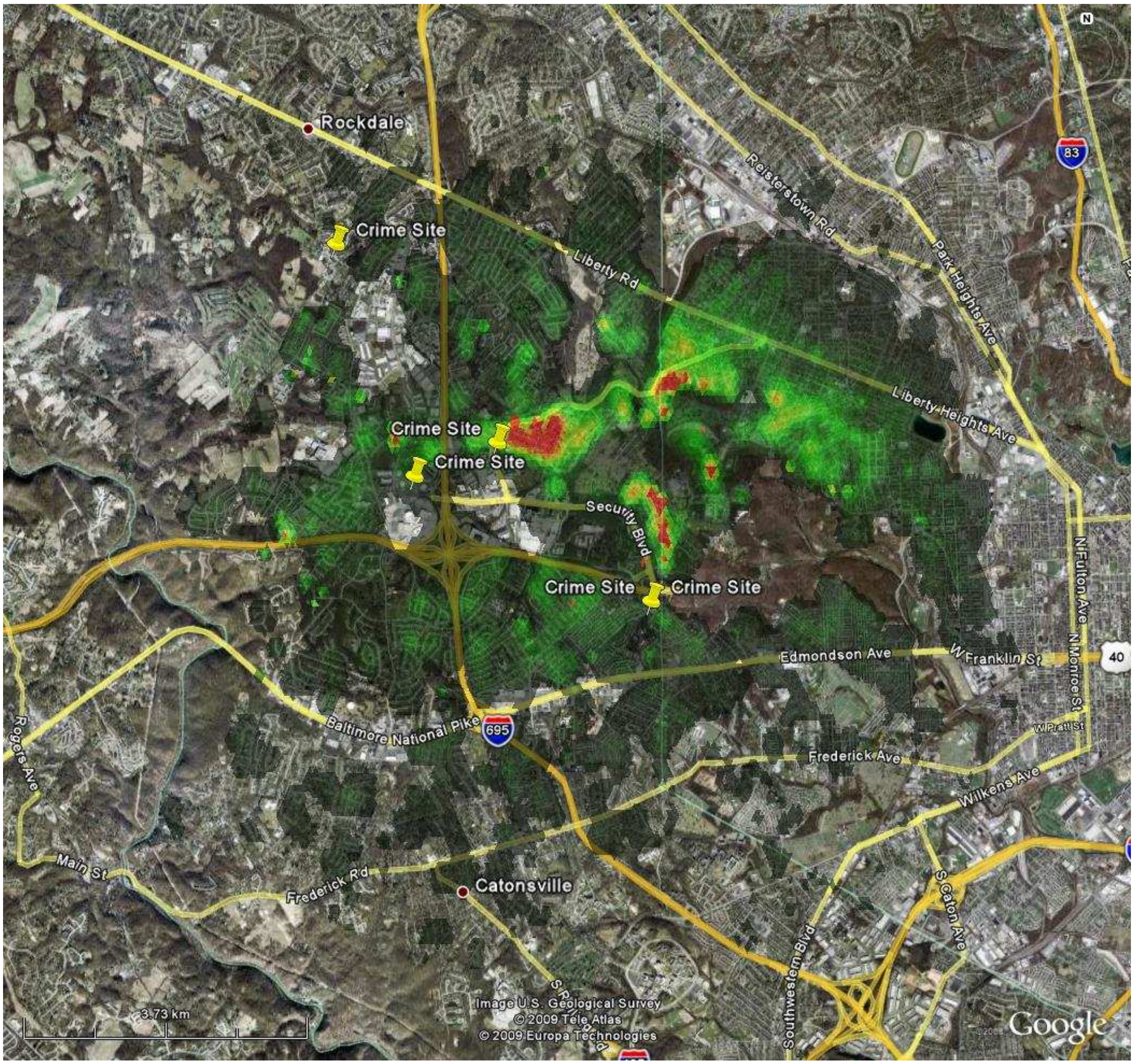

Figure 5: The result of applying this method to a series of convenience store robberies in Baltimore County. [The boundary between the county and the city can be seen in the figure as a north-south line that turns to a northwest-southeast line in the southeastern portion of the figure.]

