Following the Trail of Data

Control # 8362

February 22, 2010

Summary Statement

As contracted by the police, our task was to make a model to help apprehend criminals and prevent future offenses by predicting their locations and the locations of such offenses. We chose to generate probability plots to help assist the police in assessing which areas are of higher priority for searching and monitoring. We assume we are given crime site locations and times for past offenses of the criminal whom we are considering.

Our model determines the probability that a criminal will be found at any given location in the search area by evaluating a distance decay function for each crime scene location and adding up the probabilities, generating a distribution which can be used to prioritize a search for the criminal. The distance decay function can take various forms, several of which we present in the paper. We use a linear regression to predict the time of the next crime and compute the probability that a crime will occur in a given spot by calculating how much that crime site would change the probability distribution for the criminal's location. This finds the location which best fits with the previous crime sites. We also present a model that uses the time of crime to weight the locations. This gives the model the ability to predict changes in the criminal's patterns.

To test our model with realistic data, we complied several case studies. We asess the quality of our predictions using multiple metrics: error distance and search cost. In the case of Peter Sutcliffe, our model was able to reduce the search area to 15% of its original size, and predicted the position of his next victim with very high accuracy. We produce color-gradient plots of the probability distributions which we present in the paper.

Executive Summary

Dear Chief of Police,

We have developed a model to help aid you in investigations regarding serial criminals. I am writing to inform you of its functionality and what cases are most appropriate for its uses.

Our model requires input from any member of your squad of the time and locations, in latitude and longitude, of the criminal acts you suspect were committed by any specific criminal. It then uses the locations of these points, along with a function modeling the drop off in probability as the distance from the crime site increases, to generate a probability distribution of the criminal's location. Further, it creates a probability distribution of possible *future* crime scenes by figuring out which point would be least surprising in future probability distribution calculations. Overall, it tends to guess the locations relatively accurately, especially when it comes to future crimes.

Unfortunately, our model is not a godsend. It assumes that all input on a given run is for a single criminal, even though its inherent clustering could provide relevant data in the case of multiple criminals. Similarly, we assume that the killer has just one base location from which he or she operates. Our model does not work well for commuter criminals, or those driving long distances in order to commit crimes.

Additionally, it relies on the hypothesis that the probability of the criminal's base being at a given location decreases as you move away from the crime scenes. Furthermore, the model ignores any variations in terrain and in population density, instead assuming that they are both homogeneous.

Despite these obvious drawbacks, our model does have many strengths. In addi-

tion to the aforementioned ability to correctly handle clusters of points, it also gives a very well specified and intuitive suggested search order, helping police minimize time spent searching for the killer. By abstracting its calculation mechanism, our model is flexible enough to allow your department to select the most appropriate of many possible "submodels" for your specific case. Thus, your department need not group bank robbers and rapists in the same category: if the two criminal groups tend to have different criminal patterns, our model allows for your department to use the techniques you find most appropriate.

Furthermore, our model is relatively efficient. Despite a mediocre "classical" running time, the extremely parallelizable nature of our algorithm allows for many computers to work on the problem in parallel. With only a small number of machines, our model can complete extremely detailed (one million point) calculations in 4 hours, and ten thousand point calculations (which is still sufficient detail for most cases) in a couple of seconds. Thus, your department need not worry about waiting days or weeks for results to come in, saving invaluable time.

Our model's output consists of two maps: one with the probability distribution of the criminal's location and one with the prediction of the location of his/her next crime. The probabilities are displayed in bands of color: a red band represents an area of high priority and a blue band represents one of low priority. In the case of attempting to locate the criminal, we suggest searching the red areas first. In the case of trying to prevent a future crime, we suggest spreading as many officers as possible in the red and yellow bands, as these regions are likely to contain the criminal's next target.

We wish you luck in your investigations.

Contents

1 Background

We attempt, in this paper, to present models and future research avenues to further the effectiveness of geographic profiling in locating serial criminals and their future crimes. Motivation for modeling includes the fact that humans are affected by "prior expectations, overconfidence, information retrieval, and information processing" (Snook et al., 2004) whereas computer simulations remain unbiased and make predictions based on discrete data. While the category of serial criminals includes rapists, killers, robbers and burglars, we specifically focus our research on the effort of capturing and inhibiting the crimes of serial killers. We will also discuss the applications of our research, in our final discussion, to other types of serial criminals.

1.1 Biography of the Typical Serial Killer

When considering serial killers, we will conform to the norm set to us in previous literature (Arndt et al., 2004):

- Must have killed a minimum of three.
- Must include a cooling-off time between killings, to distinguish from mass and spree killers.

We do not consider mass murderers because they are generally classified as killing many people in a short period of time, making future crimes unlikely and capture of them relatively easier. Spree killers kill few within a short period of time; although our model may very well apply to them, they may not kill again or may reveal themselves more readily. Serial killers are distinct from these other two categories in that they typically have a defined psychological purpose for killing and have spread-out killings, providing the police with more sufficient time to assemble a search.

Usually traumatized during childhood, repercussions of abuse and homicide beginning for the average serial killer at age 27.5 (Hickey, 2002). Many have precursors to killings, including robberies or sexual abuse. This is based on Hickey's Trauma Control Model of the Serial Killer and seems to be supported statistically; most noticeably, 84% admitted to assaults on adults during adolescence (Arndt et al., 2004). When a serial killer does begin his homicidal campaign, each killing leads to more habituation and tolerance to the hormones and relief provided which do not fully alleviate the cravings, leading to decreased period between killings and a cyclical process (Arndt et al., 2004). The career length of a specific serial killer varies, ranging from around 1 to 5 years, although many may enter prison on other accounts during such a career. The number of killings also varies drastically, as with their IQ.

1.2 Locating a Killer

To account for all factors affecting specific killing sites and travel distances would be impossible, as the human decision process is dynamic and virtually boundless.

As reported by Laukkanen and Santtila, however, "87% of serial rapists in the UK committed their crimes confirming the so-called *circle hypothesis*." This technique predicts that the criminal will live within a circle drawn around the two crime locations furthest from each other. Because difficult-to-solve rapes and homicides tend to occur within similar distances from the perpetrator's residence (Santtila et al., 2007), we assume that similarly most serial murders occur within the circle.

Most serial killers and other types of criminals are also shown to not travel far from their base location, with crimes committed dropping as a function of distance (Santtila et al., 2007) - a theory we will refer to by its common name as *distance decay*. Surround the crime scenes, there also usually exists a buffer zone in which the criminal is not likely to live when avoiding detection.

2   The Problem

As contracted by the police, we are given crime scene locations and times of a specific criminal. Based on this data, we must predict the criminal's base location and their next crime's location and time.

Although humans may appear random, we must fit a geographical profile to the crime scene data for prediction purposes in a way such that our model can be used for any type of serial criminal with any personality type. This geographical profile generates a prioritized search for the police. In our plots, note that red specifies high search priority and blue low. We use bilinear interpolation to smooth the grid for ease of interpretation for the police.

3   Assumptions

As given such location and time data from the police, we assume that the police suspect all points used in a single run of our model are crime sites from a single criminal. We also assume, to simplify locating the criminal, that the specified killer has precisely one home base which we are trying to find.

To simplify the problem, we assume we are dealing with serial killers. Although this does not seriously affect our results in any way (Laukkanen & Santtila, 2006; Santtila et al., 2007), we use case studies and research to support location statistics for serial killers and later extend our model results to other types of serial criminals. Also, despite the fact that different criminal personalities will commit crimes at different distances, we assume a form of the circle theory. Research states that 87% of serial rapists, which can be extended to other types of serial criminals, reside within a circle proposed by the *circle hypothesis* (see Background - Locating a Criminal) (Santtila et al., 2007); we assume our grid, although not necessarily encompassing the full circle around the furthest two points, will act similarly in encompassing our killer's location for ease of modeling purposes. We also assume a distance decay, as

supported by research for typical stable criminals. Thus, we currently do not consider commuters who may travel far distances to commit crimes, though in the future we would like to include an option for different personality traits to compensate for distance preferences and account for such commuter criminals. We do not inherently assume a buffer zone, as will be later discussed, but instead test different models and see the accuracy of such an assumption.

Some factors which may affect a criminal's choice of crime site include population density, opportunity, and landscape. In our model, we neglect these differences and assume a homogeneous geographical area for simplicity. If actually implemented, we would like to include these descriptors in our model, similar to *CrimeStat* (Levine, 2006). Thus, unfortunately our model may predict the criminal's base location to be in an inhabitable area and may predict future crimes to be committed where there is, in actuality, little opportunity for such crimes.

For more discussion on our future extensions to dispose of such simplifying assumptions, see Future Research.

4   Metrics and Functions

4.1   Simple Spatial Metrics

When assuming we are given information including locations of N crime scenes, it is useful to look at spatial metrics to help define global traits of the locations. We define the following metrics:

Mean Center ("Center of Mass")

This is the point output when taking the average values of latitudes (y) and longitudes (x):

$$c_x = \sum_{i=1}^{N} w_i x_i, \ c_y = \sum_{i=1}^{N} w_i y_i \,.$$

It is possible to include a weight, w_i, for each point to skew the center more toward it, which may be useful when specific points are more important than others.

