The Hunt for Serial Criminals

Team 8449

Contents

0. Abstract **2**

1. Executive Summary **3**

2. Statement of the problem and approach – Hunting Serial Criminals **5**

2.1 Survey of Previous Research : Environmental Criminology 6

2.2 Assumptions 6

2.3 Propositions and Foundation 8

3. Methods **11**

3.1 Construction of the Map – Geographical Method 11

3.2 Static and Dynamic – Risk Intensity Method 18

4. Simulation Results and Discussion **30**

4.1 Results of the Geographical Method 31

4.2 Results of the Risk Intensity, “Static” and “Dynamic”, Method 36

4.3 Discussion 37

4.3.1 Sensitivity and Robustness Testing 37

4.3.2 Accuracy of the Prediction 40

4.3.3 Combination of the Two Methods 40

5. Strength and Weakness of the Model **41**

6. Conclusion and Recommendation **41**

7. Appendices **43**

A Bibliography 43

B Data 45

C Code 46

0. Abstract

The advent of computer and technological progress has introduced a new stage in the development of criminology. Investigators can now use computational techniques of geographic profiling in order to determine the patterns of movement of their suspects.

We propose a model that aims to predict areas with high probability of being the next on the criminal's target list. We have assumed that serial crimes are instrumental rather than expressive, thus ensuring that the criminal follows a predictable pattern of movement. We also assume that this pattern is characterized by a certain stability and continuity which facilitates a correlation with the actions of other criminals in that area.

Our model first uses an initial "geographical method" which reduces the areas under consideration based on parameters such as location coordinates, area, population and criminal rate, as well as the history and psychological value variables which are derived from a specific criminal pattern. This input is used to determine the shape of a Gaussian 2D function showing the distribution of the areas with the highest probability of becoming the location of a future crime. We improve these results by using the risk intensity method, a combination of two schemes, a "static" and a "dynamic" one. The static method consists of first generating the risk intensities of different locations based on variables such as crime rates and distances from the anchor point, by using tools such as the distance-decay function. We then assign crime coefficients, which indicate the extent to which the crime can be categorized as murder, rape, arson or robbery. In the dynamic model, we categorize the static parameters into homotypic, heterotypic and cumulative types by computing the mean and covariance matrix of these parameters. We apply different algorithms: logistic regression, linear regression and nearest neighbor algorithm respectively to these types and then weigh them differently to obtain a parameter probability. This is then combined with the results of the static process to generate the probability of a crime at a certain location.

We tested our model using examples from different categories of serial crimes: robbery, murder, arson, which demonstrated distinct criminal patterns. The surfaces generated using the geographic method and the final predicted probabilities generally agreed with our expectations of areas where the criminal will attack again. The test for sensitivity suggested that parameters such as crime rate or population density (area and location) are well taken into consideration by our model. Small changes in location, however, affected to a significant extent our results, probably because the differences in coordinates of the locations were not large to begin with.

By analyzing the accuracy of our results, we conclude that this model is an efficient way of minimizing the range of possible crime locations. By taking into account all the variables that can be quantified, the geographical and the risk intensity methods achieve their goal of assigning probabilities to high-risk areas. However, evidence such as similarities of a criminal's victims was not taken into consideration. Furthermore, the model can only be applied to criminals who observe a predictable pattern of movement, in spite of the randomness parameters introduced in the model.

1. Executive Summary

The model we propose proves to be a generally efficient tool in criminal investigation. Once a sequence of crimes attributed to the same suspect is given, the model offers a good estimate of the possible areas where the next crime might be committed. This area is determined using an initial geographic method which indicates high-risk areas. The second part of the model (risk-intensity method) considers both static variables such as crime rate or population density of the location and variables characteristic to the movement pattern of that particular serial criminal. The areas are eventually assigned probabilities based on all the relevant parameters. The area with the higher probability is the one where the future crime is most likely to occur.

The tests performed to check that the model works are a good indication of its efficiency. The model was used to predict future crime locations for different serial crime categories, such as murder, arson and robbery. Therefore, the model could be used in the future for any type of crime of serial nature. Moreover, the tests covered spatial patterns ranging from offenses occurring in multiple states to crimes happening in different and not necessarily neighbor counties. A potential limitation of the model could be serial crimes committed in different countries, since this possibility was excluded in the assumptions.

The model was also tested for its response to small changes in parameters. The result is that these parameters are well taken into account in predicting the place where the next crime is committed. However, law enforcement officers might want to consider other pieces of evidence such as characteristics of the victims which could suggest the places where the criminal prefers to attack. Our model does not take into account such aspects, but it covers all the variables that can be quantified.

One situation in which this model cannot output accurate prediction is the expressive serial crimes, which are defined as more spontaneous and emotional than instrumental crimes. It is not surprising that these crimes are more difficult to predict. Under such conditions, the prediction made by the model is either too broad or cannot fully cover the high-risk area. Therefore, when we deal with expressive serial criminals, we need to consider more parameters.

Some degree of randomness already exists in the model, however we do not expect it to be able to predict the actions of a disorganized serial criminal. The model might not be able to even constrain the range of locations to be considered.

We conclude that our model could be a great aid to law enforcement officers in their criminal investigations. Given that the assumptions stated above are satisfied, the prediction provided by the model could limit the range of locations from 5 locations to 2 or 3 areas by using a geographical and risk intensity method. Other pieces of evidence could further give a more exact estimate of the future crime location, thus simplifying the work of law enforcement officers to a considerable extent.

2. Statement of the problem and approach – Hunting Serial Criminals

Criminology faces a difficult task in today's world. Not only does this scientific study have to control crime rates, but it also has to deal with serial criminals. This is a category of offenders who generate tremendous fear in numerous communities and require significant resources and effort from police, courts and prisons. In order to do that, criminology developed a series of methods called *geographic profiling,* which uses the locations of a sequence of crimes in order to determine what the most probable area of residence of the offender might be. The process of crime analysis has used the traditional method of pin mapping for more than a century. However, since the early 1990s, the increase in the speed of computers has allowed more and more police departments and institutions to make use of hardware and software that is the basis of crime mapping nowadays [1].

While the residence of a serial criminal is important in criminal investigations, a potentially more crucial prediction would be that of the possible location of the next crime. The problem here lies in identifying the spatial patterns of serial criminals and has been the focus of many research studies throughout the time.

The results of most of these studies agree on the fact that humans seem to follow predictable patterns of movements. Each individual possesses an awareness space, which includes their home, work, school, shopping areas or the commuting routes between these points. This space contains, for most people, an anchor point, which represents the most important place in a person's spatial life. Research studies have shown that this anchor point is, for the vast majority of people, their residence. It is thus not surprising that, in generating a geographical profile, we keep in mind the offender's anchor point as a point of interest with a potential influence on the prediction of the next crime's location [2].

Serial crime is usually defined as crime of a repetitive or serial nature and can be one of the following: serial murder, serial rape, serial arson or even serial robbery. While these offenses share characteristics such as the extent to which they affect communities, they prove to be sometimes different in the behavioral profile of the criminal. It can then be inferred that different types of serial crime may at times generate different geographical profiles in terms of the residence or activity space of a criminal. For example, locations with high population density prove to be more prone to property crimes such as serial robberies. On the other hand, rapists seem to prefer both low-density areas, which are characterized by less surveillance and high-density areas, where they have a bigger chance of finding a suitable target.

2.1 Survey of Previous Research : Environmental criminology

1. Studies in this field focus on spatial patterns in offender and target movement on the basis of broader social routines. This theory has first been introduced by professors Paula and Patricia Brantingham at Simon Fraser University in the 1980s and has since shifted in a way the focus of the study of criminology to the environment and spatial patterns that influence criminal activity [1]. Their work has dealt into trying to predict both the most important place in a criminal's life (the anchor point) and the most vulnerable "hot spots" where the next crime could take place.
2. The geographic profiling technology was developed by Dr. Kim Rossmo at Simon Fraser University based on the theory of environmental criminology proposed by the Brantinghams [3]. By analyzing journey-to-crime models, this technology uses sets of linked crime locations in order to indicate possible activity nodes of the criminal and thus make criminal investigations more efficient.

2.2 Assumptions

Several key assumptions were necessary in order to streamline our model:

1. The serial criminal follows *a predictable pattern of movement,* which identifies their residence as an important spot on the geographical profile. There exists a buffer zone, centered at the criminal's home, in which crimes are less likely to occur because of the risk associated with the proximity of the residence, as well as a distance decay

function meaning that the criminal will not reach out too far away from his awareness space [2]. While we include a randomness variable in our model, this cannot account for the actions of any disorganized criminal.

2. The sequence of crimes under consideration has been attributed to *a single suspect* using pieces of evidence that the creators of the geographical profile need not necessarily be aware of. This is necessary because otherwise we cannot be sure we are tracking the same criminal and thus the same pattern of movement.
3. Criminals whose *activity takes place in their residence* or in a single familiar place in their awareness space are not taken into consideration in our model. Examples of this include rapists who lure their victims to their homes or nurses who kill their patients at their workplace. The serial crimes with a single crime location are difficult to identify by law enforcement officers and are beyond the scope of the problem.
4. Suspects are *unlikely to use air travel* due to the high chance of being apprehended. This assumption eliminates the possibility of a suspect traveling from one state to another in a matter of hours and influences the method of estimating distances between crimes.
5. *Country borders cannot be crossed* and state borders are not likely to be crossed in the case of past crimes committed within the same state. Although this assumption excludes situations that have been observed in the past, it nonetheless illustrates the patterns of the majority of serial crimes and helps predicts these cases more accurately.
6. The *official statistics* (population, crime rates, etc.) *are accurate* and correct enough for the purpose of this study.
7. The suspect *does not have the knowledge of being investigated* (or wanted) by the police nor does he feel the urgency of being tracked, or anything that might lead the suspect to digress from his usual criminal behavioral patterns. This is an essential assumption in trying to predict his future actions.

8. Serial crimes are instrumental rather than expressive crimes.

Definition 1.1 Expressive crimes (or affective crimes) are “more spontaneous, emotional, and impulsive crimes that are done in anger. These include domestic violence, some forms of rape, and assaults” [2].

Definition 1.2 Instrumental crimes are crimes committed in order to achieve a goal, such as money, status, or other personal gain [4]. Since expressive crimes are by definition more emotional (and hence less rational) in nature, it is almost impossible that a serial criminal’s act is expressive, unless he or she suffers from some mental disorder (we exclude the discussion of such crimes by Assumption 8).

2.3 Propositions and Foundation

Stability of Criminal Behavior

“Insanity is doing the same thing over and over again and expecting different results.” Albert Einstein (1879—1955)

The intriguing words of wisdom from Einstein suggest a sarcastic yet common truth: whatever a person considers extraordinary of himself can often be routine work in the eyes of others. The same truth applies in the psychological analysis of serial criminals: he might consider what he does creative or different than everyone else, but in reality their crimes are likely to be similar to those committed by other offenders. This form of similarity facilitates our approach. By studying the similarities between a certain criminal’s behaviors and those of others, we are able to construct mathematical models that help us predict this behavior.

In the sense of criminology, the stability of criminal behavior is defined as “the persistence in a behavior or style of interacting over time. There are two components important to the stability of criminal behavior: time and ‘persistence in a behavior or a style of interacting’.” [5]

Indeed, in study after study, “the variable that emerges as the strongest predictor of future criminal behavior is past criminal and delinquent behavior.” [6]. There are mainly two types of stability defined within the realm of criminology:

1. Normative Stability: The preservation of a set of individual ranks on a quality within a constant population over a specified amount of time.
2. Molar Stability: The persistence of a behavior or behavioral orientation as expressed in the rate of change in that quality for an age-homogenous cohort over a specified period of time.

The first type is rather easy to interpret and states the nature of stability in certain, if not all, characteristics of the serial crimes one commits. The second type, on the other hand, considers the usually left out situation where the criminal attentively alters his behavior in a series of crimes. However, the property states that the stability still holds in the way the behavior changes. The stability property enables our model to make assumptions on the future behavior of the criminal based on an aggregation of all previous crimes he or she has made.

Continuity of Criminal Behavior

“You can travel along 10,000 miles and still stay where you are.” Harry Chaplin, Sequel

For many people, continuity seems similar with stability. In criminology terms, stability refers to the “relative consistency over time in rankings of delinquency and crime.”[6] On the other hand, continuity reflects the “strength of previously achieved states, and therefore the probability of their repetition; it implies sameness, familiarity, and predictability.” [7]

For example, stability suggests that it is not quite possible for a serial burglar with no records of murder to suddenly kill someone, although it is possible for him to develop that taste gradually (as denoted by the Molar Stability). Stability also suggests the general accordance of the crimes made by the same person, while continuity emphasizes more on the sequence in which the crimes are committed. For instance, given a different sequence of three serial criminals, we might achieve different prediction results over where the next crime is most probably going to happen.

One example of continuity is the case of Theodore Robert ("Ted Bundy"), who committed over 30 murders between 1973 and 1978. He started off his murders from Washington State where he went to college, and then to Idaho, Utah and Colorado. In Utah he was noticed by the police where they started to investigate him. In the end he was caught in Florida. From Figure XX we can see that he followed a clear path from Northwest to Southeast, which makes his track much easier to be predicted.