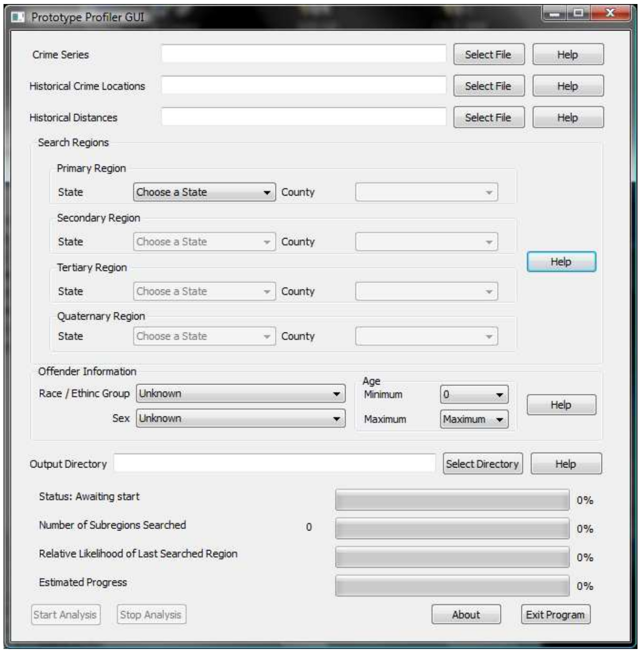

Figure 4: Screenshot of the prototype software.

In particular, this means that the offender may have committed additional crimes in the series inside Baltimore City, but they are unknown to us. Through the target attractiveness $G(\mathbf{x})$, the algorithm is able to recognize that, if the offender wanted to commit an offense that we would know about, then it would have to take place in the county. Thus, it stands to reason that the offender is more likely to have an anchor point closer to the city where crimes would not be known to us, than in the suburbs where we know the offender does not have any other series elements.

We also note that our method is able to account for a number of the salient features of the local geography. Indeed, unlike a probability distance strategy, the proposed search area rates commercial areas near the crime site quite low, and avoids the large park inside the city. Instead it gives high ratings to regions near major streets with significant population.

9. ACKNOWLEDGMENTS

This work has been supported by the National Institute of Justice, through grants 2009-SQ-B9-K014, 2007-DE-BX-K005, and 2005-IJ-CX-K036.

We express out thanks to Phil Canter of the Baltimore County Police Department for providing the data and for many helpful conversations.

10. REFERENCES

[1] 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.

[2] D. Canter and L. Hammond. A comparison of the efficacy of different decay functions in geographical profiling for a sample of us serial killers. *J. Investig. Psych. Offender Profil.*, 3:91–103, 2006.

[3] R. Chandrasekaran and A. Tamir. Open questions concerning Weiszfeld's algorithm for the Fermat-Weber location problem. *Mathematical Programming*, 44(1):293–295, 1989.

[4] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM, 1998.

[5] J. Kent, M. Leitner, and A. Curtis. Evaluating the usefulness of functional distance measures when calibrating journey-to-crime distance decay functions. *Computers, Environment and Urban Systems*, 30(2):181–200, 2006.

[6] N. Levine. Update notes to version 3.2. CrimeStat: A spatial statistics program for the analysis of crime incident locations. Online at http://www.icpsr.umich.edu/crimestat, (Accessed March 2010), 2009.

[7] G. O. Mohler and M. B. Short. Geographic profiling from kinetic models of criminal behavior. Online at http://www.math.ucla.edu/ gmohler/geopro2.pdf, (Accessed March 2010).

[8] National Institute of Justice. Online at http://www.ojp.usdoj.gov/nij/maps/gp.htm, (Accessed March 2010).

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

[10] D. J. Paulsen. Connecting the dots: assessing the accuracy of geographic profiling software. *Policing: An International Journal of Police Strategies & Management*, 29(2):306–334, 2006.

[11] K. Rossmo. *Geographic Profiling*. CRC Press, 2000.

[12] B. Silverman. *Density Estimation for Statistics and Data Analysis*. CRC Press, 1986.

[13] B. Snook, M. Zito, C. Bennell, and P. J. Taylor. On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology*, 21(1):1–26, 2005.

[14] A. Tseloni, K. Wittebrood, G. Farrel, and K. Pease. Burglary victimization in England and Wales, the United States, and the Netherlands. *British Journal of Criminology*, 44(1):66–91, 2004.

[15] C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, 2002.