

# Research of Crime Prediction Technology Based on Mathematical Model

Zhu Ke and Zhang Jin\*

Department of Educational Technology, Henan Normal University, Xinxiang, Henan, 453002, China

**Abstract:** The main tasks of this study are to ascertain the geographic profile of the criminal to find out the anchor point based on the time and locations of previous crimes and to predict the potential locations of the next crime. Firstly, this paper utilizes probability and statistics method to calculate the possibility of every point becoming the criminal's anchor point, from which this paper develops two schemes—Distance Function Method and Distribution Function Method to generate the geographic profile. Then with the result of the Distance Function Method as a reference point, the criminal's anchor point could be pinpointed from the result of the Distribution Function Method. Secondly, by Multivariate Analysis Method, this paper defines the Euclidean and Manhattan Distances between the anchor point and the locations of the previous crime sites, and then, according to distribution features of these distances, selects the corresponding distribution function to get the concrete expression. Thirdly, this paper adopts Fuzzy Mathematical Method to get quantified and normalized index factors and use Analytic Hierarchy Process to compute different weights of social index factors of different areas in the region. Thus, the weight scores of every area are gotten. Finally, considering the natural factors and social factors simultaneously, this paper determines the final probability distribution of the locations of the next crime, by which this paper could predict the most likely location of the next crime.

**Keywords:** Data mining, mathematical modeling, crime prediction.

## 1. INTRODUCTION

Geographical profiling is an information management system and investigative methodology that could determine the "geographical profile" to pinpoint the most probable area of criminal anchor point or base by evaluating the locations of connected serial crimes. It's the fusion of criminology, geography, forensic psychology, cognitive mapping, mathematical modeling, statistical analysis and investigative techniques, which is initially coined by the Canadians criminology Kim Rossmo in the early 1990s [1]. Moreover, geographic profiling can be applied to any circumstance where an unidentified person is known to have carried out activity at a series of geographic points [2,3].

### 1.1 Problem Restatement

We start our journey by concretely stating what the problem is and what we wish to examine.

So far, many sophisticated geographical profiling techniques have been developed to determine the "geographical profile" of a suspected serial criminal based on the locations of the crimes. Our work is to develop a method to aid in a local police agency's investigations of serial criminals.

- First of all, we would like to develop at least two different schemes to generate a geographical profile.
- Secondly, we would like to develop a technique to combine the results of the different schemes and generate a useful prediction for law enforcement officers.

- Thirdly, we would like to provide some kind of estimate about how reliable the estimate will be in a given situation, including appropriate warnings.
- At last, we would like to write an additional two-page executive summary to the chief of police.

### 1.2 Previous Research

There are some existing methods as follows:

Spatial distribution [4]

- Estimate the anchor point with the centroid of the crime series locations
- Estimate the anchor point with the center of minimum distance from the crime locations
- Canter's Circle hypotheses:  
They have been implemented in the following soft wares:
  - CrimeStat (Ned Levine)
  - Dagnet (David Canter)
  - Rigel (Kim Rossmo)

In general, existing algorithms begin by first making a choice distance metric  $d$  [5]; they select a decay function  $f$  and construct a hit score function  $S(y)$  by computing

$$S(y) = \sum_{i=1}^n f(d(x_i, y)) = f(d(x_1, y)) + \dots + f(d(x_n, y)) \quad (1)$$

Regions with a high hit score are considered to be more likely to contain the offenders anchor point than regions with a low hit score.