Center of Minimum Distance

This is the point where the summed distances to all of the crime scenes is minimal, or the minimum considering every (x, y) point of

$$\sum_{i=1}^{N} \sqrt{(x_i - x)^2 + (y_i - y)^2}$$

.

Standard Deviational Circle (SDC)

Here, a circle is drawn at the mean center, assuming $w_i = 1$, such that the radius is one "standard distance" (related to standard deviation) to cover 68% of the data points (assuming randomness) and 95% for a radius of two standard distances. The larger the radius, the more spread out the data is considered to be.

Standard Deviational Ellipse (SDE)

While the standard deviational circle ignores skew of the data, the ellipse allows stretching in two dimensions along any two perpendicular lines on our plane.

4.2   Simple Cases and Their Spacial Metrics

In order to gather intuition about what we would expect from a model given crime sites, we illustrate a few simple examples of possible killing distributions with expectations of where we would like to see predictions of home bases and future killings.[1] We analyze some simple spatial statistics for each basic case: mean center ("center of mass") and standard deviational circle.

Uniform Spread

First, we look at a uniform distribution of killings within some shape - a square, specifically, for its ease of modeling on a grid. While we might expect the killer to live anywhere in the boundary of this square, it would not be surprising if the killer lived directly in the center for the comparative ease of travel to all points on the square. For future killings, it may be difficult to discern a specific location as the spread is homogeneous.
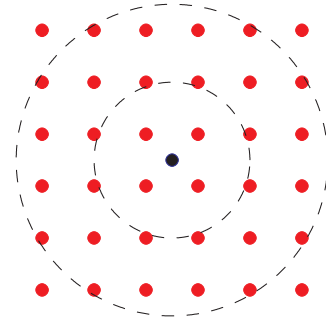
   The center of mass is directly in the center, as expected. The SDCs are efficient in grouping the data, meaning that our data is not very spread out.

Outline Spread

Here we consider killings along the edges of a square. It is not clear where the killer may live, as this killing trail may be part of his daily route or he may live closer to the center for ease of travel, but we would expect future killings to happen also along the boundary.
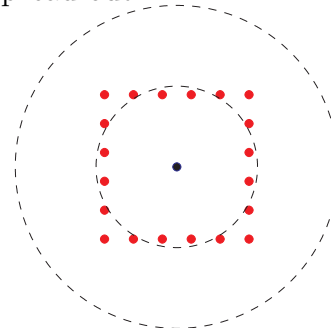
[1]These models were designed for their intuitive simplicity and to test the accuracy of our models later.

Our mean center is again in the center, but our SDCs are very large, indicating that the data is very spread out from the mean center.

Linear Spread

Here, we look at a linear killing distribution. While the future killings may depend on the times of each crime committed, specifically in this example, if we assume the times are irrelevant then we may expect the killer to live near at least a few of the points.

Again, our mean center is in the middle and our SDCs are very large, as the points are spread far from the center. The SDE may be more effective in analyzing this case as there is no spread in latitude.
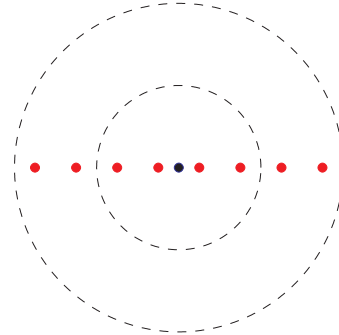
Clustering

This clustering example may be most prevalent when considering a killer with two main base locations (for example, one who lives in one city and works in another). It again may not be clear where the killer lives, but we should predict killings to occur within the two clusters.

Again, we have a large SDC. These may indicate that when looking for next killer location, it may be beneficial to look at SDC radii and SDEs.
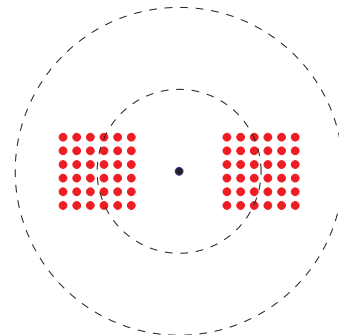
4.3   Distance Decay Functions

When calculating probability distributions, we test different distance decay functions. Recall that the theory of distance decay assumes drop-off of likelihood of our killer living at any given location as distance from crime scenes increase (see Figure 1). We consider Linear, Normal, Negative Exponential, Truncated Negative Exponential, and Plateaued Negative Exponential drop-offs. Of particular distinction between these are their behaviors directly surrounding the crime scenes and the rates of decay with increased distance.

Linear

With this model, we assume a linear drop-off rate of the form:

$$f(x) = \begin{cases} a - bx & \text{if } x \le a/b \\ 0 & \text{if } x > a/b. \end{cases}$$
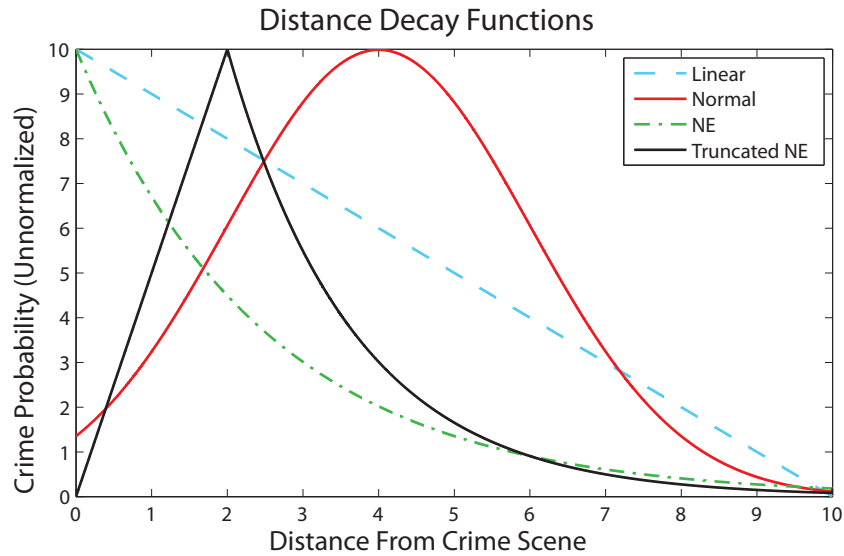
Figure 1: A decay plot.

This simple model assumes that the expected probability of finding the criminal at a given location decreases linearly with distance from the crime site until the distance is greater than a/b, after which it remains zero. Despite the simplicity of this model, its behavior is clearly somewhat unrealistic.

Normal

This model was suggested by Brantingham and Brantingham (1981) to include their hypothesized buffer zone. The function exponentially increases and decreases around such a radius as is maximized by the buffer zone:

$$f(x) = \frac{ae^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\ .$$

Although this does not work well with our idea of serial killers generally living close to their crime scenes, it may be useful when considering other types of serial criminals such as robbers and burglars are expected to on average live further from their crimes.

Negative Exponential

Here we assume a negative exponential drop-off rate of the form:

$$f(x) = e^{-cx}$$

where f(0)=1 is the maximum value. Thus, this function assumes that the probability of killer location is highest at the specific crime scenes from which it decays

exponentially. Although this contradicts our idea of a "buffer zone", presented by Brantingham & Brantingham (1981), this model is shown to work well in its predictions. According to Rhodes, it showed a much better than the Normal curve fit when compared with actual serial burglar, robber, and rapist locations versus their crimes (Rhodes & Conly, 1981; Kent, 2003).

Truncated Negative Exponential

Here we assume a linear growth from each crime scene to a certain point where the probability is maximal and then drops exponentially:

$$f(x) = \begin{cases} 1 + b(x - x_0) & \text{if } x \leq x_0 \\ e^{-c(x-x_0)} & \text{if } x > x_0. \end{cases}$$

This does support the idea of a buffer zone and has a quick drop-off after some maximal distance. In this model, we are able to specify the parameters b, c, and x_0 which may help when considering different types of criminals and their average found distance from crime scenes, or when considering different social profiles of criminals and the effects of such on distance.

Plateaued Negative Exponential

A special case of the truncated negative exponential occurs when we let $b = 0$. Instead of the usual rise-and-fall, the function stays level when $x \leq x_0$ and from there drops off like a regular negative exponential. The function takes the form

$$f(x) = \begin{cases} 1 & \text{if } x \leq x_0 \\ e^{-c(x-x_0)} & \text{if } x > x_0. \end{cases}$$

This supports the idea that, although there exists a buffer zone around the crime scenes where the killer is less likely to live, the area spanned when moving outwards from the crime scene squares and so it is innate in the spread from the crime scene that the probability is higher as distance increases when maintaining a constant probability at each grid point. It considers that the buffer zone is not derived from criminals preferring crimes past a certain distance but rather that opportunities increase with distance since the area increases.

5   Our Approach

5.1   Predicting Criminal Location

To predict the criminal's location, we employ a strategy that, while similar in some superficial ways, greatly generalizes the strategy of locating the criminal by investigating the center of mass. Whereas the center of mass approach returns a single point

(around which one can center a Gaussian distribution, for instance), our method attempts to utilize the locations of the criminal acts to the fullest by instead generating a distribution, assigning each point within the given region a probability. These probabilities returned can in turn be used to generate a map, showing which locations the algorithm considers to be of high interest and which can probably be ignored. By transforming the set of crime data points into a specially-constructed distribution rather than a handful of values, our algorithm allows for much more specialized, and in many cases much more accurate, results.