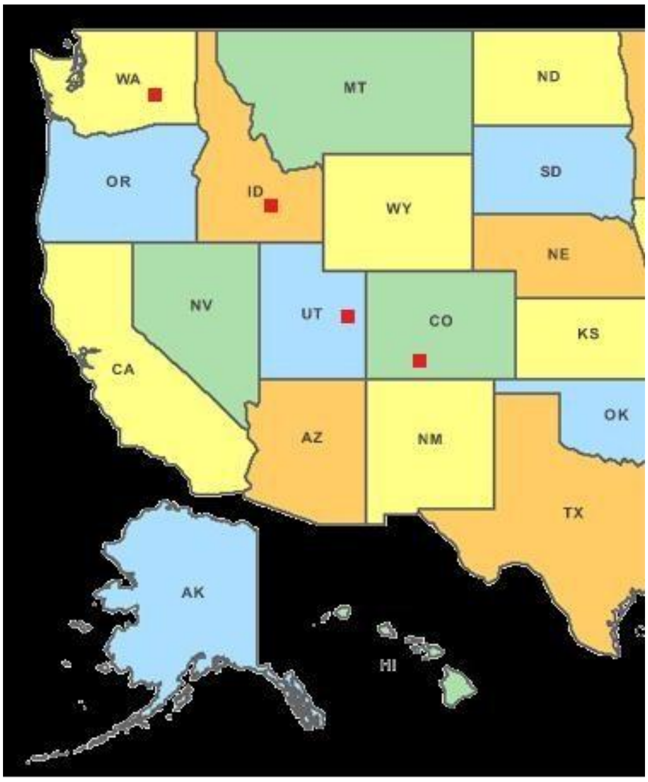

Figure 0. Ted Bundy's Continuous Criminal Pattern

There are generally speaking three types of continuity that can be represented by the behavior of the criminal:

1. Homotypic Continuity: Continuity "over time in the same types of behaviors, such as hitting, kicking, and punching, or traits, such as intelligence or impulsivity."[6]
2. Heterotypic Continuity: Continuity in which "behaviors or traits take different forms over time, but are caused by the same underlying characteristics"[6].
3. Cumulative Continuity: Continuity in which someone's earlier social behavior might affect his later behavior, and the influence is cumulated over rage.

In the first continuity type, there are strong correlations between the serial crimes in the criminal's style, object (victims), psychological states, etc. In the second type of continuity, the criminal's style and object (victims) might be different from case to case, but there are always some initial factors that are the same about the crimes he chooses to make (for example, what kind of crime would he usually commit). In the third case, the criminal actually goes through a "learning" process where his experience accumulates

3. Methods

3.1 Construction of the map – Geographical Method

The first scheme we use in finding the possible locations and the possible home locations is constructing a map based on the existing crime locations and other characteristics of this particular criminal. The output of this method includes the potential target locations and the possible home location of the criminal.

In this construction, our input includes the existing crime locations which are described by latitude and longitude, as well as other characteristics of the location, such as population, criminal rate, history value, psychological value based on the criminal and randomness. Definitions of these other characteristics and how to calculate them will be explained in a later section. Locations help us locate the center of the "criminal circles", areas where the crime is more likely to happen. Using the other characteristics, we determine the radius of these "criminal circles". With these areas, we can then get a set of possible locations where the next crime may happen and the possible home locations of the criminal.

We call this method the geographic method since we are using geographical ranges to determine the possible target locations. The detailed steps are as follows. We first locate the existing crime locations on the map. They are going be the centers of the "crime circles". Then we find the longest distance between two crime spots. Thirdly, we determine the radius of the "crime circles" based on the behavior of the criminal. Fourth, we select the locations within these "crime circles" and identify these locations as the potential target locations. Last, we locate the possible home location in the intersection of these crime circles (the intersection area based on the radium will then naturally be based on those parameters determining the radius).

The figure below shows the process of this geographical method.

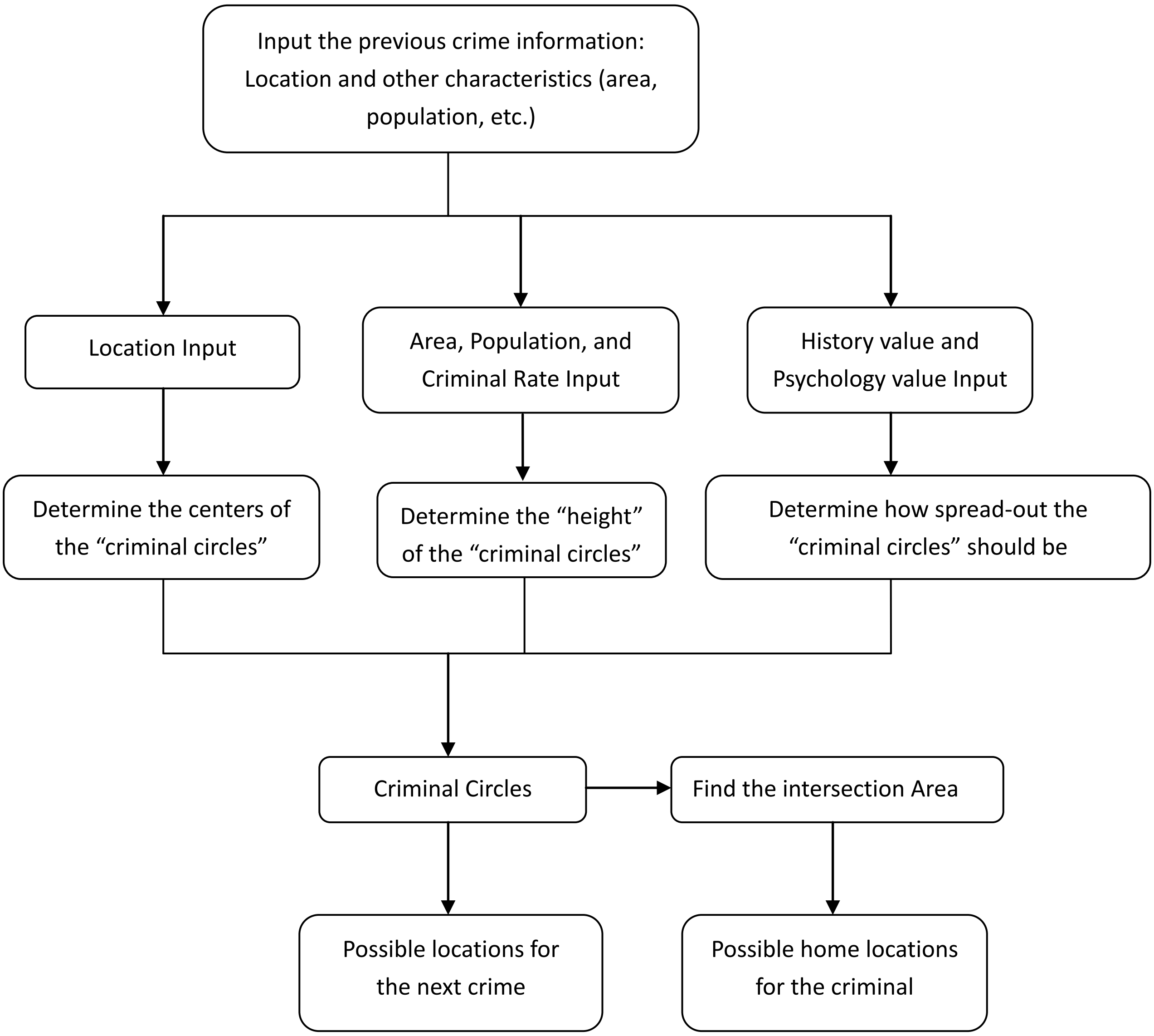

Figure 1. The Process of the Geographical Method

Previous studies have found that time is a commodity so that crimes often occur in nearby areas. The history of serial crimes reveals the inverse relationship between distance from a previous crime spot and the probability that the next crime happens at that distance. In other words, the probability that some later crime happens in some area will be lower as the distance from the previous crime location becomes larger [1]. Therefore, we can use a 2D Gaussian function (though the name is 2D, it is actually a 3D figure in the space) to determine the probability of the next crime happening in an area.

Definition 3.1.1

We define "criminal circles" as some 2D Gaussian functions. Therefore, "criminal circles" are not real 2D circles on some plane, but rather some mountain-shapes existing in the 3D space. The 2D Gaussian function used is,

$$f(x, y) = \alpha e^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2}+\frac{(y-y_0)^2}{2\sigma_y^2}\right)},$$

where α is the peak height of the Gaussian function, the height for the center location, (x_0, y_0) is the location of the center points, and σ_x and σ_y are the standard deviation of the x's and y's respectively, so basically they tells how much the Gaussian function will spread out. The height of these Gaussian functions represents the probability level of points in the x-y plane (so it is not necessarily between 0 and 1 but rather some number whose relative value to the other numbers determines whether the probability is higher or lower than the others').

Definition 3.1.2

We define H to be the history value of one location based on the criminal. H equals the number of times the criminal has committed crimes in that location.

Definition 3.1.3

We define Psy to be the psychological value of the criminal. Psy is a number between 0 and 1.

$$\text{Psy} = 1 - \frac{\text{the number of cities the criminal has conducted crimes} - 1}{\text{the number of crimes he has conducted} - 1}.$$

For example, if some serial killer has killed 10 people and he conducted all his 10 crimes in the same city, his Psy = $1 - \frac{1-1}{10-1} = 1$. However, if another serial killer has murdered 10 people but committed all these crimes in different cities, his Psy = $1 - \frac{10-1}{10-1} = 0$. Therefore we can see that the higher the value is, the more the criminal prefers to conduct crimes in the same location.

Definition 3.1.4

We define $c = (La, Lo, A, P, Cr, H, Psy)$, where c is a set of parameters associated with a previous crime location, La is the latitude of the location, Lo is the longitude of the location, A is the total area of this location, P is the population of the location, Cr is the criminal rate of the location, H is the historical value of the location based on the criminal, and Psy is the psychological value of the criminal.

The first two parameters in c: La and Lo are used to determine the center of these "criminal circles". The centers of the criminal circles are just the previous crime locations since this is where the crime was committed by this criminal. Therefore,

$$\text{Center of "criminal circle"} = (x_0, y_0) = (La, Lo)$$

One thing we should notice is that the latitude of the location in the western hemisphere goes up from right to left instead of left to right. We should take that into consideration when we graph the Gaussian function in order to make it agree with the real locations of these spots on the geographical map.

The Gaussian function then becomes

$$f(x,y) = \alpha e^{-\left(\frac{(x-La)^2}{2\sigma_x^2}+\frac{(y-Lo)^2}{2\sigma_y^2}\right)}$$

Next, we need to figure out how to determine the height of the Gaussian function.

In high population density areas, due to congestion, people cared less about others and usually, most residents in these areas are people with low incomes and of low social status. As a result, the empirical data shows that criminal rates in every category of violent crimes (serial crimes include, as mentioned in the Introduction, serial murders, serial rapes, serial arsons, and some serial robberies and are all violent crimes) are higher than in other areas [1].

Statement 3.1.5

The probability that one particular violent crime happened in one location depends positively on the population density of that location.

Therefore, in our model, we use the population density $= \frac{P}{A}$ and make sure it is positively related to the probability of the next crime.

Another parameter that affects the height of the Gaussian function is the general criminal rate. The population density should have some kind of weight on the general criminal rate since the population density usually serves as a parameter in finding the rate of violent crimes.

Thus, we set the height of the Gaussian function, $\alpha = \frac{P}{A} \times \mathrm{Cr}$. Then the Gaussian function becomes,

$$f(x, y) = \frac{P}{A} \times \mathrm{Cr} \times e^{-\left(\frac{(x-La)^2}{2\sigma_x^2} + \frac{(y-Lo)^2}{2\sigma_y^2}\right)}$$

Our next step is therefore to determine how spread-out the Gaussian function should be.

Assumption 3.1.6

The behavior of one particular criminal is either to commit crimes in one or several locations familiar to him or to commit crimes in random locations. In other words, the criminal either has “preferred” locations (the location where they always commit crimes) or they prefer to change locations every time (the criminal does not have “preferred” locations).

With this assumption, we can then use the history value of the location and the psychological value of the criminal to construct the standard deviations of the Gaussian function.

Since Psy measures whether the criminal has or does not have a “preferred” location and its value ranges from 0 to 1, and H measures how many times the criminal has conducted crimes in this city, these parameters can be used to determine the extent to which the Gaussian function should be spreading out. If the criminal prefers to visit the same place and he has visited this place for several times, there should be a higher possibility that he will come back again. If the criminal prefers to visit the same place but he rarely visited this location in the past, there should be a lower possibility that he will conduct his next crime in this location. If the criminal acts randomly (he prefers not to visit the same place) and he has visited some location many times, he may then not want to come back to the same place again. If the criminal acts randomly and he has just visited some place for just one or two times, he probably will come back.