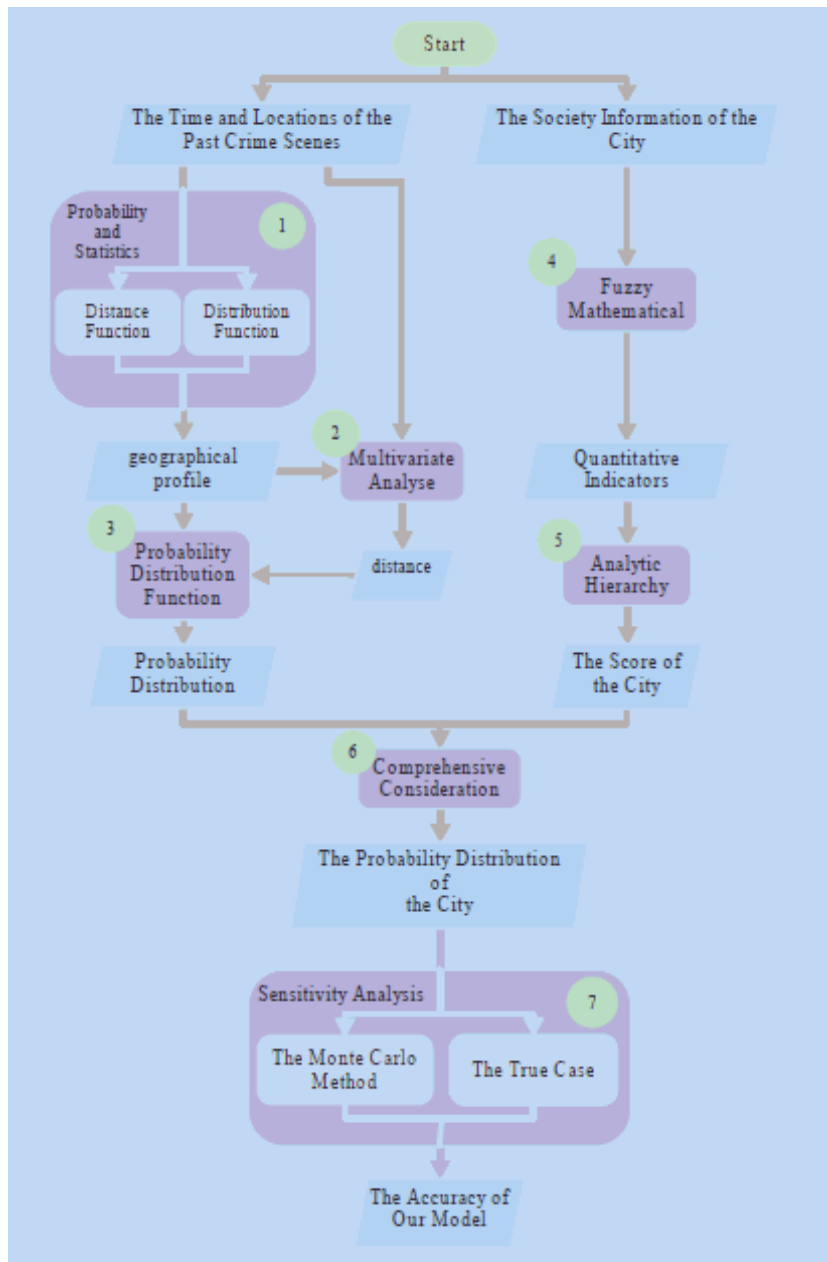


Fig. (1). The flowchart of our mathematical model start.

**1.3 Assumption**

We make the following assumptions:

1. The principles of the serial criminal selecting a crime site are constant. One principle is not too far and near.
2. The serial criminal’s anchor point is constant.
3. The area where the crime has happened is more vulnerable to next crime.
4. The factors affecting the choice on the crime site include natural and social factors. The natural factors refer to traffic conditions, terrain, the locations of the crimes, and so on. The social factors include area, population, population density, economy, race, religion, law-and-order situation, etc.

5. In the initial model traffic conditions and terrain in every two points in the region is indistinctive.

Additional assumptions are made to simply analysis for individual sections. These assumptions will be discussed at the appropriate locations.

**1.4 Notation**

We will define the following notations here as they are used widely throughout our paper. Additional notations may be defined in particular positions later.

- Region refers to the scope of all potential crime sites.
- Area refers to a divided part of the region.
- Point refers to the location of the crime.

The places surrounded by yellow line are region. The places surrounded by blue line are area. The red point refers to the point.

**2. BUILD MODEL**

We use probability statistics method, Level Analysis and simulation to study the process of generating the geographic profile and predicting the locations of the next crime. Firstly, based on the time and locations of the previous crimes, from the view of probability statistics we adopt Distance Function method and Distribution Function method to generate two geographic profiles. Then with the unique minimum point yielded by Distance Function method as a reference point, the criminal’s most likely anchor point could be pinpointed from the maximum points generated by Distribution Function method.

Secondly, having known the coordinates of the most likely anchor point and the locations of the previous crimes, we can get the Euclidean distances and the Manhattan distances between the anchor point and the locations of the crimes.

Thirdly, by analyzing distribution features of the previous crime sites, we choose an appropriate distribution function and then use the distances solved to compute the detail expression of this function. Thus, the probability distribution of the potential crime sites in the region is determined.

Next, by vague mathematics method, social factors affecting the distribution of the potential crime sites can be quantified in every area. To follow, we utilize Level Analysis to assign every area’s weight based on quantified social index factors and get every area’s weight score.

Considering natural and social factors simultaneously, we gain the final distribution of crime incidence rates in every area. Then from this distribution the locations of the next crime can be predicted.

Then, we use Monte Carlo Method and the true cases to test the sensitivity of the model.

At last, based on this model, we generate a Graphical User Interfaces (GUI) program which can be run in MATLAB. Thus, readers can apply this model to practical problems.

**2.1. One Scheme to Generate a Geographical Profile**

Before building model, we make an explanation that a point discussed represents a small area around the certain coordinates other than an abstract point of no size in geometric meaning [6].

This scheme is called Distance Function method. The basic principle is that the sum of the squares of the distances between the criminal’s anchor point and the locations of the crimes is smallest, which is caused by the assumption that one of the principles of the criminal selecting the crime sites is not too far and near.

At first, the locations of the previous crimes could be established:

$$D(x, y) = \sum_{i=1}^n [(x - x_i)^2 + (y - y_i)^2] \tag{2}$$

The point where the value of the distance function is smaller is more related to the previous crime sites. So the possibility that this point is the criminal’s anchor point is larger. Let

$$\frac{\partial D(x, y)}{\partial x} = 0, \frac{\partial D(x, y)}{\partial y} = 0 \tag{3}$$

and then get

$$\begin{cases} x = \frac{\sum_{i=1}^n x_i}{n} \\ y = \frac{\sum_{i=1}^n y_i}{n} \end{cases} \tag{4}$$

Which is denoted  $(x_0, y_0)$ .