The main idea driving the algorithm is the idea that any point p can be assigned a relative "criminal location probability" by taking the sum of some function (mentioned in Metrics and Functions) of the distances between p and prior incidents. It might help to imagine this as the sum of non-canceling "gravitational pulls" of the various locations: any point with a comparatively high total gravitational pull is going to be near many of the crime scenes, and therefore should rightly be assigned a high criminal location probability. Furthermore, the actual strengths of the pulls from each point can be weighted depending on its recency, allowing for the algorithm to more quickly adapt to the changing habits of the criminal. We will currently treat all weights as equal to one, but we will discuss weighting in more detail in Time Coupling.
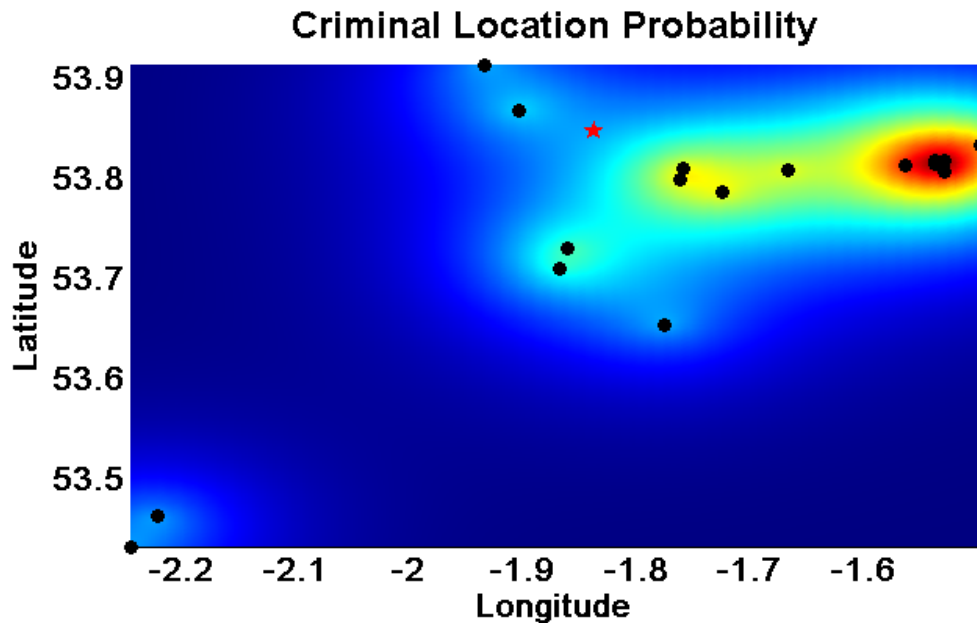

Figure 2: Example of a probability distribution. Red star signifies actual criminal's base.

5.1.1 Computation

As it is infeasible to calculate the value at every single point and extremely difficult to model analytically, our algorithm instead discretizes the problem by transforming the candidate search area into a grid and calculating the criminal location probability at each grid point. When increasing the fineness of the grid, which is allowed for in the input to our model, our algorithm better approximates the ideal, smooth, solution. Because the algorithm makes l calculations of f per grid point, it has $O(nml)$ runtime where n is the number of lines on the grid parallel to the Prime Meridian, m is the number of lines parallel to the equator, and l is the number of unlawful acts perpetrated by the criminal. As l is generally relatively small and, at least in our simulations, we often set m and n to be equal, this algorithm can essentially be considered to be $O(n^2)$, which is going to be very fast for any reasonable mesh size — consider that a single 2GHz machine can use this algorithm to partition an area the size of England into 36×36 meter blocks and create a corresponding probability distribution using some interesting distance functions in only a matter of minutes!

5.1.2 Pseudocode

To clear up any ambiguities, we provide some pseudocode below for the criminal-finding function.

Program 1 Criminal Location Finding Algorithm

G is the set of grid points after the search area is partitioned by the mesh

C is the set of past crimes, each of which has a location and date

$f(d)$ is the distance decay function which returns a higher number for more probable locations

Prob is the probability of the killer location based on distance decay affects for the grid

---

Function criminalLocationProb (crimes C, function f):

    Let Prob be a mapping from G into $\mathbb{R}$

    for $g \in G$ **do**

        /*Calculate cumulative distance effect from all $c \in C$*/

        $\text{Prob}[g] = 0$

        for $c \in C$ **do**

            $\text{Prob}[g] = \text{Prob}[g] + f(\text{distance between } c \text{ and } g)$

    Normalize the values of Prob so that the sum of the values is 1

   **return** Prob

---

5.1.3 Why it Works

Our algorithm has many properties that lead to the generation of useful results.

Prioritized Search

The probability distribution that our algorithm returns lends to a very simple algorithm for determining in what order to search various locations for the criminal. One can simply search the possible locations in descending order of their probabilities! This simple strategy often leads to good results, requiring us to go through about 15% of a certain search area to find Peter Sutcliffe, one of our case study serial killers, for most distance decay functions.

Point Clustering

One of the nicest features of our algorithm is its handling of point clusters without any outside assistance. Because all crime scenes exert a "gravitational pull," points near any crime scenes will have a probability boost and will create localized "hot-spots" where the algorithm thinks the criminal has some reasonable chance of being caught. This effect can be seen on the figure on the left/right, in which two different clumps form two hotspots, with their relative sizes and magnitudes proportional to the number of points inside.
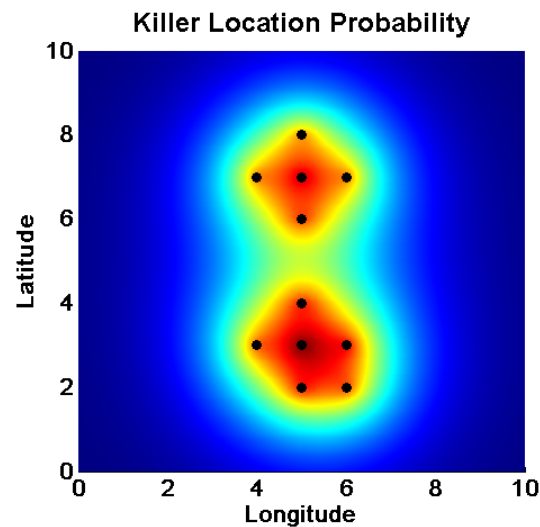
Function/Parameter Independence

Another major strength of our algorithm is its generality. Because the function works on any reasonable arbitrary function with arbitrary parameters, the police can hand-pick the functions based on statistics and their historical successes in catching the perpetrators of various types of crimes. For example, because burglars and serial killers generally exhibit different behaviors in selecting how far they are willing to travel to commit their crime, it could be that a normal distance decay function would be superior for catching burglars whereas a negative exponential distance decay function could be superior for catching serial killers. As our algorithm is extremely flexible, the police can quickly and easily adapt their search techniques to best match the criminal in question.

5.2 Predicting Location of the Next Crime

5.2.1 Method

In order to find the location of the next crime, we look for the point on the grid which most closely matches the pattern of the other crimes. If we consider the criminal location probability distribution produced after adding a crime event to our grid, the most probable location for the next crime would be the spot which generates a distribution most similar to the one produced before adding it. This gives us the most likely location for a future crime. Furthermore, for potential crime scenes other than the most probable, we estimate how much worse they are by measuring how much the produced distribution deviates from the original. This gives us a distribution over our search area which can be used to allocate resources for crime prevention.

Our algorithm works as follows: First we compute the criminal location probability distribution for the search area. Next we iterate through all of the grid points in the search area, adding a new "virtual crime scene" at each point, calculating the new criminal location probability distribution with the additional crime scene. We then compare each of these distributions to the original one by summing the squares of the differences between corresponding grid points, then assigning this pseudo probability deviation number to the location of the "virtual crime scene." This gives us a distribution over the search area, where lower numbers mean the point is a more likely spot for the criminal's next crime because adding this point as a crime location deviates less from the original probability distribution.

5.2.2 Edge Behavior

If all of the points on the mesh are tested as possible crime scenes, then there is a tendency for points near the edge to be rated higher than appropriate. This is because a point near the edge has fewer other points around it, so when a crime scene is added there it is effecting the criminal location probability distribution of fewer points, and thus has a smaller total effect on the produced distribution. To compensate for this, we add a buffer zone around our search area where the criminal location probability distribution is calculated but no crime scenes are added. This effectively removes the bad edge behavior.

5.2.3 Complexity

When adding each point as a virtual crime scene, we must calculate a different criminal location probability distribution for every point in our search area, thus requiring approximately n^2 operations, each of which is $O(n^2)$ (see Predicting Criminal Location - Computation). Therefore the algorithm for determining the location of the next crime has a complexity of $O(n^4)$. This is a polynomial time algorithm, but it can still become very slow as n increases. To keep our calculation time from getting out of control, we implement various optimizations.