As a result, we can then express the standard deviation as the product of Psy and H. Also, without loss of generality, $\sigma_x = \sigma_y$. Therefore, we have

$$\sigma_x = \sigma_y = \frac{1}{\mathrm{Psy} \times \mathrm{H}}$$

and the Gaussian function turns out to be

$$\begin{aligned} f(x, y) &= \frac{P}{A} \times \mathrm{Cr} \times e^{-\left(\frac{(x-\mathrm{La})^2}{2/(\mathrm{Psy} \times \mathrm{H})^2} + \frac{(y-\mathrm{Lo})^2}{2/(\mathrm{Psy} \times \mathrm{H})^2}\right)} \\ &= \frac{P}{A} \times \mathrm{Cr} \times e^{-\left(\frac{(x-\mathrm{La})^2 + (y-\mathrm{Lo})^2}{2/(\mathrm{Psy} \times \mathrm{H})^2}\right)} \end{aligned}$$

From previous studies, we know the special mean and standard distance of the crime sites are used to predict the location of the next crime as the next crime will be within this distance. Thus in the last step, we add this weight to this Gaussian function to make it more accurate.

Statement 3.1.7

The special mean and standard distance of the crime sites in a connected series are used to establish the most probable region for the next offense occurrence. [2]

Then we need to add more weight to the areas within the average criminal distance. Let $\bar{r}$ denote the average criminal distance so our model will then become:

$$f(x, y) = \begin{cases} \frac{P}{A} \times \mathrm{Cr} \times e^{-\left(\frac{(x-\mathrm{La})^2 + (y-\mathrm{Lo})^2}{2/(\mathrm{Psy} \times \mathrm{H})^2}\right)} \times 1.1 & \text{when } |x - \mathrm{La}| < \bar{r} \text{ and } |y - \mathrm{Lo}| < \bar{r} \\ \frac{P}{A} \times \mathrm{Cr} \times e^{-\left(\frac{(x-\mathrm{La})^2 + (y-\mathrm{Lo})^2}{2/(\mathrm{Psy} \times \mathrm{H})^2}\right)} & \text{otherwise} \end{cases}$$

The serial murderer usually premeditates his crimes, often fantasizing and planning the crime in every aspect, with the possible exception of the specific victim. This type of killer needs a lot of time to plan their next crime and they are more picky next time they commit a crime. Serial murders usually have an emotional cooling-off period between homicides. It can be days, weeks, or months. The type of multiple murders which happen in just one night or in one week are not considered as serial murders since the criminals are lacking the emotional cooling-off period. The choices for their victims are random and the locations for crimes are usually at the same place or nearby locations [8].

Serial robberies are always armed bank or shop robberies which, like serial murders, require careful planning. For serial rapes, the planning may not be as careful as for serial murders, but the criminals require an emotional cooling-off period as well. For serial arsons, the crimes usually have an emotional cooling-off period too.

Therefore, we can see that cooling-off is a substantial characteristic of serial criminals. Due to the psychological needs associated with it, time intervals between serial crimes are not negligible. We can then assume that it is likely that the serial criminals will go back to their anchor point after one crime and depart from this point for the next crime as well.

Thus, we have the corollary below deduced from Statement 231.7.

Corollary 3.1.8

We can use the special and standard distance of the crimes from the home location to determine the possible target locations for the next crime.

As a result of Corollary 3.1.8, we know that within an average distance of crimes from this anchor point (usually the home), the next crime will be most likely to happen. So the area from the center of the home to the average criminal distance should be weighted more than the others. However, we need to consider a buffer zone as well. The buffer zone is defined as a small area around the criminal's home location in which the criminal prefers not to commit any crime. Since the buffer zone is really small, we just ignore it in this method. Therefore, if we denote (x*, y*) as the home location, the area $(x - x^*)^2 + (y - y^*)^2 \le R$ should be weighted more in the next target probability, where R is the range we should use and it is based on $\bar{r}$ (some function of $\bar{r}$). We can use this as a "learning parameter": we do not use this parameter in building our models to find the next possible locations. However, we find R's value every time after we catch the criminal and find his home location. After many times of application, we can then get a function of R in terms or $\bar{r}$ and then R will help us in the future prediction.

3.2 Static and Dynamic – The Risk Intensity Method

In the first part, we greatly simplified the problem by narrowing down the targeting areas from a general set of locations as well as reducing the size of each unit (atomic) area that we are going to focus on. In the second part, we will synthesize the prediction results from two perspectives (two sources of information) in general:

1. The properties of each atomic area (whether it's a city, a county, or a region in a larger metropolitan region) which was constructed from the first graph. These properties include the statistical data of the area of parameters related to our analysis as well as the geographical information of and between these locations
2. All the information of the previous records of the serial crimes, including the types of crimes, the location and information about the victims

Predictions targeted at the behavior of human beings can be really hard to make due to human nature. However, as mentioned previously in the paper, there has been a substantial amount of criminological studies on the stabilities and continuities of serial crimes which facilitates our modeling. After taking the information from the two areas mentioned above, we decide to take two approaches to the final result:

1. The statistics (parameters) of the locations visited by the criminal provides us with plenty of useful information about the patterns of the criminal. We can use this information to figure out the pattern of the criminal and predict the probability for the next crime to happen in a particular location by examining the agreement of the parameters of that location with the pattern of the parameters we observed in the criminal's previous crimes.
2. Each criminal, no matter how different from other criminals, has some correlation with other criminals that commit the same kind of crimes in the same location. Therefore it is sensible to calculate the risk intensity of all the possible locations we

are interested in and correlate them with the criminal's tendency to commit a certain type of crime in the area to generate the probability for the next crime to occur in that location.

At last, combine the results of these two analyses to synthesize the final probability for the next crime to happen in any possible location determined in geographical method.

How to find the pattern of a serial criminal?

According to our assumptions of stability and continuity, the suspect's crimes observe certain perceptible similarities. Since these similarities usually fall in one or more categories of crimes, it is sensible to interpret that characteristic as a mixture of different types of "ideal" crimes. Each "ideal" exclusively represents one specific type of crime (murder, rape, etc.).
In the real world, however, crimes are more often consisted of only one type of these "ideal" categories. For example, murder usually comes along with rape or burglaries, etc. Thus we will need to reconstruct the serial crimes made by a criminal from the matching results of those committed by him and the typical "ideal" crime.

Categorizing is not only important in the deciding which types of crimes the criminal tends to make, but it is also essential in predicting the sequence of the crimes (specifically the next crime). From the categories of the properties of continuity we can see that there are three basic types of continuous sequences of serial crimes: homotypic, heterotypic and cumulative. Therefore we also need an "ideal" model for each of these types. Here we are not simply classifying the entire series of crimes into one of these models. Instead, we fit different characters which are related to the crimes into those model types.

Another aspect in the analysis of the previous data of the series of crimes is of more importance than the analysis of any static variables of the locations: the type of crimes committed by the suspect. Often times we observe rapes that come together with murders, burglaries that comes together with murders, etc. Figure 2 illustrates the design flow or this part of algorithm.

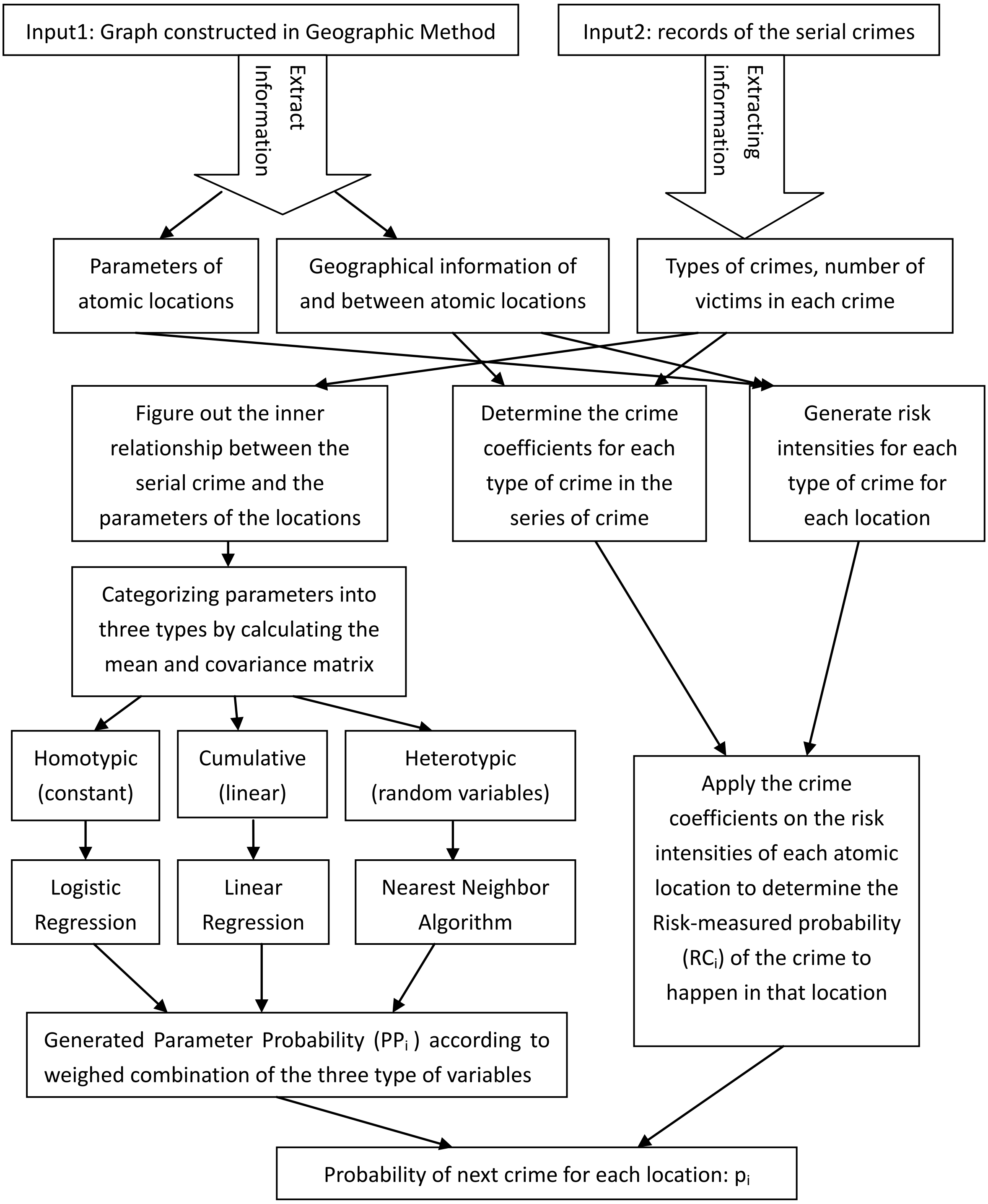

Figure 2. The Process of the Risk Intensity Method

Setting Up

In order to set up the model, we need to introduce certain tools that will help us.

1. L = $\{L_1, L_2, L_3, \ldots, L_k\}$ is the given set of spots (areas, cities, etc) of the unit (atomic) region of previous and next possible crime locations in the graph constructed in geographical method.
2. S = $\{S_1, S_2, S_3, \ldots, S_n\}$ is the given set of spots of the unit regions where the previous crimes are committed. The locations are stored in order of the sequence of the crimes, and repetitions are allowed.
3. For each element S_i in S, there exists a vector P_i $\{P_{i1}, P_{i2}, P_{i3},\ldots, P_{im}\}$ of parameters of normalized variables P = {P1, P2, P3, …, Pm} ($0 \leq Pi \leq 1$) related to each element of S (each atomic location), where the variables are the statistical data of interest to us of that location. All the variables are normalized so that

$$0 \leq Pij \ \leq 1 \ (i \in [1, n], j \in [1, m])$$

4. For each element S_i in S, there exists a subset V_i $\{V_{i1}, V_{i2}, V_{i3}, V_{i4}\}$ in which V_{ij} contains the information of the j^{th} victim in the i^{th} location in the serial crimes.
5. D = $\{D_1, D_2, D_3, \ldots, D_{n-1}\}$ are the set of distances between adjacent crime locations. Also calculate $E_d = \text{mean}\{D\}$ and SD_d which is the standard deviation of the distances in D.
6. DL = $\{DL_1, DL_2, DL_3, \ldots, DL_k\}$ where DL_i is the distance of location Li in L to the last visited crime location.

Stage One: Calculating the Parameter Probability

1. Categorizing the variables

From the foundation part, we can basically categorize the static parameters into three categories: the ones that stay stable over the course of the serial crime; the ones that change in a constant rate over the course of the serial crime; and the ones that do not observe any rules suggested by criminology. The first type is static, which we can treat as constants; the second type is cumulative, which mostly alters in a linear fashion. The third part is rather random, yet we will still be able to determine roughly the trend of the variables by training the data.

Therefore to categorize the variables, we would need to calculate the Mean vector and the Covariance matrix of the variables:

$$E = \{E_1, E_2, E_3, \ldots, E_m\},$$

where E_i is the average of the each variable within the entire set S for $1 \leq i \leq m$:

$$\mathrm{Ei} = \frac{(\mathrm{P1i} + \mathrm{P2i} + \mathrm{P3i} + \cdots + \mathrm{Pni})}{\mathrm{n}}$$

And a covariance matrix C{C1, C2, C3, …, Cm}, where

$$\mathrm{Ci} = \{\mathrm{Ci1}, \mathrm{Ci2}, \mathrm{Ci3}, \ldots, \mathrm{Cim}\} \;\; (\mathrm{i} \in [1, \mathrm{m}])$$

And $\mathrm{Cij} = \mathrm{Cov}(\mathrm{Pi}, \mathrm{Pj}) \;\; (\mathrm{i}, \mathrm{j} \in [1, \mathrm{m}]);$

We denote this matrix C as the **main covariance matrix**.