Assumed the n crime site is indistinctive and the weights of the crime sites are equal, this point is just the "center of mass" of the locations of the n crimes.

Meanwhile, because

$$\begin{cases} \frac{\partial^2 D}{\partial x^2} \Big|_{(x_0, y_0)} > 0 \\ \frac{\partial^2 D}{\partial y^2} \Big|_{(x_0, y_0)} > 0 \\ \left[ \frac{\partial^2 D}{\partial x \partial y} \Big|_{(x_0, y_0)} \right]^2 - \frac{\partial^2 D}{\partial x^2} \Big|_{(x_0, y_0)} \frac{\partial^2 D}{\partial y^2} \Big|_{(x_0, y_0)} < 0 \end{cases} \tag{5}$$

The point  $(x_0, y_0)$  is also the minimum point of the distance function, which means the possible that this point is the criminal’s anchor point, is largest.

We use MATLAB to count the distance function of every point in this region and make different colors show different values of the distance function. Thus a geographic profile is generated. Obviously, this geographic profile has one and only one minimum point.

To sum up, the possible that this minimum point is the criminal’s anchor point is largest and this point is just the "center of mass" of the locations of the previous n crimes.

**2.2. The Other Scheme to Generate a Geographic Profile**

This scheme is named Distribution Function method, which considers the time and space simultaneously. The basic principle is to make the midpoint of the straight line joining two continual crime sites as an origin, and the probability of the anchor point whose distance from the origin is  $r$  observes certain distribution function with a variable  $r$ .

Assumed the n crime sites have been arranged in chronological order, we firstly make the midpoint of the straight

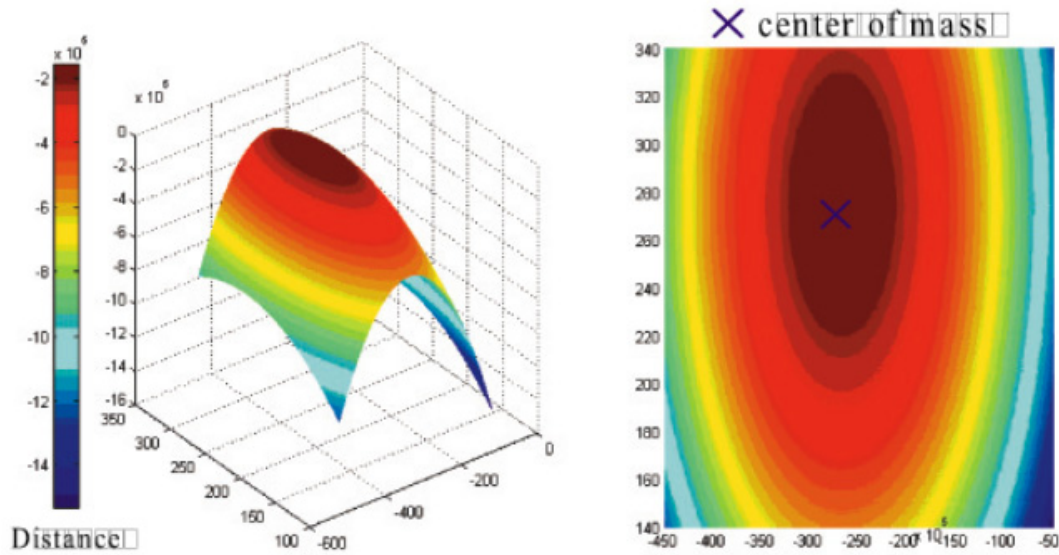


Fig. (2). The process of distance function method.

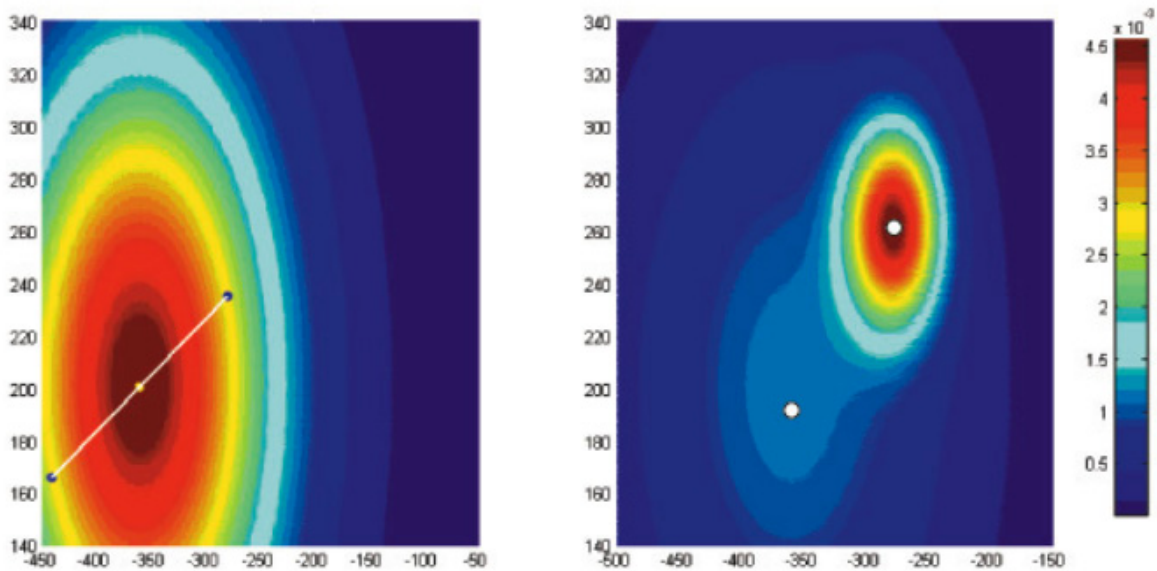


Fig. (3). The generated geographical profile based on distribution function method.