Program 2 The killer-finding and future-predicting algorithm

G is the set of grid points after the search area is partitioned by the mesh

G' is G augmented with a 40% boundary buffer zone

C is the set of past crimes, each of which has a location and date

$f(d)$ is the distance decay function which returns a higher number for more probable locations

Prob is the probability of the killer location based on distance decay affects for the grid

ΔProb measures the change in distribution of Prob when adding a new crime scene

---

Function criminalLocationProb (crimes C, function f):

    Let Prob be a mapping from G' into $\mathbb{R}$

    for $g \in G'$ **do**

        /**Calculate cumulative distance affect from all $c \in C$*/

        Prob[g] = 0

        for $c \in C$ **do**

            Prob[g] = Prob[g] + f(distance between c and g)

    Normalize the values of Prob so that the sum of the values is 1

    return Prob

Function nextCrimeProb (crimes C, function f):

    killer_prob_distrib = criminalLocationProb(C,f)

    Let ΔProb be a mapping from G' into $\mathbb{R}$

    for $g \in G'$ **do**

        Let C' be C augmented with a "virtual crime" at g

        V = criminalLocationProb(C',f)

        /**Calculate sum of squared differences between killer_prob_distrib and V*/

        ΔProb[g] = 0

        for $h \in G'$ **do**

            ΔProb[g] = ΔProb[g] + (V[h] − killer_prob_distrib[h])^2

    Normalize the values of ΔProb so that the sum of the values is 1

    return ΔProb

(a) No buffer.   (b) 30% buffer.
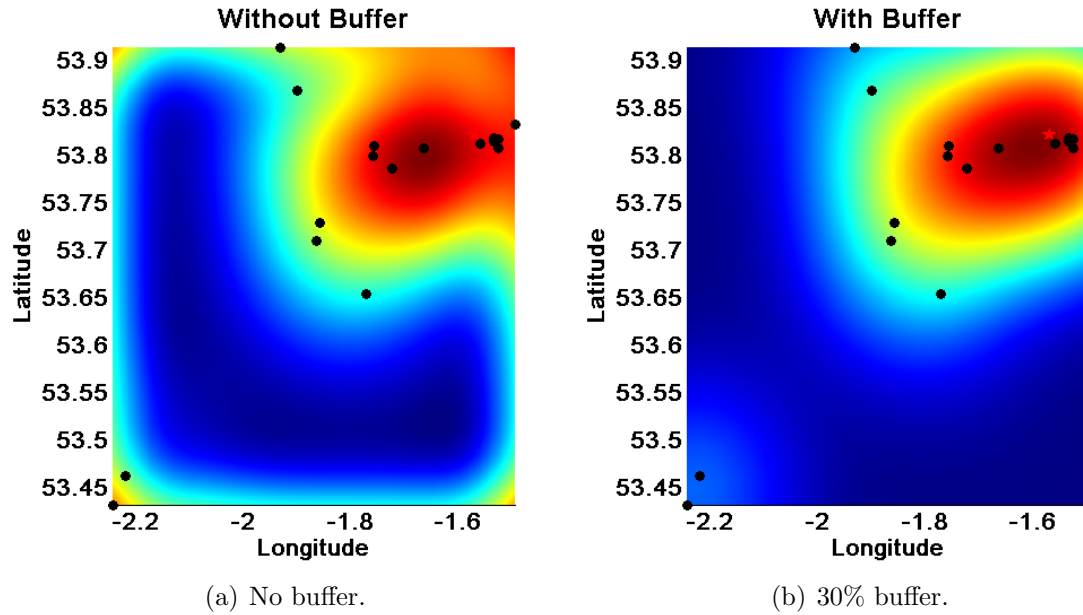
Figure 3: Varying buffer percentages to illustrate boundary conditions.

The single most important optimization implemented in our algorithm is caching. Suppose our search area has N crime scenes. For each point on the mesh we add a virtual crime scene and then calculate the criminal location probability distribution over the search area using the $N + 1$ points. When calculating the new probability distribution with $N+1$ points, instead of doing the full calculation, we use the original distribution for the first N points and combine it with a distribution for only the new point in a way that maintains normalization. This gives the algorithm a factor of N speedup.

Our algorithm is also nearly perfectly parallelizable. By multi-threading the code, it can be run over a number of cores with a speedup equal to the number of cores used. This makes running very high resolution calculations practical if a multicore or computing cluster is available.

5.3   Predicting Time of Future Crimes

5.3.1   Method

When considering a serial killer's next time or date of killing, we look at a basic linear regression model. We find the average time between killings to predict the next killing from the last. To test our model, we remove the last point from the N killing times which we attempt to predict using our regression:

$$t_N = t_{N-1} + \frac{t_{N-1} - t_1}{N-1} \, .$$

5.3.2 Motivation

Although a linear regression may not be realistic for every killer, it applies more generally to all as some may increase their frequency of killings as habituation sets in and some may de-

Figure 4: Predicted Nth crime in red; actual Nth crime in black.

crease their frequency. Hypothesized by Hickey's Trauma Control Model and verified by Arndt's analysis of Newton's *Hunting Humans* which is a collection of data of serial killers, the killing rate of serial killers generally increases with time; however, there are many sources which state that the killing rate may also decrease in some cases (Arndt et al., 2004). Because other interpolation techniques such as exponential or quadratic may greatly over-predict future crimes because of strong end-behavior, we choose to use the linear regression for its simplicity and universal niceness.

5.4 Time Coupling

Having developed methods for predicting both the criminal location and where the next crime will occur, we now extend these methods by adding in the concept of time. As described so far, the algorithm neglects all chronological properties of the killings, discarding one of the most important pieces of information available to us. Ideally, we want to be able to quickly adapt to the criminal's motion if they were to move to a different location or otherwise alter behavior over time.

To account for these possible changes, we weight events more heavily the more recently they occurred. Doing this allows for recent changes in behavior to be more strongly reflected in future predictions, which allows for better calibration with the criminal's latest motives. To keep the algorithm as general as before, we will modify the algorithm to take a weighting function as an input, which is then multiplied by $f(\text{dist})$ when we calculate $\text{Prob}[g]$.

Although the composition of the weighting function w can be anything, we notice that

$$w(t) = \frac{(d - c)a}{\sqrt{t} + a} + c$$

yields reasonable results. Due to time constraints, this w was the only reasonable weighting function which we managed to study in depth, and therefore any future

---

Program 3 The killer-finding and future-predicting algorithm with time-based weighting

---

G is the set of grid points after the search area is partitioned by the mesh

G' is G augmented with a 30% boundary buffer zone

C is the set of past crimes, each of which has a location and date

$f(d)$ is the distance decay function which returns a higher number for more probable locations

$w(t)$ is the weight function which returns a higher number for smaller inputs

Prob is the probability of the killer location based on distance decay affects for the grid

ΔProb measures the change in distribution of Prob when adding a new crime scene

---

Function criminalLocationProb (crimes C, decay_function f, weight_function w):
    Let Prob be a mapping from G' into $\mathbb{R}$
    for $g \in G'$ **do**
        /*Calculate cumulative distance effect from all $c \in C$*/
        $\text{Prob}[g] = 0$
        for $c \in C$ **do**
            $\text{Prob}[g] = \text{Prob}[g] + w(\text{time since } c_{\text{time}}) \cdot f(\text{distance between } c \text{ and } g)$
    Normalize the values of Prob so that the sum of the values is 1
    return Prob

Function futureCrimeProb (crimes C, decay_function f, weight_function w):
    killer_prob_distrib = criminalLocationProb(C,f,w)
    Let ΔProb be a mapping from G' into $\mathbb{R}$

    for $g \in G'$ **do**
        Let C' be C augmented with a "virtual crime" at g
        V = criminalLocationProb(C',f,w)

        /*Calculate sum of squared differences between killer_prob_distrib and V*/
        $\Delta\text{Prob}[g] = 0$
        for $h \in G'$ **do**
            $\Delta\text{Prob}[g] = \Delta\text{Prob}[g] + (V[h] - \text{killer\_prob\_distrib}[h])^2$

    Normalize the values of ΔProb so that the sum of the values is 1
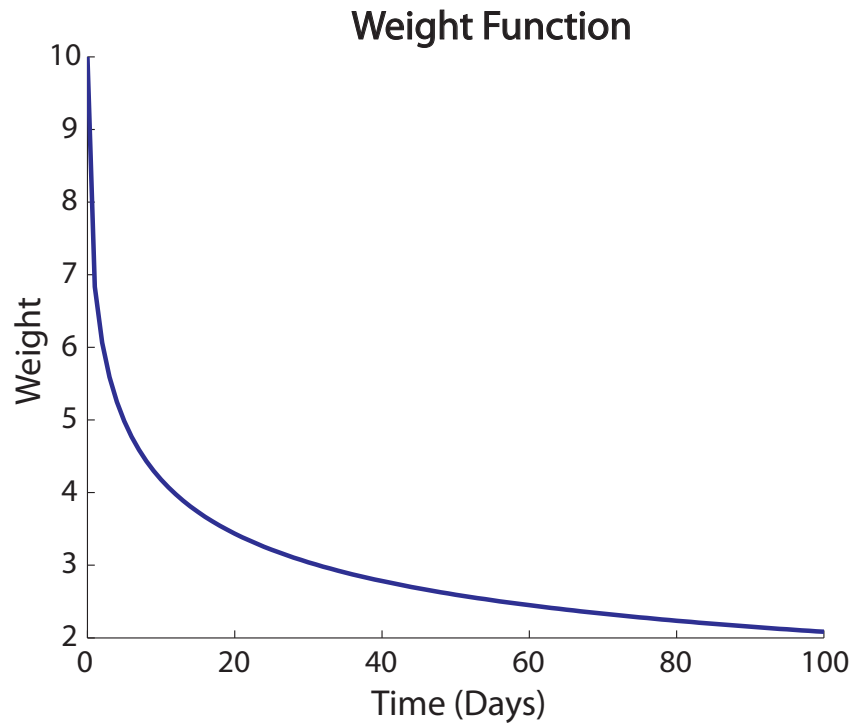   **return** ΔProb

---

Figure 5: A decay plot.

time-based models use this w as the time weighting function.