The really useful information contained in the main covariance matrix is its diagonal, which represents the standard deviation of the variables we are interested in.

1. Denote the vector SD = $\{C_{11}, C_{22}, C_{33}, \ldots, C_{mm}\}$ which is the standard deviation of the variables according to the normalized data. We take all the elements in SD and take out those with values $\leq$ r1, where r1 is defined as the upper limit of the 10% smallest elements in SD, to form a new array Ps = $\{Ps1, Ps2, Ps3, \ldots, Ps_{(ls)}\}$ where ls is the length of array Ps, which is a collection of all the variables which stay stable over the course of the serial crime. And for each element in Ps, store the average of that variable: $\overline{\mathrm{Psi}}$.
2. After excluding the stable elements in SD from set P, perform **linear regression** on the rest of the variables in S.

Those with norm $\leq$ r2 (where r2 is defined as the upper limit of the 40% smallest elements in P) will be recorded in the vector Pc = $\{Pc1, Pc2, Pc3, \ldots, Pc_{(lc)}\}$ where each lc is the length of array Ps, which is a collection of all the related variables whose value accumulates in a linear fashion over the course of the crime. Besides the vector Pc, for each element Pc_i in Pc, we maintain the linear fit function of that variable:

$$\mathrm{f(n)} = \alpha \times \mathrm{n} + \epsilon$$

where α is the coefficient, n is the ranking of the crime in the sequence, and ϵ is the correction coefficient. Also store the standard deviation for the cumulative variables:

$SD_c = \{SD_1, SD_2, SD_3, ..., SD_{(lc)}\}$.

3. After the first two types of variables are determined, store the rest of the variables in $Pr = \{Pr1, Pr2, Pr3, \ldots, Pr_{(lr)}\}$ where lr is the length of the array Pr.

2. Calculate the Parameter Probability based on the patterns of variables

Now that we have successfully categorized the three types of variables, we can make future prediction of those variables based on their different types. In predicting the next possible locations of the crime, each possible location can either have the next crime happen there or not.

1. Therefore for each location, the determination can be viewed as a Bernoulli process for the stable type of variables since the variables usually have a fixed relationship with the probability of the occurrence of the crime. Therefore we include a **Logistic Regression** process with the static variables stored in Ps.

Denote $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$, where x1, x2, ..x_k are the stable variables in Ps. And

$$\beta_i = \frac{1}{SD(x_i)} \ (i \in [1, ls])$$

So that the greater stability x_i has, the bigger β_i will be. Also β_0 is defined as:

$$\beta_0 = -(\beta_1 \overline{x_1} + \beta_2 \overline{x_2} + \cdots + \beta_k \overline{x_2})$$

Then according to Logistic Regression, define:

$$f(z) = \frac{1}{1 + e^{-z}}$$

Then calculate the f(z)'s for all the atomic regions determined in geographical part using their corresponding variables in Pi, and store them in the new set $Lz = \{Lz_1, Lz_2, Lz_3, \ldots, Lz_k\}$. After that normalize the coefficients in Lz in scale such that:

$$\sum_{i=1}^{k} Lz_i = 1$$

2. For the cumulative variables, the best prediction of a certain variable will be determined by its corresponding linear function. Hence for a variable Pc_i in Pc, suppose the number of known murder is N, make the next prediction of the mean

$$Ec_i = f(N) = \alpha \times N + \epsilon$$

Then for the same reason as explained above, the determination can be viewed as a Bernoulli process. Therefore assign the probability of the crime happening in location Lc_i determined from variable Pc_i as a normal distribution:

$$Lc_i \sim No(Ec_i, SDc_i), (i \in [1, lc])$$

Similarly with the previous case, normalize the coefficients in Lc such that:

$$\sum_{i=1}^{k} Lc_i = 1$$

3. For the random variables, we know from the continuity property that there are still inner connections between them and crime probability. Here we use the **nearest neighbor algorithm** in determining the best possible distribution of the same type of variables in all possible locations in L. Since $Pi \in [0,1]$ in P, it is possible to divide the region [0,1] evenly into ten segments, and then take 10 most recent cases of the variable in Pr, and categorize the prediction into the most populated section of [0,1]. If a tie occurs, then take the average of the sections as a prediction to retrieve the average Er_i. Again, for the same reason explained earlier, the determination is still a Bernoulli process, which observes a normal distribution (storing the values in Lr):

$$Lr_i \sim No(Er_i, 0.5), (i \in [1, lr])$$

Similarly, normalize the coefficients in Lr such that:

$$\sum_{i=1}^{k} Lr_i = 1$$

Since stable and cumulative variables are much more precise in the predictions, they are weighed more in the final calculation. Therefore for every possible location in the map Li in L, calculate the Parameter Coefficient PP_i:

$$PP_i = Ls_i \times 45\% + Lc_i \times 45\% + Lr_i \times 10\%$$

Stage Two: Calculating the Risk-measured Probability

Besides calculating the deterministic statistical data from the locations, it is equally important to find patterns within the crimes themselves in the series. One important measurement is the type of crimes committed in the series and the number of victims in each crime. From previous analysis we have found out that criminals, no matter how different they try to be from others, have some basic similarities with those who commit the same type of crimes with them. Therefore we can divide the criminal into composites of several types of "ideal" criminals who are purely dedicated to one type of crime.

Also since the criminal may gradually develop a "taste" for a certain kind of crime, the crimes that happened more recently have more significance than those happened earlier. According to our previous definition, for location Si in set S, there exists a vector V that records the number of victims of different type of crimes in this event. Therefore we create another vector C = {C1, C2, C3, C4} where C1 is the coefficient of murder, C2 for rape, C3 for arson and C4 for burglary. Then

$$\mathrm{C1} = \mathrm{r}^{(\mathrm{n}-1)}\mathrm{V}_{11} + \mathrm{r}^{(\mathrm{n}-2)}\mathrm{V}_{21} + \mathrm{r}^{(\mathrm{n}-3)}\mathrm{V}_{31} + \cdots + \mathrm{rV}_{(\mathrm{n}-1)1} + \mathrm{V}_{\mathrm{n}1}$$

Similarly,

$$\mathrm{C2} = \mathrm{r}^{(\mathrm{n}-1)}\mathrm{V}_{12} + \mathrm{r}^{(\mathrm{n}-2)}\mathrm{V}_{22} + \mathrm{r}^{(\mathrm{n}-3)}\mathrm{V}_{32} + \cdots + \mathrm{rV}_{(\mathrm{n}-1)2} + \mathrm{V}_{\mathrm{n}2}$$

$$\mathrm{C3} = \mathrm{r}^{(\mathrm{n}-1)}\mathrm{V}_{13} + \mathrm{r}^{(\mathrm{n}-2)}\mathrm{V}_{23} + \mathrm{r}^{(\mathrm{n}-3)}\mathrm{V}_{33} + \cdots + \mathrm{rV}_{(\mathrm{n}-1)3} + \mathrm{V}_{\mathrm{n}3}$$

$$\mathrm{C4} = \mathrm{r}^{(\mathrm{n}-1)}\mathrm{V}_{14} + \mathrm{r}^{(\mathrm{n}-2)}\mathrm{V}_{24} + \mathrm{r}^{(\mathrm{n}-3)}\mathrm{V}_{34} + \cdots + \mathrm{rV}_{(\mathrm{n}-1)4} + \mathrm{V}_{\mathrm{n}4}$$

After calculating the coefficients, we can calculate the risk of the next crime according to different weights of C1, C2, C3 and C4.

1. **Generating the Risk Intensities**

A second part in the model is trying to restrict the areas of interest identified in part A by assigning risk intensities to each of them. These risks are static in that they only depend on the crime type and characteristics of the locations under consideration and not on the criminal's pattern (see Figure 2).

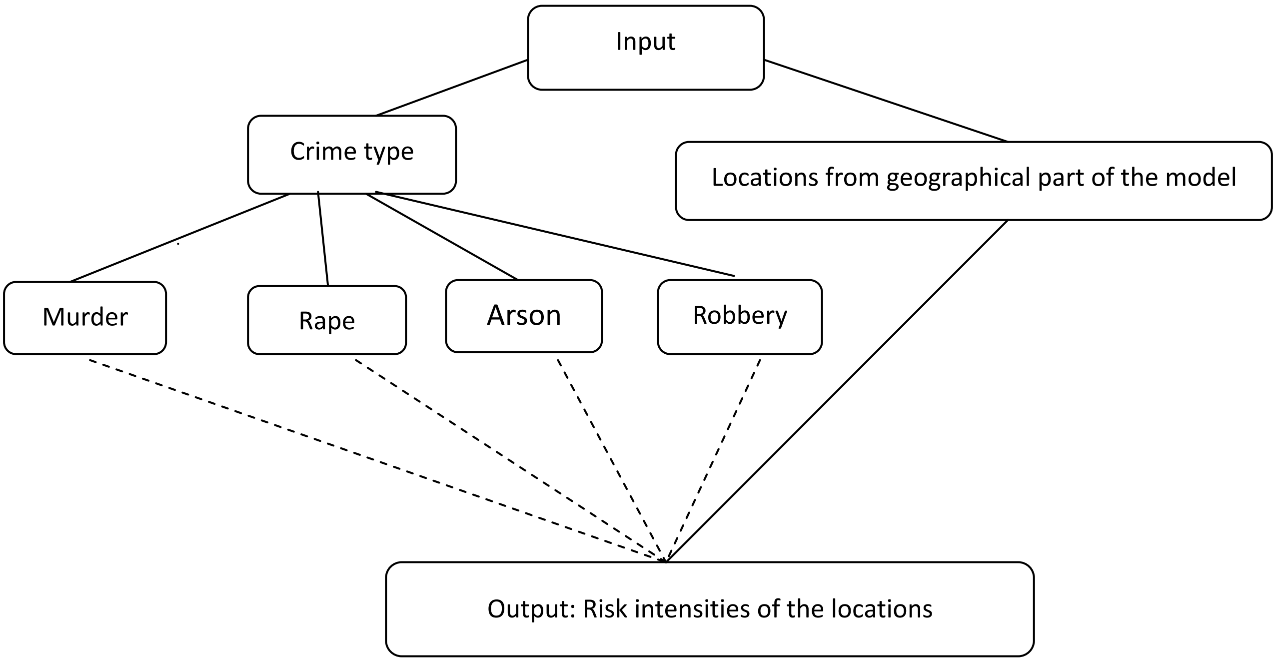

Figure 3. The process of the static method

The intensity risks will further translate into the probability of these areas to become the target of the next crime. In order to be able to identify the risk associated with a particular region, we need to find the variables that can significantly influence this risk. Different variables characteristic to the location under consideration are used according to the nature of the crime. These variables, along with our assumptions, are listed below.

1. Murder
   - Murder rate
   - Economic status: communities with the highest levels of crime also have the highest rates of poverty. There is no suggestion of the fact that poverty in itself causes crime, but inequality does. Due to the limited data that can be found on this topic, it is agreed that this aspect has already been taken into consideration in determining the murder rate.
   - Ethnic/racial heterogeneity: Due to limited statistics available on this subject, it is agreed that this aspect can be associated with other factors such as income

or unemployment rate which have already been taken into consideration in determining the murder rate.

- Distance from anchor point: for murder, trends in serial criminal activity show the need for a buffer zone, because we are not dealing with murders committed exclusively at home or at the criminal's anchor point

The risk intensity is thus defined as:

$\sigma_{murder} = e^{mr} + dec,$

where σ_{murder} is the risk intensity associated with murder,

mr = murder rate of the location

dec = distance-decay function defined as:

$$dec = \begin{cases} 0, if\ d\ is\ within\ a\ small\ range\ (buffer\ zone) \\ \frac{\bar{r}}{\sqrt{d},}, otherwise \end{cases}$$

with $\bar{r}$ the mean of the distances between the given cities

and d the distance from the anchor point to the location being considered

The formula uses the exponential for the murder rate of the location because we consider this to be the most important factor (at this stage) in determining the risk intensity of that location. The distance-decay function used illustrates a buffer zone whose radius needs to be determined using the criminal's psychological profile and patterns.