line joining  $i$  th and  $i+1$  th crime sites as the origin  $O_i$  and the probability of the anchor point whose distance from the origin is  $r$  observes certain distribution function  $P(r)$ . Then according to the research on psychological characteristics of the criminal and the locations of the  $i$  th and  $i+1$  th crime sites, we can gain  $O_i$ 's concrete distribution function  $P(r)$ . Similarly, the other origins' concrete distribution function could be gotten. Then summing these  $n-1$  probability distributions, we could get a geographic profile.

Usually, the possibility of the midpoint of the straight line joining two continual crime sites becoming the anchor point is largest. Then with the increase of the distance from the origin, the possibility to be the anchor point decreases and at the two crime sites the possibility is nearly reduced to 0. This is just a normal distribution:

$$P_N(r) = \begin{cases} 0, & r > R \\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right), & r \leq R \end{cases} \quad (6)$$

In which  $\sigma = d/2$ .  $d$  represents the distance of the two crime sites.

Different colors in the geographic profile stand for the different possibility. This geographic profile has at least one maximum point in general, where the percentage is largest.[8]

At first, we analyze the two schemes generating geographic profiles. The first scheme doesn't take the time factor into consideration, so there will be certain limitations. However, this scheme yields an only point which is most likely to be the criminal's anchor point. So this point could

be regard as a reference point. Then, the second scheme combines the time and space factors, but its maximum points are not unique. Thus we use the reference point generated in the first scheme, count the distances between the reference point and maximum points yielded by the second scheme, and then choose the point yielded by the second scheme whose distance from the reference point is smallest as the most likely anchor point.

**Definition of Distance** We let Center Point denote the most likely anchor point whose coordinates are expressed in  $(X, Y)$  which have been known. Then by multivariate analysis build two models to depict the distances between the locations of the crimes and Center Point.

- The Euclidean distance:

$$r_i = \sqrt{(x_i - X)^2 + (y_i - Y)^2} \tag{7}$$

- The Manhattan distance:

$$r_i = |x_i - X| + |y_i - Y| \tag{8}$$

Ignoring any factors, the Euclidean distance computes straight-line distance between the two places, which is easy to understand. Considering most of streets in cites is approximately distributed along the directions of the latitude and longitude lines, the Manhattan distance which computes the distance between the two places is more realistic. Overall, they adapt to different conditions. For example, when the region's traffic conditions are good, we adopt the form. Otherwise, we use the latter.

Despite the differences between two distance formulas, the later operation is fully consistent. For simplicity, we define  $r_i$  either are thought of as the Euclidean distance, but also serves as the Manhattan distance. The Most Probable Radius  $r_0$  now every crime site  $(x_i, y_i)$  has a corresponding known distance  $r_i$ . Based on the assumption that the area where the crime has happened is more vulnerable to next crime, we adopt least square method to ascertain the most probable radius  $r_0$ , that is, the minimum of  $r$  that satisfies

$$f(r) = \sum_{i=1}^n (r - r_i)^2 \tag{9}$$

Then make a derivative of  $r$  on both sides, we can get:

$$\frac{df}{dr} = 2 \sum_{i=1}^n (r - r_i) \tag{10}$$

Let

$$\frac{df}{dr} = 0 \tag{11}$$

And obtain

$$r = \frac{\sum_{i=1}^n r_i}{n} \tag{12}$$

Thus, because of

$$\frac{d^2 f}{dr^2} = 2n > 0 \tag{13}$$

The most probable radius  $r_0$  is the average of the total  $r_i$ . Distribution Function of Crime Incidence Rate First of all, we assume crime incidence rates of the points whose distances from Center Point are equal are same. (We will improve this assumption later.) Dividing  $r$  in appropriate interval and counting the number of previous crimes in every interval, we could draw the corresponding image of the relationship between the frequency and the interval. Based on the shape of the image, we can select suitable distribution function, e.g. Rayleigh distribution, gamma distribution, lognormal distribution, F distribution; we bull distribution, to fit.

For instance, we choose Rayleigh distribution:

$$P_R(r) = \begin{cases} \frac{r}{\mu^2} \exp^{-\frac{r^2}{2\mu^2}}, & r \geq 0 \\ 0, & r < 0 \end{cases} \tag{14}$$

Whose mean value is  $\sqrt{\frac{\pi}{2}}\mu$ , then make  $r_0 = \sqrt{\frac{\pi}{2}}\mu$ , and

compute  $\mu$ . Thus, the concrete expression of distribution function, that is, the probability density of crime incidence rate in the region, is known. Consider Different Weights of Different Areas The assumption that crime incidence rates of the points whose distances from Center Point are equal are same is flowed. Considering social factors around Center Point, we adopt vague mathematics and Level Analysis to calculate weight scores of how the criminal selects different areas in the region to commit the crime.