5.4.1 Why It Helps

To help emphasize the benefits of time coupling, we consider a randomly generated scenario. The points in the following graphs were automatically generated to be close-to-linear, and some randomly selected set of times generated by a Poisson process were matched to these points in one of two ways:

- In the "correlated" case, the times were chosen to increase monotonically with x. That is, the larger the value of x, the larger the time that was assigned to it.

- In the "uncorrelated" case, the various times were assigned randomly to the points.

Notice how the addition of time-based weighting in the uncorrelated case had almost no effect when compared to the correlated case, in which the effect of the weighting was pronounced. As there was no correlation in the former case, any small perturbations will essentially cancel each other out, leaving each point with no expected
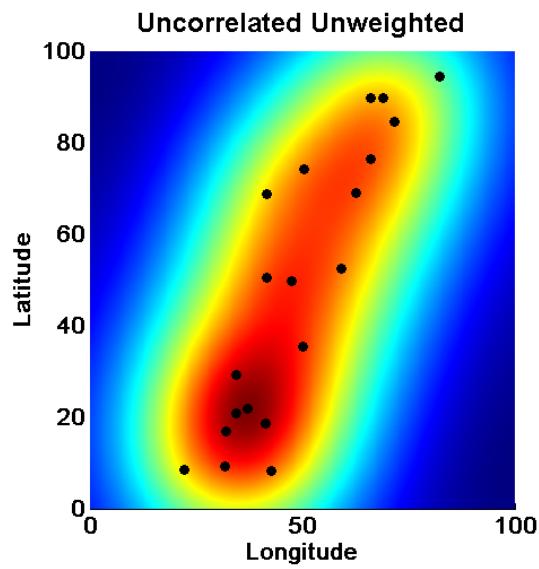
(a) Next Crime Probability Plot with unweighted time, uncorrelated points.
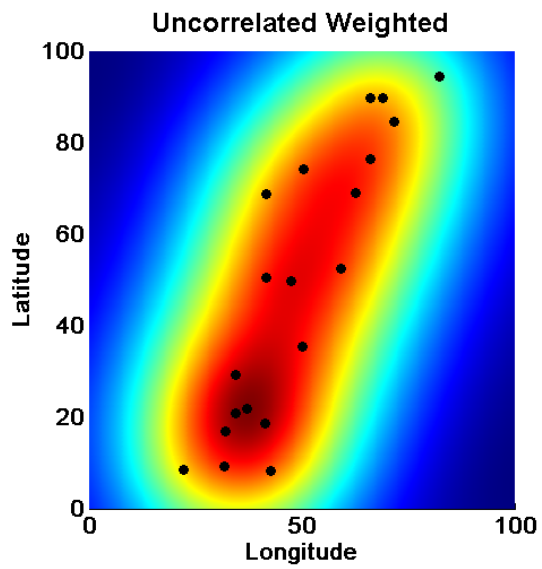
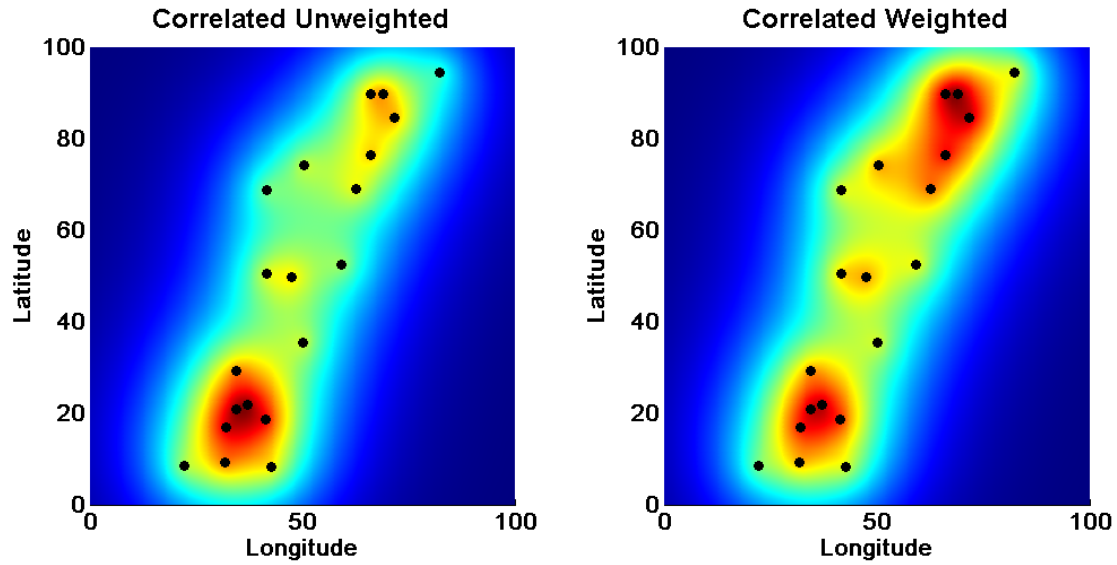(b) Next Crime Probability Plot with weighted time, uncorrelated points.
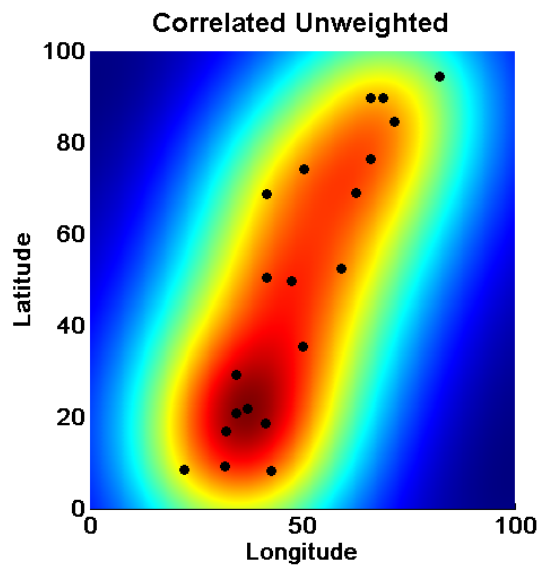
(c) Next Crime Probability Plot with unweighted time, correlated points.
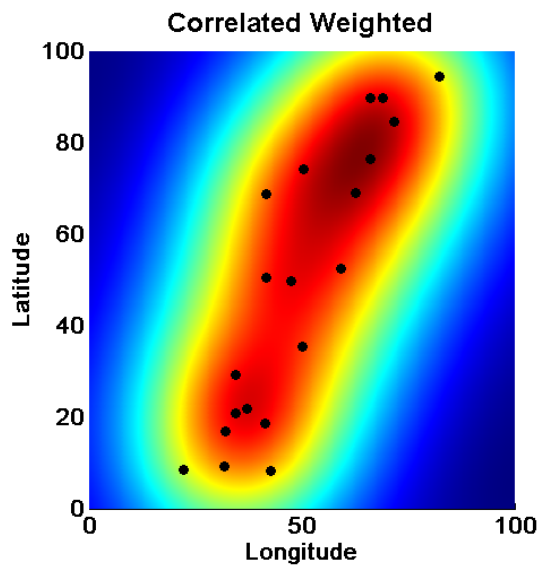
(d) Next Crime Probability Plot with weighted time, correlated points.

(e) Criminal Location Probability Plot with unweighted time, correlated points.

(f) Criminal Location Probability Plot with weighted time, correlated points.

change. Thus, the correctness of the plot is not drastically affected, if at all. In the latter case, however, the correlation was captured by the algorithm and the expected location of the next killing shifted rather drastically to accommodate this change. Intuitively, if a serial killer has been moving north one mile every day for a while, it is reasonable to assume that his or her future kills will be more north than average.

Perhaps even more interesting, the location of the next crime adapted much more quickly than the estimated location of the criminal. Again, this is sensible: a gradual shift in the location of crimes is much more indicative of a behavior-based shift than one that is residence-based.

6 Result Discussion

To evaluate our models, including various choices for our distance decay function, we look at case study data (see Appendix A) for actual apprehended serial killers: Peter Sutcliffe and Chester Turner.

6.1 Metrics

We use two metrics in quantifying the quality of our results. The first is Error Distance and the second is Search Cost.

Error Distance

This mesaures the distance between the calculated most probable spot and the actual spot "as the crow flies". When searching for the criminal, the error distance tells us how far away our best guess was from the actual location. When predicting the next crime it tells us how far away the crime happened from the location where we allocated the most resources.