2. Rape
   - Rape rate
   - Ethnic diversity
   - Distance from anchor point: similar to murder, trends in serial criminal activity show the need for a buffer zone, because we are not dealing with murders committed exclusively at home or at the criminal's anchor point

The risk intensity is thus defined as:

$$\sigma_{rape} = e^{rr} + dec,$$

where σ_{rape} is the risk intensity associated with rape,

rr = rape rate of the location

dec = distance-decay function defined as:

$$dec = \begin{cases} 0, if\ d\ is\ within\ a\ small\ range\ (buffer\ zone) \\ \frac{\bar{r}}{\sqrt{d},}, otherwise \end{cases}$$

with $\bar{r}$ the mean of the distances between the given cities

and d the distance from the anchor point to the location being considered

3. Arson
   - Arson rate
   - Distance from anchor point: serial arsonists are usually not particularly mobile; most of their crimes occur within 1-2 miles of their residence

The risk intensity is thus defined as:

$\sigma_{arson} = e^{ar} + dec,$

Where σ_{arson} is the risk intensity associated with arson,

ar = arson rate of the location

dec = distance-decay function defined as:

$dec = \frac{\bar{r}}{\sqrt{d+\varepsilon}}$ with $\bar{r}$ the mean of the distances between the given cities and d the distance from the anchor point to the location being considered

The formula uses the exponential for the arson rate because we consider this to be the most important factor in determining the risk intensity of the region. We also consider a distance-decay function, where we need the positive constant ε in order to avoid zero terms in the denominator.

4. Robbery
   - Robbery rate
   - Distance from anchor point: in the serial robbery case, studies do not show any consistent trend in how distance from the criminal's familiar place might influence the risk of a particular area.

The risk intensity is thus defined as:

- $\sigma_{robbery} = e^{br}$,

where $\sigma_{robbery}$ is the risk intensity associated with robbery

and br = arson rate of the location

The expression of the risk intensity for robbery of a location only includes the robbery rate as a variable, since other parameters can either not be easy to find or do not show a clear influence on the risk.

Remark

Not mentioned in the above variables is the urban design or traffic pattern of a location, given that the physical layout of streets in an area can have an impact on criminal activity. For example, predictable street grid networks are more likely to attract crime, while more "organic" street layouts are generally considered safer. However, this data is considerably difficult to come across, both at the state and county level.

Once we identify the nature of the crime, the model assigns risk intensities to the locations identified by the geographical method. These intensities are then normalized to provide values in the interval (0, 1) for the risk of each location:

$$risk_i = \frac{\sigma_i}{\sum_{i=1}^{n} \sigma_i}$$

where σ_i = risk intensity of location i

n = number of locations under consideration.

2. **Calculating the Risk-measured Probability**

Given the risk intensities of the four types of crimes of an atomic location Li in set L (R_{i1}, R_{i2}, R_{i3}, R_{i4}) , C1, C2, C3, C4, we can calculate the Crime Coefficients CC_i for each atomic location:

$$\mathrm{CC_i} = \sum_{j=1}^{4} \mathrm{Rji} \times \mathrm{Cj}$$

After calculating CC_i for all $i \in [1, k]$, normalize the crime coefficients of the cities to RC_i so that

$$\sum_{i=1}^{n} \mathrm{RC_i} = 1$$

Stage Three: Integration

Since the predictions based on crime types and the predictions based on the patterns of the statistical parameters of the locations are of equal importance, the final prediction should be given equal dependence on the results of these two methods. So the combined prediction results Cp_i for the L_i in set L is:

$$Cp_i = \frac{(PP_i + RC_i)}{2}$$

There is, however, another important factor to be taken into consideration: the distance decay principle where the possibilities decay with respect to the distance from the current crime location to the possible target location. Given that we've calculated the mean E_d and the standard deviation SD_d of the distances travelled between cities in the course of the serial crime, we can add the distance weight to the probabilities previously calculated for each atomic location Li. Also according to the "buffer zone" effect, the criminal is less willing to commit crime in the same locations. According to the set DL defined above, we create a new set DP in which its element DP_i is defined as:

$$DP_i = \begin{cases} 1 & DL_i < E_d \\ \sim No(E_d, SD_d) & DL_i > E_d \end{cases}$$

Therefore the final probability of function for each location Li is p_i:

$$p_i = \begin{cases} Cp_i \times DP_i & i \neq 0; \\ Cp_i \times 0.5 & i = 0. \end{cases}$$

After that normalize the probabilities p_i such that:

$$\sum_{i=1}^{k} pi = 1$$

4. Simulation Results and Discussion

We implemented the methods and algorithms described above using Matlab and generated the graphs using the same program. We stored the output from each step described above in order to note the results of each method.

4.1 Results of the Geographical Method

The geographical method resulted in reasonably smoothed Gaussian surfaces indicating areas with high crime probability based on past crime locations and some characteristics of these areas.

We first considered the *Gary Latray case*, a serial bank robbery criminal who was recently apprehended and whose activity across the states of Pennsylvania, Maryland, Ohio and Virginia is particularly interesting. We notice that the graph generated in the geographical method for this situation illustrates a sharp peak at one of the locations visited by the criminal. The other peaks are smooth and distributed along the other areas where he committed the crime (Figure).

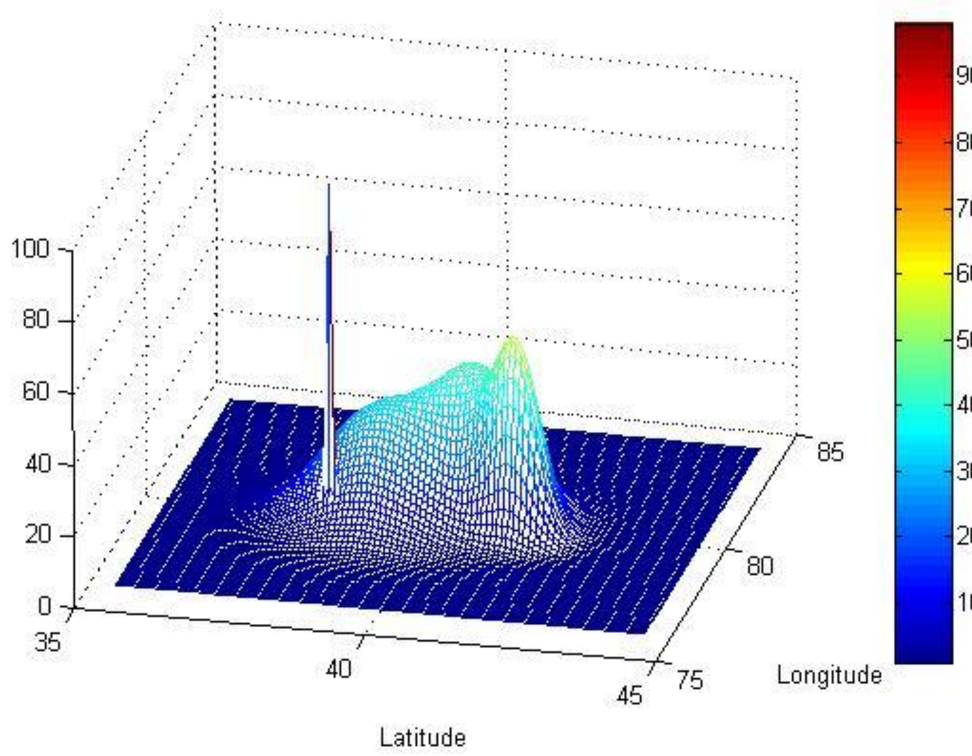

Figure 4. Distribution of high-risk areas in the Gary Latray robbery case using the geographic method

Figures and show the distribution of the potential target locations, with yellow areas suggesting a high crime risk. The red point corresponds to the peak mentioned above.

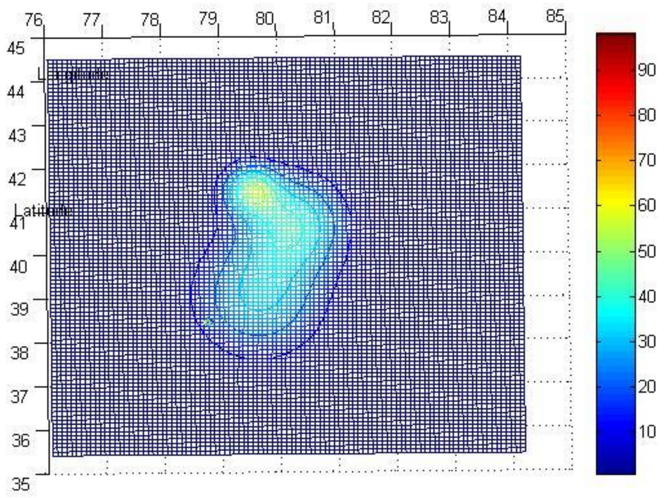

Figure 5. Distribution of potential target areas in the Gary Latray robbery case

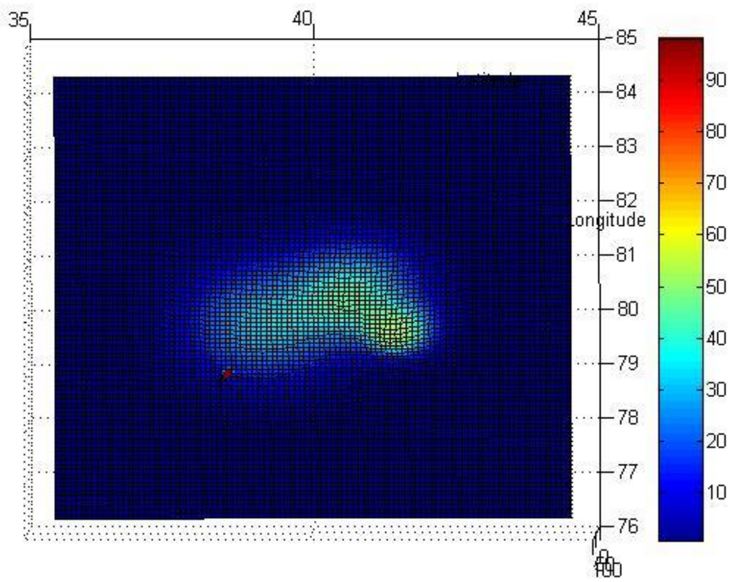

Figure 6. Distribution of potential target areas in the Gary Latray robbery case

We also tested our model on a serial rape and murder case which occurred in Florida in 2003-2004. The murderer – a female – committed the crimes in several counties in the state. However, she visited Marion County 3 times out of the seven she was convicted for. The geographical method generated the following graph for this case:

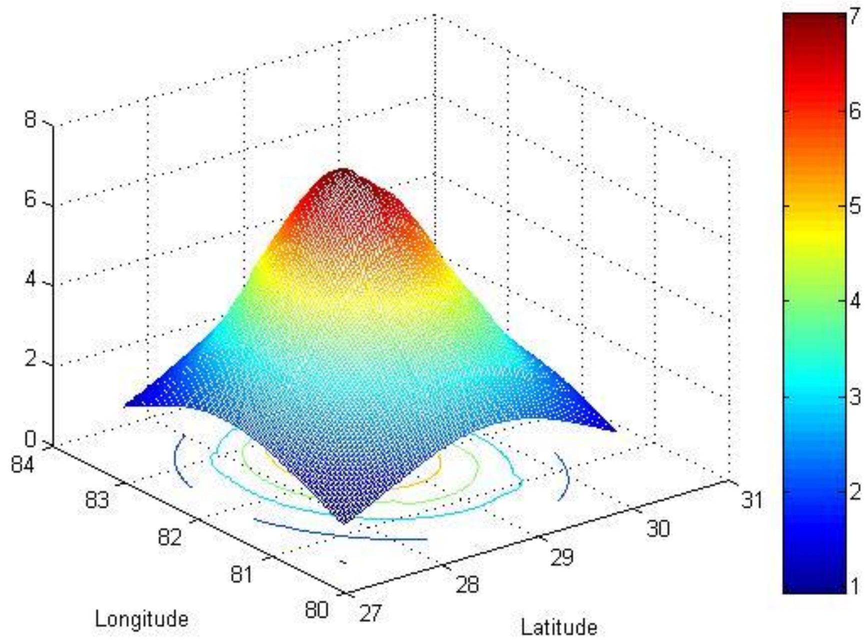

Figure 7. Distribution of high-risk areas in the Florida murder case using the geographic method

Note that he red peak on this graph is centered on the coordinates of Marion county, which agrees with our expectation of this area having a higher crime probability for this particular offender. The criminal circles essential for generating the shape can also be seen from the graph.

The distribution of potential target locations can also be observed in the picture on the next page:

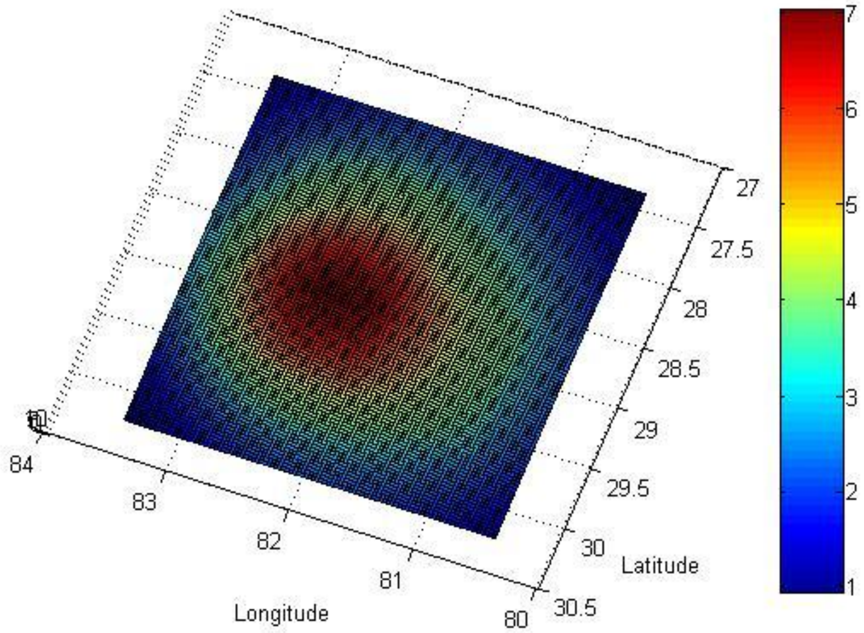

Figure 8. Distribution of potential target areas in the Florida murder case

The geographical model was further tested for a serial arson situation. Thomas A. Sweatt terrorized the Washington DC area, Maryland and Virginia with the numerous fires set there before being caught in West Virginia, which can be considered his “anchor point” as discussed in this paper. The results of this test were that the high risk areas would be closer to DC and Maryland (see figure). This confirms our expectations since these locations are the ones where he set up most of the fires. We also note the higher density of these states as opposed to Virginia.