At first, by vague mathematics method, social factors affecting the distribution of crime incidences rate can be quantified in every area. Then from top to bottom the problem is divided into goals layer, guidelines layer and programs layer. Goal layer is the priority areas where the criminal choose to commit the crime; guideline layer is social factors considered in every area, such as area, population, population density, economy, race, religion, law-and-order situation, etc. and we assume there are  $n$  index factors; program layer is all potential crime areas in the region and the number of the areas is assumed  $m$ .

Firstly, quantified index factors by Vague mathematics could are standardized or nondimensionalized. The common methods include average method, the initial value method, the standard deviation method, extreme value difference methods, efficacy coefficient method, and so on. We could select appropriate method according to the specific data types. For example, to population data, assumed that the population of  $i$ th area is  $\mu_i$ , the proportion of total population in this area

$$v_i = \frac{\mu_i}{\sum_{i=1} m\mu_i} \tag{15}$$

is regarded as a standardized population index in the area.

Secondly, construct a pairwise comparison judgment matrix  $A$  to count weight coefficients of every index factor in all factors. The details are that based on the previous crimes record, we pairwise compare all factors to get the importance of every index factor to others, which is the element of pairwise comparison judgment matrix  $a_{ij} = \frac{\omega_i}{\omega_j}$ . For instance,

the criminal is apt to offend in the area of large population, but to victims' race there is no particular demand. So the relative importance of population to race can be denoted 6: 1, while the relative importance of race to population is 1: 6. Thus generated pairwise comparison judgment matrix is a positive and negative matrix. The concrete form is

$$A = \begin{pmatrix} \omega_1 & \dots & \omega_1 \\ \omega_1 & & \omega_n \\ \vdots & \ddots & \vdots \\ \omega_n & \dots & \omega_n \\ \omega_1 & & \omega_n \end{pmatrix} \tag{16}$$

Thirdly, this matrix's eigenvalue and corresponding eigenvector can be counted. Then the eigenvector of the largest eigenvalue is normalized to be the weight of every index factor in all factors  $D = (d_1, d_2, \dots, d_n)^T$ . In addition, we need to test the consistency of the index factors. If the matrix couldn't pass the test, we need to reconstruct a matrix.

Finally, to the area  $j$ , according to the weight coefficients of index factors solved, we sum all normalized index data weighted to gain the final weight score in this area  $p_j = \sum_{i=1}^n c_{ji}d_i$ . Then we could get the weight score of every area in the region, that is,

$$\begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix} = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \dots & c_{mn} \end{pmatrix} \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} \tag{17}$$

These scores represent relative percentages of the criminal committing the crime in the areas. Predict the Possible Locations of the Next Crime Until now, to any point  $(x, y)$  in the region we could count its distance  $r$  from Center Point, then introduce  $r$  into distribution function solved, and get the percentage of next crime will happen at the point  $P(r)$ , which only considers natural conditions.

To follow, we add social factors of related weight coefficients in different areas. The point  $(x, y)$  in the region must belong to certain area. We define  $P_j$  is this area's relative percentage weight coefficient, then  $P(r)P_j$  represents the relative percentage of next crime will happen at the point  $(x, y)$ . We use this method to compute every point in the region and draw a map, from which the points whose percentages are relatively large could be pinpointed. The pinned point is likely to the location of next crime.

### 3. APPLICATION OF THE MODEL

In order to test the feasibility of the model, we apply it to the case of Peter Sutcliffe.

#### 3.1. Obtain the Data of Peter Sutcliffe's Criminal Records

Firstly, we obtain the material of Peter Sutcliffe's crimes from the chart of attacks and murders in The Yorkshire Ripper Website, including the date, date interval, time, town, place, the name and age of the victim, and outcome. From all the records we eliminate the 1st and 17th cases whose material was not reliable. The rest 21 criminal records are counted [9].

Secondly, we download the Google Earth/Maps overlay showing significant locations in the Ripper case, which lists the information of Peter Sutcliffe's 20 criminal records, including the specific geographic latitude and longitude coordinates of the crime sites. We carefully examine this data and the former chart in The Yorkshire Ripper Website for comparison, and collect other information from The Yorkshire Ripper Website Wikipedia and The Real Yorkshire Ripper Website [10]. Then we eliminate two cases in the Victims Homes folder of Google Earth/Maps, modify the latitude and longitude coordinates of the place where Barbara Leach was murdered, and add the latitude and longitude coordinates of the places where Marcella and Theresa Sykes were attacked. At last all the latitude and longitude information is consistent with the former chart material. So we obtain all the data of Peter Sutcliffe's criminal records successfully. It shows in the Appendix.

Introduce this case to our model and the notations in the model will have practical significance.

- Region refers to West Yorkshire
- Areas refer to City of Bradford, Calderdale, Kirklees, City of Wakefield and City of Leeds.
- Point refers to the concrete location in West Yorkshire.

To compute easily, we regularize the shapes of West Yorkshire, City of Bradford, Calderdale, Kirklees, City of Wakefield and City of Leeds.