Search Cost

This measures how many grid squares would need to be searched out of the total before the correct spot is found if we go through the squares in the order of their probability. When searching for the criminal this tells us how much area would need to be searched to find the criminal. We assume no preference over locations of the same color and thus this may vary in actual police searching where the initial search direction may vary. When predicting the location of the next crime it tells us what percentage of our resources were wasted on locations that had higher priority than the actual location but saw no actual crime. This metric provides a much more realistic assessment of the quality of a prediction; however, in some instances the Error Distance still provides useful information.

6.2 Procedure

We use the model without the time-dependent weighting first. Thus for these calculations the times of the killings are irrelevant to the calculated probability distributions for the criminal's location and the location of his next crime; however, the times are used to calculate when the next crime will occur.

We begin by removing the most recent crime from our datasets. This emulates the position the police would be in while searching for the criminal. Next we calculate the criminal location probability distribution using each of our decay functions. The quality of these calculations is assesed using both of our metrics, comparing the predicted locations to the actual known locations of the killers' home. Then we calculate the next crime location probability distribution for each decay function. These results are also assesed using the metrics. Finally, we repeat the calculations using the model with time dependent weights. The results is presented below. Plots can be found in the Appedix.

Method	Criminal Home Uncoupled	Criminal Home Coupled	Next Crime Uncoupled	Next Crime Coupled
Sutcliffe				
Linear	18.17/17.88	19.26/29.49	3.99/9.47	3.41/8.73
Neg. Exp.	15.97/30.24	17.05/30.24	0.27/1.46	0.26/1.46
Trunc. Neg. Exp.	14.44/24.32	15.53/24.32	1.98/8.00	1.65/7.28
Plateau Neg. Exp.	15.00/25.07	16.33/25.07	1.25/5.70	1.04/4.98
Normal	15.00/22.83	15.99/22.92	0.81/4.27	0.63/3.58
Chester Turner				
Linear	58.64/5.52	60.12/5.92	40.90/4.12	44.46/5.10
Neg. Exp.	59.80/5.07	60.79/5.07	41.06/3.12	45.93/3.26
Trunc. Neg. Exp.	59.80/4.38	60.79/4.38	40.68/3.46	45.26/4.12
Plateau Neg. Exp.	59.80/4.48	60.79/4.48	40.79/3.26	45.58/3.46
Normal	66.04/4.46	66.70/4.46	40.83/3.26	43.38/3.46

Table 1: Search Costs / Error Distance using various functions for both the Sutcliffe and Chester Turner data.

7 Conclusion

All examined distance decay functions tend to do a good job when examining "good" data, which is data in which the next kill or killer location tends to be in the general vicinity of the killings. In the case of Peter Sutcliffe, our algorithm predicted Sutcliffe's location rather accurately on any distance decay function and guessed the location of the next crime *almost exactly* in many cases.

Unfortunately, the data can be misleading. In the case of Chester Turner, who exclusively committed murders to the east of his home, the algorithm unsurprisingly failed to guess his home address accurately. This is unavoidable: there is no method (as far as we can tell) to accurately guess the killer's "hub" location if it is sufficiently removed from the locations of the crimes themselves.

In either case, it seems that the guessing accuracy for the next crime location is much better than that for the location of the criminal. While this result is somewhat surprising, as the guess for the next crime location is itself based on the estimated killer location, it is still possible to explain. One simple explanation which can be visually verified of this phenomenon is that killers tended to cluster their killings closely, allowing for easier guessing for the next crime. Another possible explanation is that the increased number of calculations required to generate the estimated next location resulted in significantly smaller random fluctuations, meaning that the data is "smoother" and therefore more likely to behave well.

Contrary to our expectations, the addition of time coupling generally, but not always, *decreased* our performance. While we don't currently have any explanation much better than "bad luck" for this behavior, this result is something that should

definitely be examined before serious use of this tool by the police. It seems that time coupling only gives superior performance when the spatial and temporal aspects of the previous crimes is correlated. Without time coupling, our model misses this pattern in the data, but the time coupled version finds it.

Even with the model's flaws, we believe that it can be a useful tool for any police station to employ. If the model managed to give us such accurate data for predicting the whereabouts of Sutcliffe's next victim, it is not unreasonable to believe that the tool could repeat this behavior for larger datasets. With this predictive power in hand, the police can better prepare for upcoming attacks and apprehend the criminal.

8   Future Research

8.1   Incorporating Statistics and Landscapes

Although our model avoids human bias toward overconfidence and predisposed predictions, one might argue "information that a person, animal or institution knows about a physical or social environment" (Gigerenzer & Selton, 2001, 187) gives humans the benefit for predicting locations; however, we would like to incorporate that information into our model: the landscape, personality profile, population density. Eventually, one may be able to also specify social personality regions - for example, rich areas versus poor, known drug-trafficking areas, etc. We believe this augmentation would not only lead to better predictions by our model but also give it a clearer advantage over human predictors. In such cases as we are considering, it is not immediately obvious where the serial criminal lives or will commit their next crime. Thus, we support probabilities and statistics against human bias.

Different personality types of criminals (e.g., team killers, sexually motivated killers, burglars, etc.) also commit crimes at different distances from their base. Team killers are less likely to kill women and live in a more geographically localized area; sexually motivated killers are more likely to use a more personal killing method, such as strangulation, and prefer strangers more, but tended to be more geographically stable as compared to non-sexually motivated killers (Arndt et al., 2004). These personality inferences may be made by surveying crime scenes. In our model, therefore, we could have optional inputs for use by police to specify personality traits, which we could then use to alter our parameters automatically to predict more probable distances with more ease for the police.

Serial Killer	Predicted Crime	Actual Crime
Sutcliffe	09-Dec-1980	17-Nov-1980
Zodiac Killer	30-Dec-1969	11-Oct-1969
Jack the Ripper	07-Oct-1888	09-Nov-1888
Grim Sleeper	24-Feb-2005	01-Jan-2007
Chester Turner	01-Jan-1999	06-Apr-1998

Appendix A

Serial Killer Case Study Data

| Serial Killer	Latitude	Longitude	Date of Killing
Sutcliffe	53°49'22.72"N	1°34'38.03"W	November 17 1980, 9:25 pm
	53°48'30.95"N	1°40'18.26"W	August 20 1980, 11:00 pm
	53°48'47.58"N	1°34'3.23"W	September 2 1979, 1:00 am
	53°42'40.28"N	1°52'18.99"W	April 4 1979, 11:55 pm
	53°27'46.12"N	2°13'35.21"W	May 16 1978, 11:00 pm
	53°39'17.04"N	1°46'46.20"W	January 31 1978, 9:25 pm
	53°47'57.94"N	1°45'59.09"W	January 21 1978, 9:30 pm
	53°25'57.94"N	2°15'2.47"W	October 1 1977, 9:30 pm
	53°49'4.56"N	1°31'58.99"W	June 26 1977, 2:15 am
	53°48'38.70"N	1°45'49.79"W	April 23 1977, 11:15 pm
	53°50'1.26"N	1°30'9.15"W	February 5 1977, 11:30 pm
	53°48'28.75"N	1°31'58.58"W	January 20 1976, 7:30 pm
	53°52'7.83"N	1°54'28.10"W	July 5 1975, 1:30 am
	53°49'7.15"N	1°32'31.71"W	October 30 1975, 1:30 am
	53°48'55.72"N	1°32'28.77"W	December 14 1977, 8:30 pm
	53°47'12.39"N	1°43'47.22"W	July 10 1977, 3:20 am
	53°43'48.65"N	1°51'54.77"W	August 15 1975, 11:45 pm
	53°54'52.78"N	1°56'16.61"W	August 27 1975, 10:30 pm
Chester Turner	33°56'36.66"N	118°16'50.04"W	March 9, 1987
	33°56'24.32"N	118°16'49.98"W	October 29 1987
	33°56'49.26"N	118°16'57.66"W	January 20 1989
	33°57'23.37"N	118°16'57.85"W	September 23 1989
	33°56'50.59"N	118°16'50.55"W	September 30 1992
	33°56'50.59"N	118°16'50.55"W	November 16 1992
	33°56'53.00"N	118°16'57.77"W	December 16 1992
	33°58'7.02"N	118°16'57.78"W	April 2 1993
	33°58'40.96"N	118°17'5.57"W	May 16 1993
	33°58'0.20"N	118°16'59.76"W	February 12 1995
	33°56'56.43"N	118°16'42.38"W	November 6 1996
	34° 2'56.29"N	118°15'20.32"W	February 3 1998
	34° 2'31.22"N	118°14'26.07"W	April 6 1998
Zodiac Killer	38° 5'49.94"N	122° 8'59.18"W	December 20 1968, 11:15 pm
	38° 7'12.90"N	122°11'30.51"W	July 4 1969, 11:55 pm
	38°34'9.57"N	122°14'7.71"W	September 27 1969, 6:15 pm
	37°47'19.34"N	122°27'25.98"W	October 11, 1969, 9:55 pm
Grim Sleeper	34° 0'17.48"N	118°19'17.57"W	August 12 1986
	34° 0'14.98"N	118°18'35.63"W	September 11 1988
	33°59'37.53"N	118°15'4.83"W	January 10 1987
	33°58'41.68"N	118°17'33.75"W	August 10 1985
	33°58'32.96"N	118°18'7.76"W	August 14 1986
	33°57'54.10"N	118°19'0.30"W	March 9 2002
	33°57'27.36"N	118°18'31.85"W	November 1 1987
	33°57'7.36"N	118°18'31.24"W	April 16 1987
	33°56'50.31"N	118°18'30.33"W	January 1 2007
	33°57'17.11"N	118°17'55.40"W	November 20 1988
	33°56'41.05"N	118°19'2.66"W	January 30 1988
	33°56'19.68"N	118°18'15.39"W	July 11 2003
Jack the Ripper	51°31'12.17"N	0° 3'37.34"W	August 31 1988
	51°31'13.51"N	0° 4'20.70"W	September 8 1988
	51°30'49.90"N	0° 3'57.16"W	September 30 1988
	51°30'49.13"N	0° 4'39.64"W	September 30 1988
	51°31'7.29"N	0° 4'30.32"W	November 9 1988

Data collected via GoogleEarth™ and CommunityWalk.