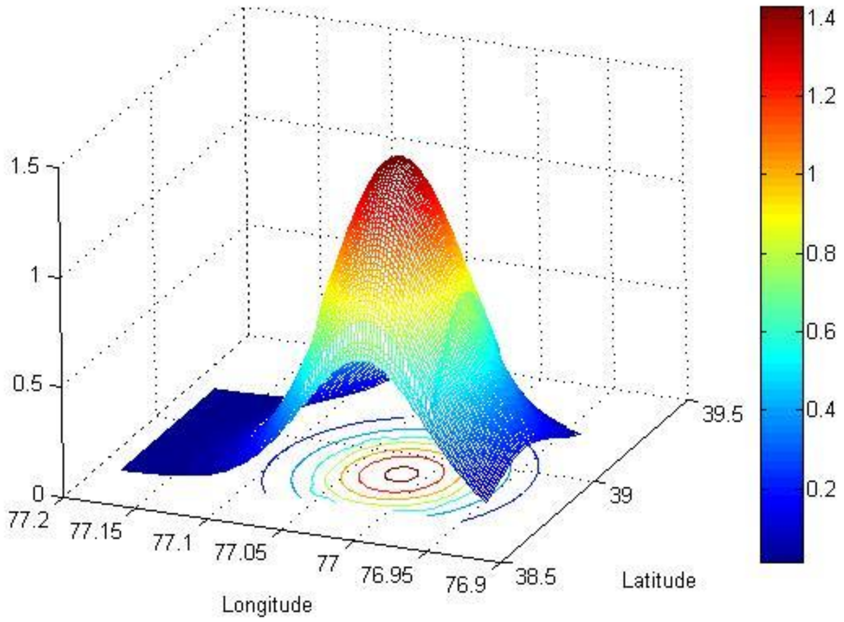

Figure 9. Distribution of high-risk areas in the Thomas Sweatt arson case using the geographic method

Note that the criminal circles are again conspicuous the generated Gaussian shape. A 2D view of the graph is provided in the picture below:

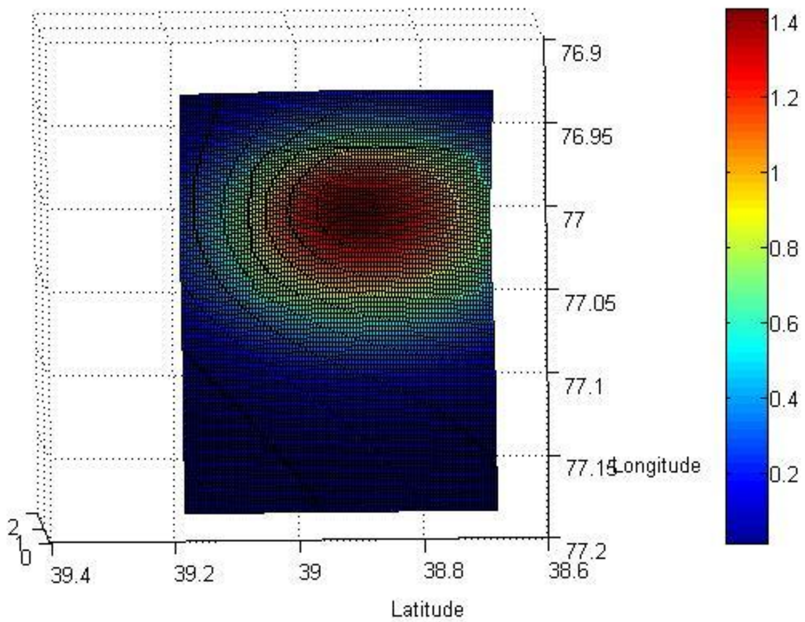

Figure 10. Distribution of potential target areas in the Sweatt arson case

4.2 Results of the Risk Intensity, “static” and “dynamic”, Methods

The output of the “dynamic method” is the parameter probability (PPi), generated by categorizing the parameters as homotypic, heterotypic or cumulative.

For the Gary Latray robbery case, the model generates the probabilities below:

	East Franklin	**St. Clairsville**	**Oakland**	**Harrisonburg**	**Belpre**
Parameter probabilities	**0.0921**	**0.2087**	**0.1812**	**0.3086**	**0.2095**

For the same robbery case, the “static” method generates risk intensities associated with the locations under consideration:

	East Franklin	**St. Clairsville**	**Oakland**	**Harrisonburg**	**Belpre**
Risk intentisites	**1.8035**	**2.2638**	**21.2636**	**22.2755**	**8.0614**

Note the considerable greater risk intensities for Oakland, MD and Harrisonburg, VA, which are most likely due to the high crime rates, population densities and influenced by their distance to the “anchor point”.

However, these intensities are not our final result. By combining them with the robbery crime coefficient and the static method, we get the probabilities of the locations to be the next target for this criminal (see next page):

	East Franklin	St. Clairsville	Oakland	Harrisonburg	Belpre
Final Probability Prediction	0.0838	0.2837	0.2850	0.0122	0.3352

Therefore, by considering all parameters and the criminal's sequence of actions, we predict the Belpre area to be the most probable future target, followed closely by Oakland and St. Clairsville.

4.3 Discussion

4.3.1 Sensitivity and Robustness Testing

A discussion of the quality of the model cannot be complete without mentioning robustness and sensitivity analysis. While robustness is used to determine whether the model will break down in extreme cases, sensitivity measures the effect of small changes in parameters. For a good model, this will induce small changes in the output data.

We tested our model for *sensitivity* by using the Gary Latray serial robbery example and varying the value of one parameter at a time.

We first tested the model's response for a minor change in the population on location 5 (Belpre, OH). This change (from 44015 to 40000 inhabitants) further resulted in a change in the population density. The results were as follows (see next page):

	East Franklin	St. Clairsville	Oakland	Harrisonburg	Belpre
Initial Probability Prediction	0.0838	0.2837	0.2850	0.0122	0.3352
Minor change population prediction	0.0852	0.2924	0.2739	0.0143	0.3343

Note that there are small changes in probabilities and the highest-probability location does not change. This suggests small changes in population do not affect the output to a great extent.

The test for a minor change in the coordinates of Location 5 (by 1 degree), however, translated in a significant change in probability:

	East Franklin	St. Clairsville	Oakland	Harrisonburg	Belpre
Initial Probability Prediction	0.0838	0.2837	0.2850	0.0122	0.3352
Minor change location prediction	0.0300	0.1289	0.5266	0.0091	0.3054

The location with the highest probability is no longer Belpre, OH but the third location (Oakland, MD), which suggests that geographic changes affect our output greatly.

We performed a similar test for a minor change in area (area of location 5 was changed from 3.5 sq mi to 4 sq mi), and noticed a small change in the final probabilities:

	East Franklin	**St. Clairsville**	**Oakland**	**Harrisonburg**	**Belpre**
Initial Probability Prediction	0.0838	0.2837	0.2850	0.0122	0.3352
Minor change area prediction	0.0838	0.2843	0.2853	0.0122	0.3343

For a change in crime rate, we obtained similar data and this concluded the analysis of sensitivity of our model:

	East Franklin	**St. Clairsville**	**Oakland**	**Harrisonburg**	**Belpre**
Initial Probability Prediction	0.0838	0.2837	0.2850	0.0122	0.3352
Minor change crime rate prediction	0.0810	0.2668	0.3067	0.0083	0.3373

An analysis of the *robustness* of the model was achieved using the Kansas City murder and rape example, where all crimes were committed inside the same city (data on the exact location of the crimes is very difficult to find). We tested this example by using as input only the variables associated with Kansas City. Our results were consistent with the expectation that the model would assign full chances to Kansas as being the next crime location as well.

4.3.2 Accuracy of the Prediction

In this section, we are going to discuss the accuracy of our model. The two main interests are whether the model is able to highlight the next murder area and whether the high-risk area has been minimized.

In the Florida serial murders case, we used the geographical method to predict the next crime. Since the murderer Aileen Wuornos was caught after the seventh murder, there is actually no “next crime”. However, we can input the first six murders’ information into our model and check if the seventh murder’s location has been covered in our prediction. As showed in 3.1, the seventh murder’s location, Dixie County, has been well covered in our prediction, being highlighted as a high-risk area.

As we noticed in the graphs above, the high-risk area has been constrained to an area about 10,000 km^2, equivalently about 5 counties in Florida. However, with the idea of “home location” and the increase in the number of data used, the range can be minimized to about 2 to 3 counties.

With the application of the second method, the range of high-risk area can be further minimized. As in the serial robberies case discussed above, after the implementation of the geographic method, five locations are highlighted as of high risk. Then we further use the risk intensity method and the final probability shows that three cities are of clearly higher risk as opposed to others: St. Clairsville, Oakland, and Belpre (the real last robbery occurs in Belpre, which has a higher risk probability). We can see that with the implementation of the second method the range of the next possible locations has been constrained even further, to 3 locations.

4.3.3 Combination of the Two Methods

We integrate the two methods, the geographical one and the risk intensity one, by using the first method as a preliminary step for the second one. In fact, the first method generates a map for the second one: when we calculate the risk intensity, we are only concerned about the locations that appear in the first map.

5. Strength and Weakness of the Model

This model is considerably accurate. With the combination of the two methods, the geographic method and the risk intensity one, we can minimize the range of possible locations to about 2 or 3 areas. With the testing of existing data, our model successfully highlighted the location for the next crime. Most of the variables that could influence the location of the next crime on the offender's target list are taken into consideration. Therefore, the model achieves the goal of predicting the location of the next crime in serial crimes and in minimizing the number of possible locations.

On the other hand, our model did not take into account the fact that sometimes the boundary of administrative regions might impede with the investigations ("edge effects"). Moreover, pieces of evidence such as similarities of the victims are not represented in the determination process, although they could affect the spatial patterns of a criminal. Another weakness of our model can be the fact that small changes in the geographic coordinates influence to a certain extent our output. However, the reason for this could be that the input geographic coordinates did not differ greatly.

We should also keep in mind the assumptions we made in order to generate this model. For instance, it is essential to assume that the criminal follows a predictable pattern of movement or that there is some kind of correlation between his actions and those of other criminals that commit the same kind of crime in that location (see risk intensity method assumptions).

6. Conclusion and Recommendation

In this paper, we have developed and discussed our model for locating the next possible locations for serial crimes. Our model takes into account the statistical information of previous crimes, such as the geographical coordinates, population densities, criminal rates, etc, as well as the criminal's previous crime behavior. We consider the crime pattern of the criminal in order to calculate the risk intensities for different areas. Our model involves the implementation of the Gaussian distribution, distance decay function, logistic regression, linear regression and nearest neighbor algorithm based on the statistical investigation of the areas and the criminal's previous records.

Our main assumption in building this model is that serial crimes are instrumental rather than expressive, thus ensuring that the criminal follows a predictable pattern of movement. We also assume that this pattern is characterized by certain stability and continuity properties which facilitate a correlation between the behavior of the suspect and that of the other criminals in concerned areas.

During the simulation process, we used the data from previous serial crimes to test and verify the prediction of our model about the next crime location. Our model's predictions match the actual location of the next crime location in the sequences we tested. It is proved that the combination of the first and the second method increases the accuracy of the model's prediction in general while at the same time decreases the amount of work needed for the separate implementation of the two methods. The first model simplifies the tasks for the second method by narrowing the range of next possible crime locations. The second method further identifies the risk intensity as well as the criminal's pattern which are then combined in synthesizing the best probable location(s).

Admittedly, our algorithm for the next crime location is only guaranteed to work for those criminals with a predictable pattern of movement. For the other criminals, however, our model is not guaranteed to determine an accurate and exact pattern of the criminal. Neither is it possible under such circumstances to minimize the range of location to a small set of areas. Also, the introduction of the home location in our model needs a considerable amount of data for the purpose of the training of the variables in the model. Therefore an aggregation of criminal data is required for the home location "learning" process to work.

In general, our model achieves the goal of predicting the criminal's next possible target locations and minimizing the range to two or three most probable areas. It is able to assist law enforcement officers in locating the next crime and putting the serial criminal in control.