#### 3.2. Distribution Function

Before choosing the distribution function, we observe the distribution of Peter's crime sites.

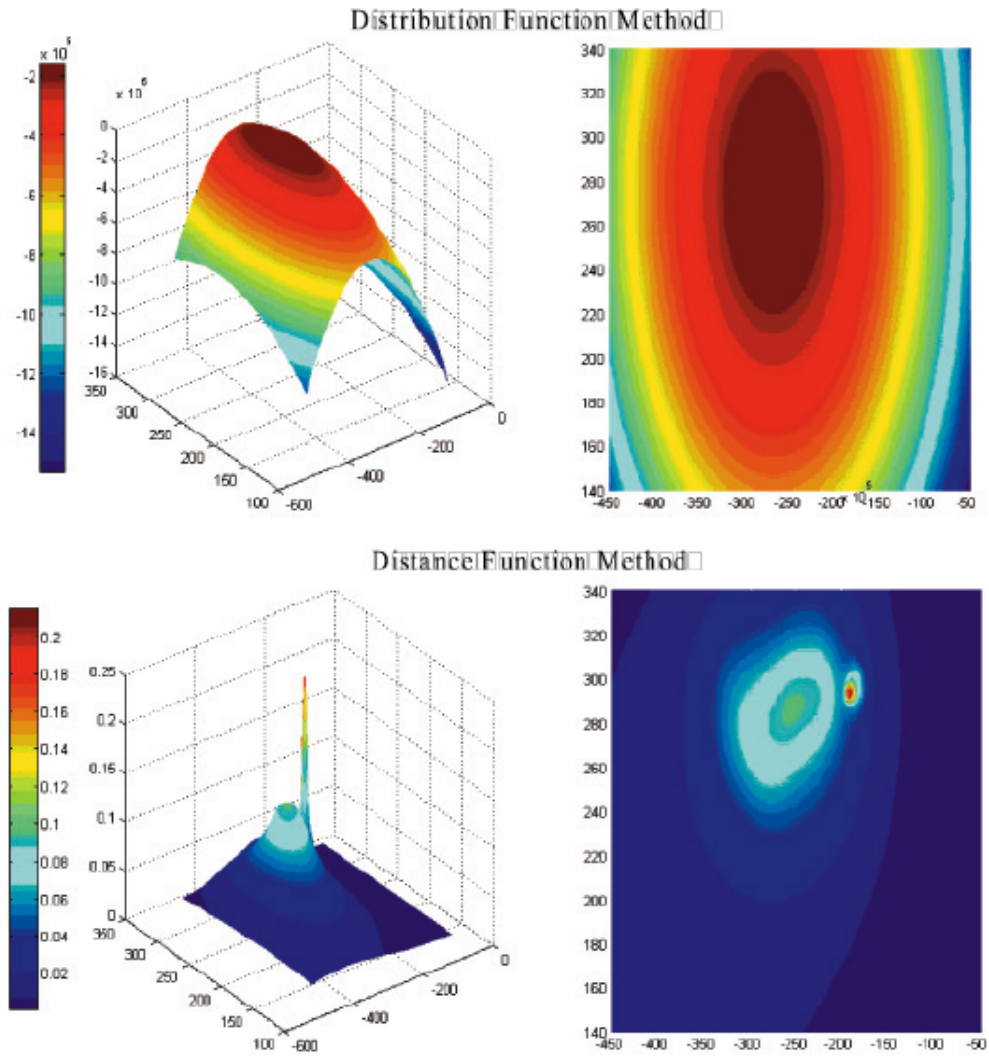


Fig. (4). The geographical profile of peter sutcliffe.

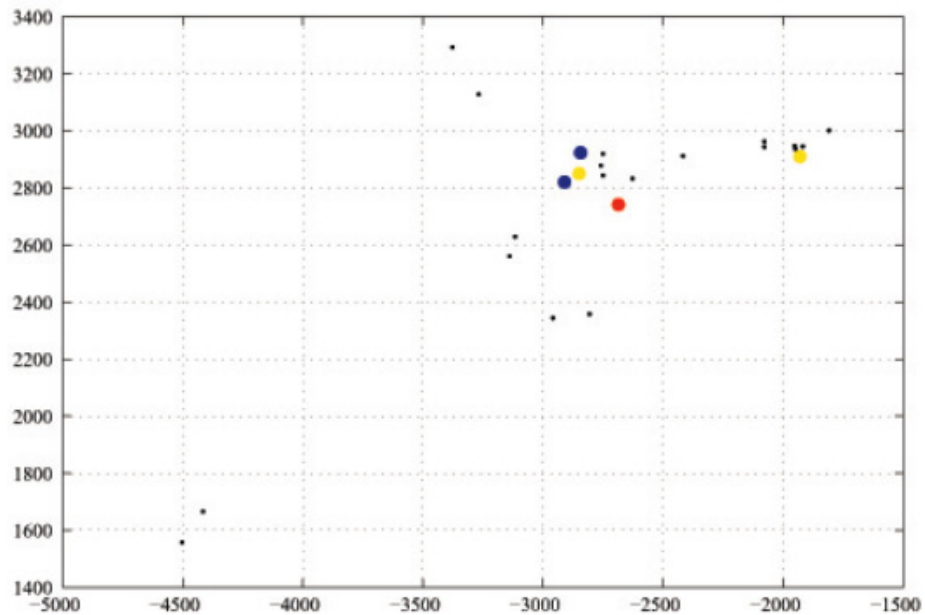


Fig. (5). The schematic drawing of the locations of the crimes and the most possible living place.