Model Plots for Sutcliffe

Here we show our presented functions for distance decay with their plots for Sutcliffe and Chester Turner. These are here to illustrate the differences between the different functions.
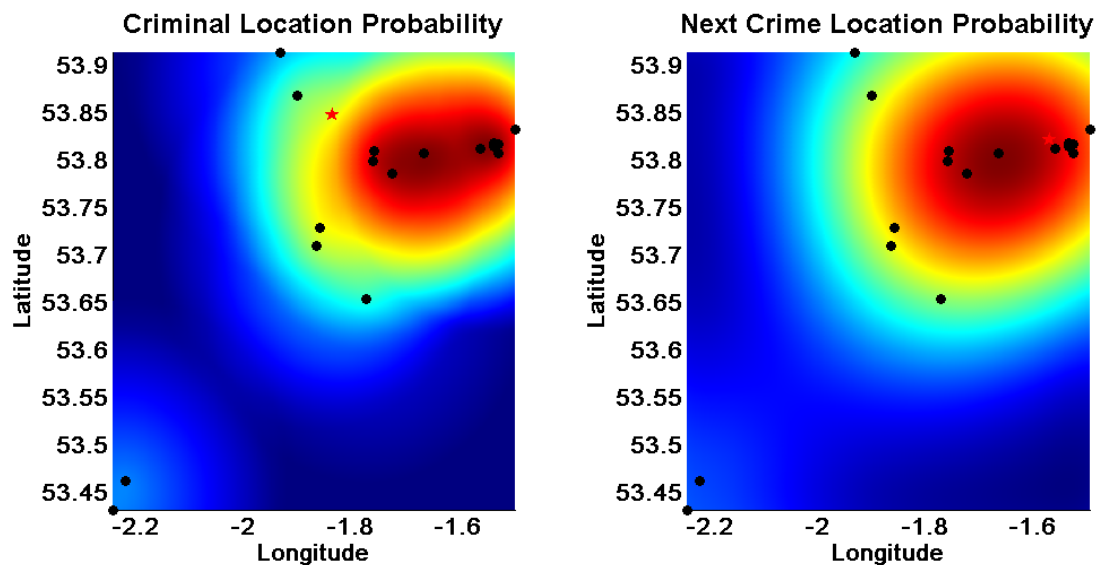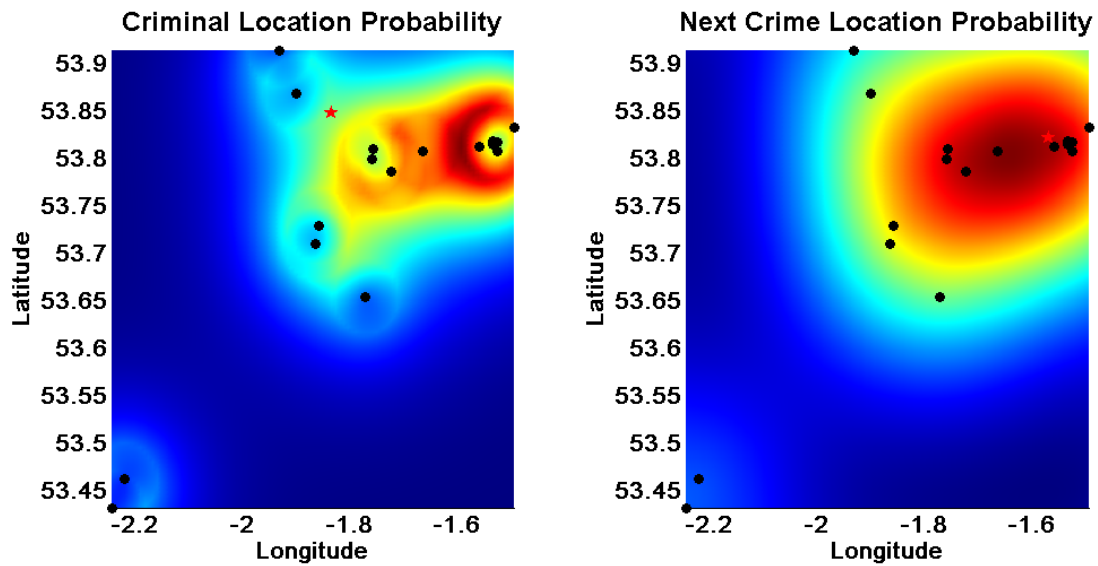

(g) Criminal Location Probability Plot of Sutcliffe using Linear decay function.

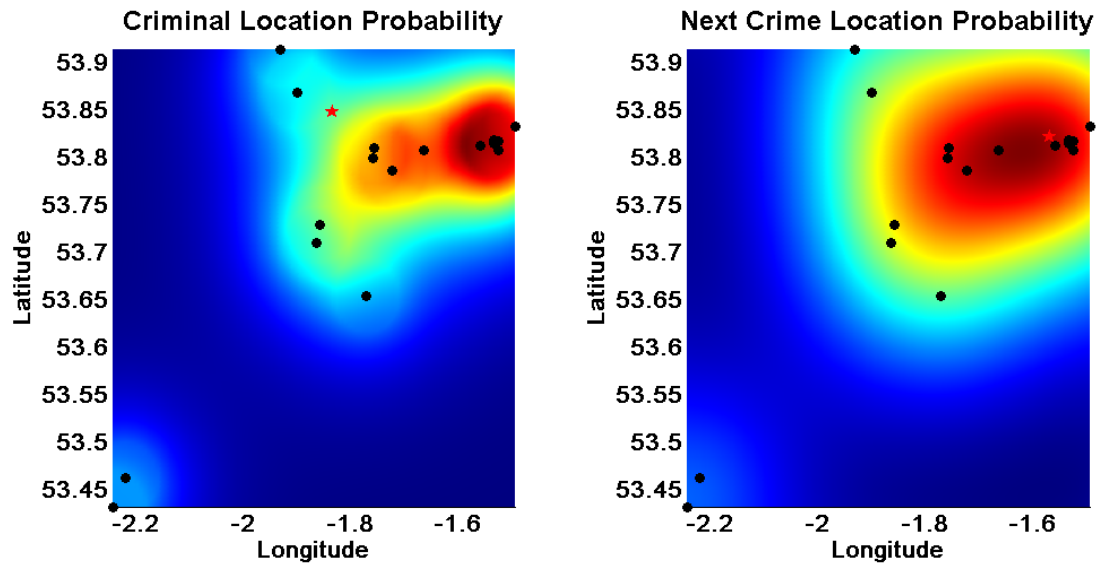(h) Next Crime Probability Plot of Sutcliffe using Linear decay function.
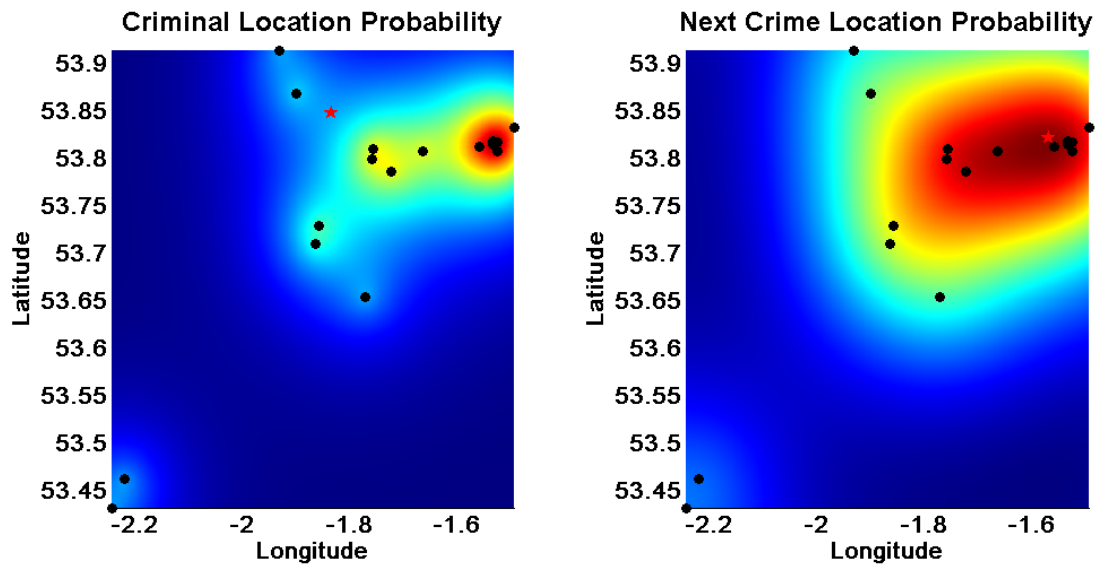
(i) Criminal Location Probability Plot of Sutcliffe using Negative Exponential decay function.