8. Appendices

A. Bibliography

[1] D. Kim Rossmo. Geographic Profiling. CRC Press, 1999

[2] Kim Michelle Lersch. Space, time, and crime. Carolina Academic Press, 2004

[3] Environmental Criminology Research Inc. Geographic Profiling: an effective tool for Serial Crime Investigation

[4] Terance D. Miethe, Richard C. McCorkle. Crime Profiles : The Anatomy of Dangerous Persons. Oxford University Press, 2001

[5] Loeber, R. Criminal Behaviour and Mental. Health, 3, 492-523. 1982

[6] John Paul Wright, Stephen G. Tibbets, Leah E. Daigle. Criminals in the Making. Sage Publications, Inc., 2008

[7] Larry J. Siegel, Joseph J. Senna. Introduction to Criminal Justice. Wadsworth Publishing, 2007

[8] FBI. Serial Murder — Multi-Disciplinary Perspectives for Investigators.

B. Data

(resource: http://www.fbi.gov/ucr/cius2008/offenses/index.html)

Florida Murder Case:

Places visited	Latitude	Longitude	Population	Area (square miles)	Density (/square miles)	Crime rate (%)	Number of times visited
Volusia	29°10’	81°31’	498,036	1,432	401	0.57	1
Marion	29°16’	82°07’	329,628	1,663	163	1.35	3
Citrus	28°53’	82°31’	141,416	773	202	0.26	1
Pasco	28°19’	82°20’	471,028	868	464	0.43	1
Dixie	28°25’	82°18’	14,957	864	21	0.095	1

Kansas City Murder and Rapist Case:

Places visited	Latitude	Longitude	Population	Area (square miles)	Density (/square miles)	Arson rate (%)	Number of times visited
Washington DC	38°51’	77°02’	591,833	68.3	9,776.40	0.015	21

Thomas A. Sweatt Arson Case:

Places visited	Latitude	Longitude	Population	Area (square miles)	Density (/square miles)	Arson rate (%)	Number of times visited
Washington DC	38°51’	77°02’	591,833	68.3	9,776.40	0.015	21
Maryland	39°01’	77°01’	5,633,597	12,407	541.9	0.0041	19
Virginia	38°52’	77°06’	7,769,089	42,774.20	193	0.0028	5

Gary Latray Robbery Case:

Places visited	Latitude	Longitude	Population	Area (square miles)	Density (/square miles)	Crime rate (%)	Number of times visited
East Franklin, PA	40°88’	79°49’	3,900	31.5	126.5	0.59	1
St. Clairsville, OH	40°04’	80°54’	5,057	2.2	2,354	1.46	1
Oakland, MD	39°24’	79°24’	1,930	2.1	915.7	3.05	1
Harrisonburg, VA	38°26’	78°52’	44,015	17.6	2,559	3.1	1
Belpre, OH	39°16’	81°35’	6,660	3.6	1,889.60	2.08	1

C. Code

(Codes are arranged alphabetically according to their name)

1. FunctionOne

```
function f = functionOne(P,A,Cr,x,y,La,Lo,Psy,H);
f = (P./A).*Cr.*exp(-((x-La).^2+(y-Lo).^2)./(2./((Psy.*H).^2)))*1.1;%(2./(Psy.*H).^2)).*1.1;
```

2. FunctionTwo

```
function f = functionTwo(P,A,Cr,x,y,La,Lo,Psy,H);
f = (P./A).*Cr.*exp(-((x-La).^2+(y-Lo).^2)./(2./((Psy.*H).^2)));%(2./(Psy.*H).^2)).*1.1;
```

3. Location

```
function location = location(degree, minute);
location = degree+minute./60;
```

4. Logistic Regression

```
function f = LogisticRegression(z);
f=1./(1+exp(-z));
```

5. Main

```
StateName = ['DC', 'Maryland', 'VA'];

LongitudeDegree = [77 77 77];
LongitudeMinute = [02 01 06];
LatitudeDegree = [38 39 38];
LatitudeMinute = [51 01 52];

LongitudeLocation = location(LongitudeDegree, LongitudeMinute);
LatitudeLocation = location(LatitudeDegree, LatitudeMinute);

Population = [591233 5633597 7769089];
CrimesStat = [89 231 219];
Area = [68.3 12407 42774.2];
Crimes = [21 19 5];
CrimeStat = [89 231 219];
CrimeRate = CrimeStat./Population;
PsycValue = 1-(Crimes-1)./sum(Crimes);
H = Crimes;



LocationSize = length(LongitudeLocation);
```

```
Distances=zeros(LocationSize);

for i = 1:1:LocationSize
    for j = 1:1:LocationSize
        Distances(i,j) = sqrt((LongitudeLocation(i)-LongitudeLocation(j)).^2+(LatitudeLocation(i)-LatitudeLocation(j)).^2);
    end;
end;

DistanceAverage = mean(mean(Distances));
Xradius = max(LatitudeLocation)-min(LatitudeLocation);
Yradius = max(LongitudeLocation)-min(LongitudeLocation);
Xmin = min(LatitudeLocation)-Xradius; Xmax = max(LatitudeLocation)+Xradius;
Ymin = min(LongitudeLocation)-Yradius; Ymax = max(LongitudeLocation)+Yradius;

Xrange = linspace(Xmin,Xmax,1e2); Yrange = linspace(Ymin,Ymax,1e2);
map = meshgrid(Xrange,Yrange);
Value = zeros(length(Xrange),length(Yrange));
Value1 = Value;


for i = 1:1:1e2
    for j=1:1:1e2
         for k=1:1:LocationSize
            if (abs(Xrange(i)-LatitudeLocation(k))<DistanceAverage && abs(Yrange(j)-LongitudeLocation(k))<DistanceAverage)
                Value(i,j) = Value(i,j)+functionOne(Population(k),Area(k),CrimeRate(k),Xrange(i),Yrange(j),LatitudeLocation(k),LongitudeLocation(k),PsycValue(k),Crimes(k));
            else Value(i,j)= Value(i,j)+functionTwo(Population(k),Area(k),CrimeRate(k),Xrange(i),Yrange(j),LatitudeLocation(k),LongitudeLocation(k),PsycValue(k),Crimes(k));
            end;
        end;
    end;
end;

figure(1)
meshc(Xrange,Yrange,Value);
xlabel('Latitude');ylabel('Longitude');
figure(2)
surf(Xrange,Yrange,Value);
xlabel('Latitude');ylabel('Longitude');
```

6. Murder

```
County = ['Volusia' 'Marion' 'Citrus' 'Pasco' 'Marion' 'Marion' 'Dixie'];
LatitudeDegree = [29 29 28 28 28];
LatitudeMinute = [10 16 53 19 25];
LongitudeDegree = [81 82 82 82 82];
LongitudeMinute = [31 07 31 20 18];

LatitudeLocation = location(LatitudeDegree,LatitudeMinute);
LongitudeLocation = location(LongitudeDegree,LongitudeMinute);

Population = [498036 329628 141416 471028 14957];
CrimeRate = [.57 1.35 .26 .43 .095]/100;
Area = [1432 1663 773 868 864];
Crimes = [1 3 1 1 1]
PsycValue = 1- Crimes/sum(Crimes);
H = Crimes;

LocationSize = length(LongitudeLocation);
Distances=zeros(LocationSize);

for i = 1:1:LocationSize
    for j = 1:1:LocationSize
        Distances(i,j) =
sqrt((LongitudeLocation(i)-LongitudeLocation(j)).^2+(LatitudeLocation(i)-LatitudeLocation(j)).^2);
    end;
end;

DistanceAverage = mean(mean(Distances));
Xradius = max(LatitudeLocation)-min(LatitudeLocation);
Yradius = max(LongitudeLocation)-min(LongitudeLocation);
Xmin = min(LatitudeLocation)-Xradius; Xmax = max(LatitudeLocation)+Xradius;
Ymin = min(LongitudeLocation)-Yradius; Ymax = max(LongitudeLocation)+Yradius;

Xrange = linspace(Xmin,Xmax,1e2); Yrange = linspace(Ymin,Ymax,1e2);
map = meshgrid(Xrange,Yrange);
Value = zeros(length(Xrange),length(Yrange));
Value1 = Value;

for i = 1:1:1e2
    for j=1:1:1e2
        for k = 1:1:LocationSize
            if (abs(Xrange(i)-LatitudeLocation(k))<DistanceAverage &&
abs(Yrange(j)-LongitudeLocation(k))<DistanceAverage)
                Value(i,j) =
```

```
Value(i,j)+functionOne(Population(k),Area(k),CrimeRate(k),Xrange(i),Yrange(j),LatitudeLocation(k),Lon
gitudeLocation(k),PsycValue(k),Crimes(k));
            else Value(i,j)=
Value(i,j)+functionTwo(Population(k),Area(k),CrimeRate(k),Xrange(i),Yrange(j),LatitudeLocation(k),Lon
gitudeLocation(k),PsycValue(k),Crimes(k));
            end;
        end;
    end;
end;
figure(1)
meshc(Xrange,Yrange,Value);
xlabel('Latitude');ylabel('Longitude');
figure(2)
surf(Xrange,Yrange,Value);
xlabel('Latitude');ylabel('Longitude');
```

7. Murder Calc

```
County = ['Volusia' 'Marion' 'Citrus' 'Pasco' 'Marion' 'Marion' 'Dixie'];
LatitudeDegree = [29 29 28 28 29 29];
LatitudeMinute = [10 16 53 19 16 16];
LongitudeDegree = [81 82 82 82 82 82];
LongitudeMinute = [31 07 31 20 07 07];

LatitudeLocation = location(LatitudeDegree,LatitudeMinute);
LongitudeLocation = location(LongitudeDegree,LongitudeMinute);

Population = [498036 329628 141416 471028 329628 329628];
CrimeRate = [.57 1.35 .26 .43 1.35 1.35]/100;
Area = [1432 1663 773 868 1663 1663];
Density = Population./Area;
%Crimes = [1 3 1 1 1];
TotalSize = length(LatitudeLocation);
SampleSize= TotalSize-1;
PopulationSample = Population(1:SampleSize);
CrimeRateSample = CrimeRate(1:SampleSize);
AreaSample = Area(1:SampleSize);
DensitySample = Density(1:SampleSize);

NormPopulationSample = normalize(PopulationSample);
NormCrimeRateSample = normalize(CrimeRateSample);
NormAreaSample = normalize(AreaSample);
NormDensitySample = normalize(DensitySample);

MeanPopulationSample = mean(PopulationSample);
```

```
MeanCrimeRateSample = mean(CrimeRateSample);
MeanAreaSample = mean(AreaSample);
MeanDensitySample = mean(DensitySample);

StdPopulationSample = std(PopulationSample);
StdCrimeRateSample = std(CrimeRateSample);
StdAreaSample = std(AreaSample);
StdDensitySample = mean(DensitySample);
x = 1:1:SampleSize;

[PopulationP,PopulationS]=polyfit(x,PopulationSample,1);
[CrimeRateP,CrimeRateS]=polyfit(x,CrimeRateSample,1);
[AreaP,AreaS]=polyfit(x,AreaSample,1);
[DensityP,DensityS] = polyfit(x,DensitySample,1);

M = [NormPopulationSample' NormCrimeRateSample' NormAreaSample' NormDensitySample'];
N = cov(M);

Distances = zeros(TotalSize);
for i = 1:1:TotalSize,
    for j = 1:1:TotalSize,

Distances(i,j)=sqrt((LatitudeLocation(i)-LatitudeLocation(j))^2+(LongitudeLocation(i)-LongitudeLocation
(j))^2);
    end;
end;
MeanDistance = mean(mean(Distances));
StdDistance = mean(std(Distances));
for i=1:1:TotalSize
    Next =
sqrt((LatitudeLocation(i)-LatitudeLocation(SampleSize)).^2+(LongitudeLocation(i)-LongitudeLocation(S
ampleSize)).^2);
    AdjacentDistances(i) = Next;
end;

RiskIntensity = MurderRisk(CrimeRate,StdDistance,AdjacentDistances);
PR = normalize(RiskIntensity);

AreaMeanPrediction = MeanAreaSample;
AreaStdPrediction = StdAreaSample;
AreaPrediction = normpdf(Area,AreaMeanPrediction,AreaStdPrediction)
NormAreaPrediction = normalize(AreaPrediction);

PopulationMeanPrediction = polyval(PopulationP,TotalSize);
```

```
PopulationStdPrediction = StdPopulationSample;
PopulationPrediction = normpdf(Population,PopulationMeanPrediction,PopulationStdPrediction);
NormPopulationPrediction = normalize(PopulationPrediction);

DensityMeanPrediction = polyval(DensityP,TotalSize);
DensityStdPrediction = StdDensitySample;
DensityPrediction = normpdf(Density,DensityMeanPrediction,DensityStdPrediction);
NormDensityPrediction = normalize(DensityPrediction);

CrimeRateMeanPrediction = polyval(CrimeRateP, TotalSize);
CrimeRateStdPrediction = StdCrimeRateSample;
CrimeRatePrediction = normpdf(CrimeRate,CrimeRateMeanPrediction,CrimeRateStdPrediction);
NormCrimeRatePrediction = normalize(CrimeRatePrediction);

PP = normalize(NormAreaPrediction*0.1+(NormPopulationPrediction +
NormDensityPrediction+NormCrimeRatePrediction)*0.9);
P = (PP+PR)./2;

% LocationSize = length(LongitudeLocation);
% Distances=zeros(LocationSize);
%
% for i = 1:1:LocationSize
%     for j = 1:1:LocationSize
%         Distances(i,j) =
sqrt((LongitudeLocation(i)-LongitudeLocation(j)).^2+(LatitudeLocation(i)-LatitudeLocation(j)).^2);
%     end;
% end;
%
% DistanceAverage = mean(mean(Distances));
% Xradius = max(LatitudeLocation)-min(LatitudeLocation);
% Yradius = max(LongitudeLocation)-min(LongitudeLocation);
% Xmin = min(LatitudeLocation)-Xradius; Xmax = max(LatitudeLocation)+Xradius;
% Ymin = min(LongitudeLocation)-Yradius; Ymax = max(LongitudeLocation)+Yradius;
%
% Xrange = linspace(Xmin,Xmax,1e2); Yrange = linspace(Ymin,Ymax,1e2);
% map = meshgrid(Xrange,Yrange);
% Value = zeros(length(Xrange),length(Yrange));
% Value1 = Value;
%
% for i = 1:1:1e2
%     for j=1:1:1e2
%         for k = 1:1:LocationSize
%             if (abs(Xrange(i)-LatitudeLocation(k))<DistanceAverage &&
abs(Yrange(j)-LongitudeLocation(k))<DistanceAverage)
```

```
% Value(i,j) =
Value(i,j)+functionOne(Population(k),Area(k),CrimeRate(k),Xrange(i),Yrange(j),LatitudeLocation(k),Lon
gitudeLocation(k),PsycValue(k),Crimes(k));
% else Value(i,j)=
Value(i,j)+functionTwo(Population(k),Area(k),CrimeRate(k),Xrange(i),Yrange(j),LatitudeLocation(k),Lon
gitudeLocation(k),PsycValue(k),Crimes(k));
%             end;
%         end;
%     end;
% end;
% figure(1)
% meshc(Xrange,Yrange,Value);
% xlabel('Latitude');ylabel('Longitude');
% figure(2)
% surf(Xrange,Yrange,Value);
% xlabel('Latitude');ylabel('Longitude');
```