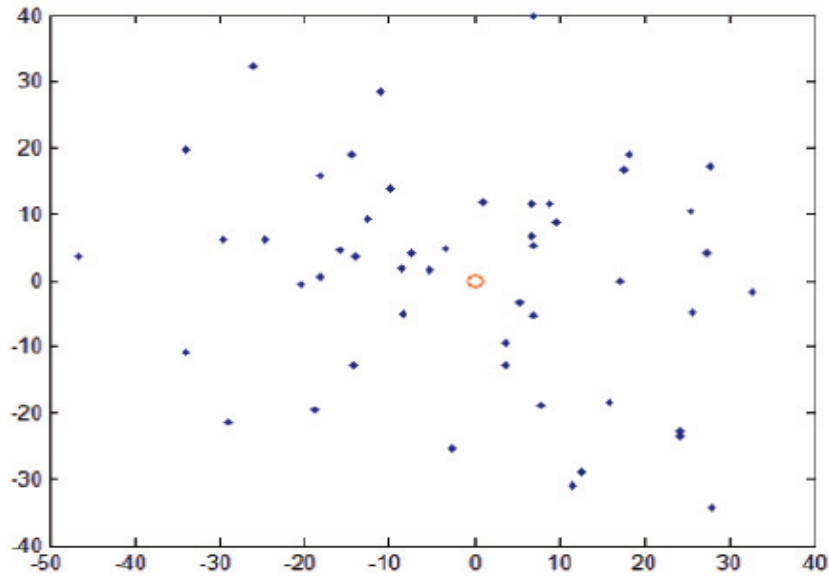


Fig. (6). The simulated points of the rayleigh distribution case.

So we adopt Rayleigh distribution. Based on the average distance between the crime sites and Center Point, the expression of Rayleigh distribution is:

$$P_R(r) = \begin{cases} \frac{r}{\mu^2} \exp \frac{-r^2}{2\mu^2}, & r \geq 0 \\ 0, & r < 0 \end{cases} \quad (18)$$

in which  $\mu=73.27$  in Manhattan distance and  $\mu=55.96$  in Euclidean distance. Based on this expression, regardless of social factors, the probability distribution of the crime incidence rates in the West Yorkshire is determined.

### 3.3. Analytic Hierarchy Process

At first, we use average method to nondimensionalize all quantified index factors. The result is here.

According to Peter’s crimes record to construct a pairwise comparison judgment matrix:

$$A = \begin{pmatrix} 1.0000 & 0.6250 & 1.5000 & 5.0000 & 3.0000 \\ 1.6000 & 1.0000 & 2.0000 & 4.0000 & 2.0000 \\ 0.6667 & 0.5000 & 1.0000 & 4.0000 & 2.0000 \\ 0.2222 & 0.1250 & 0.2500 & 1.0000 & 0.5000 \\ 0.3333 & 0.2500 & 0.5000 & 2.0000 & 1.0000 \end{pmatrix}$$

Next, this matrix’s largest eigenvalue is  $\lambda = 5.0053$ . Upon examination, the consistency ratio is

$$CR = \frac{CI}{RI} = 0.001183 \ll 0.1 \quad (19)$$

Thus this matrix is reasonable. And the corresponding eigenvector normalized is

$$D = (0.2670 \quad 0.3966 \quad 0.1912 \quad 0.0496 \quad 0.0956)^T \quad (20)$$

Which are the weights of every index factors? Thus, the final weight scores of these five areas are calculated, which

could be regarded as relative crime incidence rates in these area.

### 4. CONCLUSIONS AND FUTURE WORK

Through the established model and the following implementation methods, we can either find out a serial criminal’s most likely anchor point based on the time and locations of the previous crimes, but also predict the potential locations of the next crime. Thus, on the one hand, the police could focus investigation on serial criminals’ possible hiding places and take the initiative to arrest them as soon as possible. On the other hand, they can strengthen vigilance on the potential locations against possible trouble. However, there are also certain limitations when our model is applied to actual cases. In the following discussion, we will carry on further detailed explanation.

Firstly, the police need to gain the precise time and locations of the previous crimes. The precise location is pinpointed in a coordinate system. If the serial crime’s region is not too large, the region could be viewed as planar domain, and the longitude and latitude could be used as the coordinates. Next, if the exact time of the crimes is unknown, we at least need to know the time order of the crimes.