(j) Next Crime Probability Plot of Sutcliffe using Negative Exponential decay function.
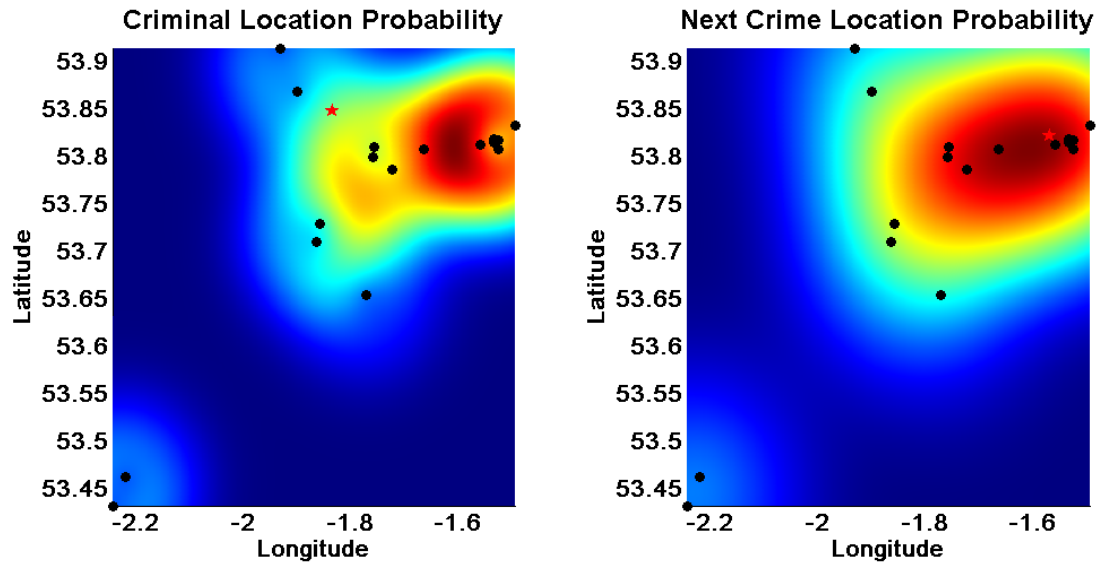

(k) Criminal Location Probability Plot of Sutcliffe using Truncated Negative Exponential decay function.

(l) Criminal Location Probability Plot of Sutcliffe using Truncated Negative Exponential decay function.
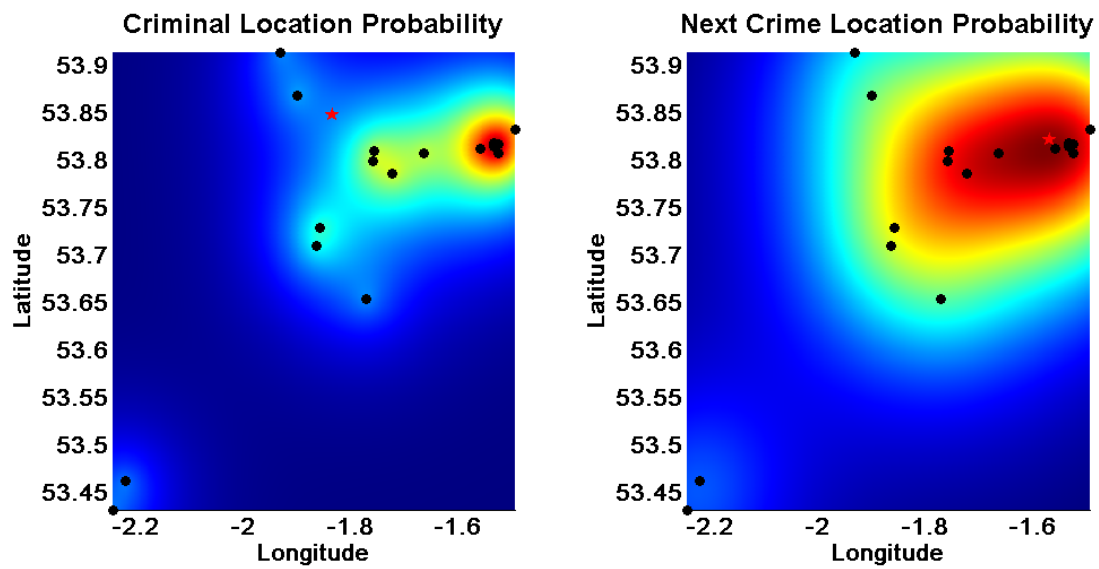
(m) Criminal Location Probability Plot of Sutcliffe using Plateaud Negative Exponential decay function.

(n) Next Crime Probability Plot of Sutcliffe using Plateaued Negative Exponential decay function.

(o) Criminal Location Probability Plot of Sutcliffe using Normal decay function.

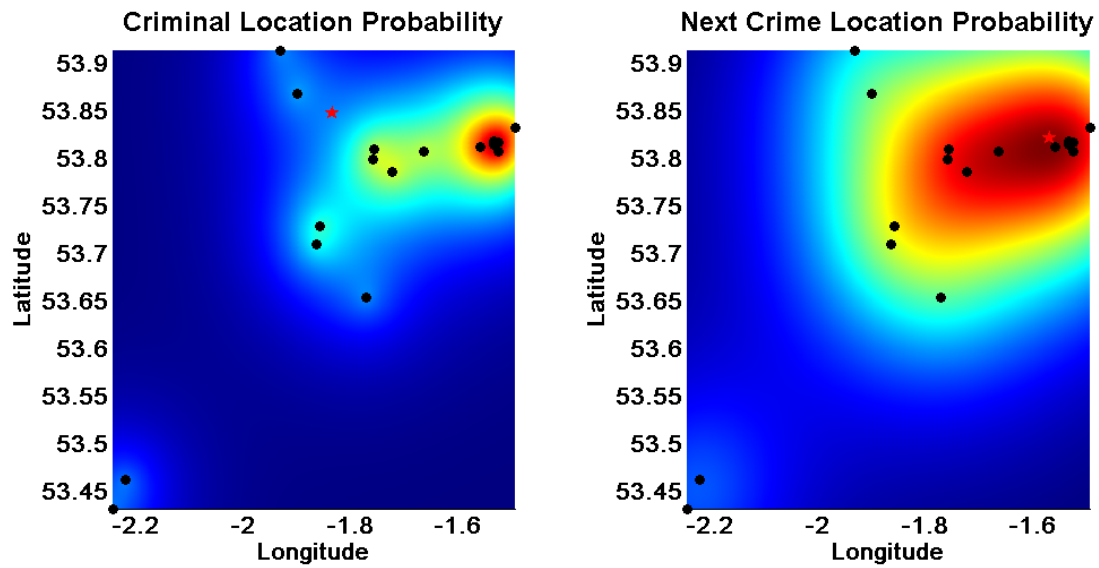(p) Next Crime Probability Plot of Sutcliffe using Normal function.

(q) Criminal Location Probability Plot of Sutcliffe using Negative Exponential decay function, time weighted.

(r) Next Crime Probability Plot of Sutcliffe using Negative Exponential function, time weighted.

Model Plots for Chester Turner

(s) Criminal Location Probability Plot of Sutcliffe using Negative Exponential decay function, time weighted.

(t) Next Crime Probability Plot of Sutcliffe using Negative Exponential function, time weighted.

References

Arndt, W. B., Hietpas, T., & Kim, J. (2004). Critical characteristics of male serial murderers. *American Journal of Criminal Justice*, *29*(1), 171–131.

Brantingham, P., & Brantingham, P. (1981). *Environmental Criminology*. Beverly Hills, CA: Sage Publications.

Gigerenzer, G., & Selton, R. (2001). *Bounded rationality*. London: MIT Press.

Hickey, E. W. (2002). *Serial murderers and their victims*. Belmont, CA: Wadsworth, 3 ed.

Kent, J. D. (2003). *Using Functional Distance Measures When Calibrating Journey-to-Crime Distance Decay Algorithms*. Ph.D. thesis, Louisiana State University.

Laukkanen, M., & Santtila, P. (2006). Predicting the residential location of a serial commercial robber. *Forensic Science International*, *157*, 71–82.

Levine, N. (2006). Crime mapping and the crimestat program. *Geographical Analysis*, *38*(1), 41–56.

Newton, M. (1990). *Hunting Humans*. Port Townsend, WA: Loompanics Unlimited.

Rhodes, W., & Conly, C. (1981). *Crime and Mobility: An Empirical Study*. Environmental Criminology. Prospect Heights, IL: Waveland Press, Inc.

Santtila, P., Laukkanen, M., & Zappala, A. (2007). Crime behaviors and distance travelled in homocides and rapes. *Journal of Investigative Psychology and Offender Profiling*, *4*, 1–15.

Snook, B., Taylor, P. J., & Bennell, C. (2004). Geographic profiling: The fast, frugal, and accurate way. *Applied Cognitive Psychology*, *18*, 105–121.