8. Murder Risk

```
function output = MurderRisk(mr,StdD,d);
dec = zeros(1,length(d));
for i=1:1:length(d)
    if (d(i)<0.7*StdD)
        dec(i)=0;
    else dec(i) = (StdD./sqrt(d(i)));
    end;
end;
output = exp(mr)+dec;
```

9. Normalize

```
function output = normalize(x);
output = x./sum(x);
```

10. Prob Calc

```
LongitudeDegree = [79 80 78 79 81];
LongitudeMinute = [49 54 52 24 35];
LatitudeDegree = [40 40 38 39 39];
LatitudeMinute = [88 04 26 24 16];

LongitudeLocation = location(LongitudeDegree,LongitudeMinute);
LatitudeLocation = location(LatitudeDegree,LatitudeMinute);

Population = [3900 9057 44015 1930 6660];
Area = [31.5 2.2    17.6 2.1 3.5];
Density = [126.4 2354.2    2559.0 915.7 1889.6];
```

```
Crimes = [23 74    1366 59 139];
CrimeRate = Crimes*100./Population;
SizeOfSample = length(Population)-1;
TotalSize = length(Population);


PopulationSample = (Population(1:SizeOfSample));
PopMean = mean(PopulationSample);
AreaSample = (Area(1:SizeOfSample));
AreaMean = mean(AreaSample);
DensitySample = (Density(1:SizeOfSample));
DensityMean = mean(DensitySample);
CrimeRateSample = (CrimeRate(1:SizeOfSample));
CrimeRateMean = mean(CrimeRateSample);

NormPopulationSample = normalize(PopulationSample);
NormAreaSample = normalize(AreaSample);
NormDensitySample = normalize(DensitySample);
NormCrimeRateSample = normalize(CrimeRateSample);

M = [NormPopulationSample' NormAreaSample' NormDensitySample' NormCrimeRateSample'];
N = 1e4*cov(M);
CovDetermin = zeros(1,SizeOfSample);
x = 1:1:SizeOfSample;
X = 1:1:TotalSize;

for i=1:1:SizeOfSample
    CovDetermin(i) = N(i,i);
end;
minStd = min(CovDetermin);
minPos = ((CovDetermin-min(CovDetermin))==0);


[PopulationP,PopulationS,PopulationMu] = polyfit(x,PopulationSample,1);
[AreaP,AreaS,AreaMu] = polyfit(x,AreaSample,1);
[DensityP,DensityS,DensityMu] = polyfit(x,DensitySample,1);
%[CrimeRateP,CrimeRateS,CrimeRateMu] = polyfit(x,CrimeRateSample,1);

[populationP,PopulationS,PopulationMu] = polyfit(x,NormPopulationSample,1);
[areaP,AreaS,AreaMu] = polyfit(x,NormAreaSample,1);
[densityP,DensityS,DensityMu] = polyfit(x,NormDensitySample,1);

beta1 = 1/minStd;
beta0 = -(beta1*CrimeRateMean);
```

```
for i = 1:1:TotalSize
z(i) = beta0+beta1*CrimeRate(i);
end;
Ps = normalize(LogisticRegression(z));

StdVector = [std(PopulationSample),std(AreaSample),std(DensitySample)];
PopPredictMean = polyval(PopulationP,TotalSize);
AreaPredictMean = polyval(AreaP,TotalSize);
DensityPredictMean = polyval(DensityP,TotalSize);
for i = 1:1:TotalSize
  p1(i)=normpdf(Population(i),PopPredictMean,StdVector(1));
  p2(i)=normpdf(Area(i),AreaPredictMean,StdVector(2));
end;
Pc = normalize(normalize(p1)+normalize(p2));


RVMean = mean(DensitySample(SizeOfSample/2:end));
for i=1:1:TotalSize
p3(i) = normpdf(Density(i),RVMean,StdVector(3));
end;
Pr = normalize(p3);
PP = Pr*0.4 + Pc*0.3+ Ps*0.3;

RiskIntensity = RobberyRisk(CrimeRate);
PR = normalize(RiskIntensity);
P = (PR+PP)/2;

Distances = zeros(TotalSize);
for i = 1:1:TotalSize,
    for j = 1:1:TotalSize,

Distances(i,j)=sqrt((LatitudeLocation(i)-LatitudeLocation(j))^2+(LongitudeLocation(i)-LongitudeLocation
(j))^2);
    end;
end;
MeanDistance = mean(mean(Distances));
StdDistance = mean(std(Distances));
for i = 1:1:TotalSize
    Next = sqrt((LatitudeLocation(i)-LatitudeLocation(SizeOfSample)).^2 +
(LongitudeLocation(i)-LongitudeLocation(SizeOfSample)).^2);
    if (Next>MeanDistance)
    NextDistance(i) = Next;
    elseif (Next==0)
        NextDistance(i)=-MeanDistance;
```

```
        elseif(Next<0.7*MeanDistance)
                NextDistance(i)= 0;
        else NextDistance(i)=MeanDistance;
        end;
end;

DP = normpdf(NextDistance,MeanDistance,StdDistance);
P = P.*DP;
p = normalize(P);
```

11.Robbery

```
LongitudeDegree = [79 80 79 78 81];
LongitudeMinute = [49 54 24 52 35];
LatitudeDegree = [40 40 39 38 39];
LatitudeMinute = [88 04 24 26 16];

LongitudeLocation = location(LongitudeDegree,LongitudeMinute);
LatitudeLocation = location(LatitudeDegree,LatitudeMinute);

Population = [3900 9057 1930 44015 6660];
Area = [31.5 2.2 2.1 17.6 3.6];
Density = [126.4 2354.2 915.7 2559.0 1889.6];
Crimes = [23 74 59 1366 139];
CrimeRate = Crimes./Population;
PsycValue = [1 1 1 1 1];
H = [1 1 1 1 1];

LocationSize = length(LongitudeLocation);
Distances=zeros(LocationSize);

for i = 1:1:LocationSize
        for j = 1:1:LocationSize
                Distances(i,j) =
sqrt((LongitudeLocation(i)-LongitudeLocation(j)).^2+(LatitudeLocation(i)-LatitudeLocation(j)).^2);
        end;
end;

DistanceAverage = mean(mean(Distances));
Xradius = max(LatitudeLocation)-min(LatitudeLocation);
Yradius = max(LongitudeLocation)-min(LongitudeLocation);
Xmin = min(LatitudeLocation)-Xradius; Xmax = max(LatitudeLocation)+Xradius;
Ymin = min(LongitudeLocation)-Yradius; Ymax = max(LongitudeLocation)+Yradius;

Xrange = linspace(Xmin,Xmax,1e2); Yrange = linspace(Ymin,Ymax,1e2);
```

```
map = meshgrid(Xrange,Yrange);
Value = zeros(length(Xrange),length(Yrange));
Value1 = Value;

for i = 1:1:1e2
    for j=1:1:1e2
        for k = 1:1:LocationSize
            if (abs(Xrange(i)-LatitudeLocation(k))<DistanceAverage && abs(Yrange(j)-LongitudeLocation(k))<DistanceAverage)
                Value(i,j) = Value(i,j)+functionOne(Population(k),Area(k),CrimeRate(k),Xrange(i),Yrange(j),LatitudeLocation(k),LongitudeLocation(k),PsycValue(k),Crimes(k));
            else Value(i,j)= Value(i,j)+functionTwo(Population(k),Area(k),CrimeRate(k),Xrange(i),Yrange(j),LatitudeLocation(k),LongitudeLocation(k),PsycValue(k),Crimes(k));
            end;
        end;
    end;
end;
figure(1)
meshc(Xrange,Yrange,Value);
xlabel('Latitude'); ylabel('Longitude');
figure(2)
surfc(Xrange,Yrange,Value);
xlabel('Latitude'); ylabel('Longitude');
```

12. Robbery Risk

```
function output = RobberyRisk(br);
output = exp(br);
```

13. Simulation

```
Home = 'Volusia';

County = str2mat('Volusia', 'Marion',    'Citrus',    'Pascal', 'Marion', 'Marion');
t = linspace(0,2*pi,1e5);
LocationCalc = @(Degree, Minute) Degree+Minute./60;
CircleCalcX = @(x,r) x+r*cos(t);
CircleCalcY = @(y,r) y+r*sin(t);

CountyLocationYDegree = [29 29 28 28 29 29];
CountyLocationYMinute = [10 16 53 19 16 16];
CountyLocationXDegree = [81 82 82 82 82 82];
CountyLocationXMinute = [31 07 31 20 07 07];
Population = [498036 329628 141416 471028 329628 329628];
```

```
CriminalRate=[0.57 1.35 0.26 0.43 1.35 1.35];
XLocation = LocationCalc(CountyLocationXDegree, CountyLocationXMinute);
YLocation = LocationCalc(CountyLocationYDegree, CountyLocationYMinute);

LocationSize = length(XLocation);
Distances = zeros(LocationSize);


for i = 1:1:LocationSize,
    for j = 1:1: LocationSize,
        Distances(i,j)=sqrt((XLocation(i)-XLocation(j))^2+(YLocation(i)-YLocation(j))^2);
    end;
end;

MaxDistance = max(max(Distances));
Xmin = min(XLocation)-MaxDistance; Xmax = max(XLocation)+MaxDistance;
Ymin = min(YLocation)-MaxDistance; Ymax = max(YLocation)+MaxDistance;
Xrange= linspace(Xmax,Xmin,1e2);
Yrange= linspace(Ymin,Ymax,1e2);
Xaverage=(Xmin+Xmax)/2;
Yaverage=(Ymin+Ymax)/2;
MaxRange=max(Ymax-Ymin,Xmax-Xmin);


map = meshgrid(Xrange,Yrange);
mapValue = meshGrid(Xrange,Yrange);
SpotValue = meshGrid(Xrange,Yrange);
Xinterval = Xrange(2)-Xrange(1);
Yinterval = Yrange(2)-Yrange(1);

for i = 1:1:1e2
    for j = 1:1:1e2
      Coverage = ((Xrange(i)-XLocation).^2 + (Yrange(j)-YLocation).^2)-MaxDistance^2>0;
      mapValue(i,j) = sum(Coverage);
%          for k = 1:1:LocationSize
%
if(sqrt((Xrange(i)-XLocation(k)).^2+(Yrange(j)-YLocation(k)).^2)<=3*sqrt(Xinterval^2+Yinterval^2))
%                SpotValue(i,j) = 10;
%          else SpotValue(i,j)=mapValue(i,j);
%          end;
%        end;
    end;
  end;
```

```
% x = zeros(length(t),LocationSize);
% y = zeros(length(t),LocationSize);


figure(1)
surf(Xrange,Yrange,mapValue);
figure(2)
hold on
axis([(Xaverage-MaxRange/2) (Xaverage+MaxRange/2) (Yaverage-MaxRange/2)
(Yaverage+MaxRange/2)]);
for i=1:1:LocationSize
x = CircleCalcX(XLocation(i),MaxDistance);
y = CircleCalcY(YLocation(i),MaxDistance);
contourf(x,y)
plot(x,y);
end;
text(XLocation,YLocation,County);
plot(XLocation,YLocation,'.');
```