Secondly, in the simplest condition, that is, only considering natural factors of the locations of the crimes, we count the sum of the squares of the distances between any point in the region and the crime sites to find out the point whose value is smallest. This point is regarded as a reference point of the most likely anchor point. This scheme is to adopt Distance Function Method to generate the geographic profile. If considering social factors such as the condition of economic, population density, etc. in different areas, the corresponding weight scores of different areas must be multiplied. We will discuss the details later. The other method to generate the geographic profile is Distribution Function Method. We firstly make the midpoint of the straight line joining 1st and 2nd crime sites as center and the distance of the two sites as diameter to draw a circle. Assumed that the criminal’s an-



chor point inside this circle obeys certain distribution, e.g. Normal Distribution, and outside this circle the probability of the anchor point is 0, then a distribution color map is created. Similarly, the circles and the corresponding distribution color maps of 2nd and 3rd crime sites and others can be made. Thus a geographic profile is formed, which has at least a maximum point. Compare the distances between the maximum points and the reference point. Then the point nearest to the reference point is the most likely anchor point. The police could pay more attention to this location. In the process of determining anchor point, we recommend observing generated geographic profiles carefully. If the profiles are sharp and have evident peak values, the solutions are credible; however, if the profiles are gentle and the differences are small, the solutions may be incredible. For the former case the criminal is likely to have an anchor point. But for the latter the criminal may be mobile and an anchor point couldn't be determined. In this case, we advise the police to focus on the prediction of the next crime.

Thirdly, count the distances between the previous crime sites and the anchor point. We use two methods to compute the distances: The Euclidean Distance and the Manhattan Distance. When the region's traffic conditions are good, we adopt the former. Otherwise, we use the latter. Then based on the distribution of the distances, we select the corresponding distribution function and get the concrete expression. Thus we get the distribution of crime incidence rate in the range.

Next, consider social factors affecting criminals' choice on the locations of the crime. For example, if a criminal is apt to commit crimes in one area where there is more White and the economic is more developed, we need to increase the crime incidence rates in the corresponding area. At first, according to the previous crime records, we could find out some index factors, utilize Fuzzy Mathematical Method to quantify and normalize the index factors in different areas, and then use Analytic Hierarchy Process to get different areas weight scores. This need to construct a pairwise comparison judgment matrix. Based on the previous crimes record, we pairwise compare all factors to get the importance of every index factor to others, which is the element of the matrix. Then we compute the corresponding eigenvector when the eigenvalue is largest, and normalize the eigenvector to get the weights of every index factor to others. At last, we sum the weighted index factors in every area to gain the final weight scores that reflect the crime incidence rates of different areas.

To follow, considering the social factors and the natural factors simultaneously, we get the final distribution map of the crime incidence rates in the range. From this distribution map the points whose percentages are relatively large could be pinpointed. The pinned point is likely to the locations of the next crime. So the police can strengthen vigilance in these points.

At last, based on the built model, we generate a Graphical User Interfaces (GUI) program, which could be run in MATLAB. The instruction of this program is in Appendix. Thus, readers could apply this model to the real problems easily.

In addition, if the criminal has no anchor point, even if the criminal has anchor point, there are no particular principles when he or she chooses the locations to commit crimes, or the criminal has no evident purpose or objective to commit crimes, our model is inappropriate. An improved scheme is not to emphasize finding an anchor point but to utilize the previous time and locations of the crimes to line different crimes sites to a directed polyline in chronological order. Then consider the time interval in every two crimes to predict the potential locations of the next crime by Random Walk Model. This will be our research in future.

### CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

### ACKNOWLEDGEMENTS

This work is supported by Science and technology research projects of Educational Commission of Henan Province (No. 14A880019).

### REFERENCES

- [1] LIU, B., HSU, W. AND MA, Y. "Integrating classification and association rule mining." *Knowledge Discovery and Data Mining*, 80-86,2010.
- [2] Ozaki, K., & Toyoda, H. "Paired comparison IRT model by 3-value judgment: estimation of item parameters prior to the administration of the test." *Behaviormetrika*, 33(2), 131-147,2006.
- [3] Reynolds, P. A., Rice, S., & Uddin, M. "Online learning in dentistry: the changes in undergraduate perceptions and attitudes over a four year period." *British Dental Journal*, 203(7), 419-423,2007.
- [4] *Mathematical Statistics and Probability*. Vol 1, California, USA pp. 281-297,2011.
- [5] Davies, J., Grobelnik, M., & Mladeni c, D. "Semantic knowledge management: Integrating ontology management, knowledge discovery, and human language technologies." *Springer*,2009.
- [6] Meyer, S. M., Degener, J., Giannandrea, J., & Michener, B. (2010). "Optimizing schema-last tuple-store queries in graphd." *In Proceedings of the 2010 international conference on Management of data, SIGMOD'10* .pp. 1047-1056. New York, USA: ACM, 2010.
- [7] Tallent-Runnels, M., Thomas, J., Lan, W., Cooper, S., Ahern, T., Shaw, S., et al. "Teaching courses online: a review of the research." *Review of Educational Research*, 76(1), 93-135,2010.
- [8] Cornell, R. M., Eining, M. M., & Hu, P. J. "The effects of process accountability on individuals' use of a familiar technology." *Journal of Information Systems*, 25(1), 109-128,2011.
- [9] S. K. Meador, et al., "Using Response Latency Within a Preference Assessment," *The Behavior AnalystToday*, vol. 8, pp. 63-69, 2007.
- [10] L. F. Hornke, "Response Time in Computer-Aided testing: A "Verbal Memory" Test for Routes and Maps," *Psychology Science*, vol. 2, pp. 280-293, 2005.