

Series Crime Prediction Research Based on Regional Weighted Models

Zhijie Sun^{1,2,*}, Xueyan Zhang¹

¹ Policing Information Department, Shandong Police College, Jinan, China, 250200

² Research Team of Science and Technology Application Innovation on Public Security, Shandong Police College, Jinan, China, 250200

* Corresponding Author Email: szj_sdpc@126.com

Abstract. Serial criminal cases directly threaten the security of the people and even the country and society. At present, grass-roots public security work still adopts the method of "handling cases by experience", which is difficult to popularize. In view of this situation, combined with the existing scientific researches, we innovate and optimize the regional coverage weighted model on the basis of the technology of criminal geographical portrait and the high regularity of serial criminal cases. Through data fitting and geographical risk probability weighting, we have carried out geometric mapping and constructed a visual serial criminal prediction models, which effectively break through the singularity and limitations of the original crime prediction and significantly improve the prediction accuracy.

Keywords: Serial crime, Criminal geographical portrait, Region-weighted model, Geometric mapping.

1. Introduction

Serial criminal cases often have the characteristics of strong concealment and frequent occurrence. Moreover, the suspect has strong anti reconnaissance awareness and involves a variety of criminal acts. The traditional investigation model has gradually revealed its drawbacks and is difficult to handle it. At the same time, the crime in the information age also makes a lot of crime information preserved, which brings an important turning point to the detection of serial criminal cases and provides a prerequisite for the application of data-based modeling for crime prediction.

In 1981, Peter Sutcliffe, an Englishman known as the "Yorkshire butcher", was convicted of a series of malicious injuries and 13 serial murders. In the process of solving the case, it was proposed that the next scope of suspect's activities should be determined more accurately from the "centers" of these crime locations. This attempt has become the beginning of a new field. Since then, more and more technologies and methods have been created at home and abroad to determine the "geographical outline" of the suspects activities in serial crimes. Later, in order to master the distribution law of serial killers in the regional space, Fujian Police College, after recording the crime locations of the suspects for many times, obtained the criminal center of the suspects as a possible "geographical outline", according to the spatial analysis method of crime hot spots in criminology. Theoretically, this method is more suitable for the comparison of different crime types or the change of the same crime in time and space. The police can deploy more police forces in the region according to the actual situation [3]. In addition, Canadian criminologist Kim Rossmo proposed the Rossmo formula method [4], which can also be used to directly calculate the probability of criminals living in various places, generate an intuitive probability distribution map of prisoners' living points, and compare it with the standard distance circle generated by the previous method, so as to more accurately predict the possible criminal areas of criminals.

In this paper, we study the suspects' psychological situations and life preferences, depict their unique criminal patterns, and reasonably predict their criminal tracks with the help of the traditional positive detection thinking for the serial criminal cases. Through the above methods, we can roughly predict the next high-risk crime places according to the crime time and location, but the reference

factors are relatively single, and the scope of the income is broad, which to some extent can provide a reference for this paper to make a comprehensive and effective prediction of serial crime.

2. Serial crime prediction model based on region-weighted model

2.1. Model Data Preprocessing

The case is a serial crime committed by a criminal, and the crime is based on his subjective will; The criminal is completely free to commit a crime, and lives in the same place for a long time; The criminal always turn the window and enter the house. The methods and means have certain characteristics, which meet the research requirements and the application requirements of the regional risk classification of the serial criminal case prediction model based on geographical portrait technology.

Therefore, we use the 2012 YUAN Danping serial homicide case in Changsha, Hunan Province as an example to establish a model data analysis, as shown in Table 1 [6][7].

Table 1. Data sheet of Yuan Danping serial homicide

Time of crime	Guilty Place	Latitude	Longitude
2003/06/01	Zhenggong middle school in Zhangzhuangpu town, Hubei Province	29.833157	112.030169
2003/09/19	Songxi River, Zhenggongdu market town, Gonggan County, Hubei Province	29.834757	112.026738
2004/02/23	Zhenggongdu town, Gonggan County, Hubei Province	29.834757	112.026738
2005/01/11	Riverbank of group 9, Shuangxing village, near Zhenggongdu market town, Hubei Province	29.815816	112.017399
2009/12/07	Yangjia village, Nanping Town	29.778296	112.02729
2009/12/25	Xinghuo seed farm, Mengjiaxi Town, Gonggan County, Hubei Province	29.906426	112.10820
2010/09/22	Xiaodukou Town, Li County, Hunan Province	29.633714	111.92186
2011/10/24	Dongsheng Village, Ganjiachang Township, Gonggan County, Hubei Province	29.763149	112.07961
2013/03/23	Fenglin Village, Rudong Township, Li County, Changde City, Hunan Province	29.751645	111.99739

$$u_k = \sum_{j=1}^p w_{kj} x_j, \tag{1}$$

$$v_k = net_k = u_k - \theta_k, \tag{2}$$

$$y_k = f(v_k), \tag{3}$$

2.2. Establish a serial criminal prediction model based on region-weighted model

Establish a criminal geographical portrait model, and analyze the geographical portrait of the serial criminal cases by using the methods of geographic information system and criminal intelligence analysis, so as to study the characteristics and laws of the temporal and spatial distribution of the serial criminal cases and the criminal's activity track, and then judge the area where the criminal may live and the probable place where the criminal will commit the next crime [6].

The main basic theory of this model is "circle hypothesis theory". According to the above analysis, the point where the average sum of distances to each effective crime location is the smallest is the possible residence or stay of the suspect. Calculate the minimum value, calculate the radius r, and construct a circle with the minimum value point as the center and radius r, which is the circle of the crime area. According to this model, the radius of the crime area and the place where the criminal

may live or stay are calculated and predicted, and the geographical portrait is predicted on the basis of morphology [7].

Locate the location of the suspect on the map with longitude and latitude and mark the location of the suspect, then mark the location of the crime as C_1, C_2, \dots, C_M according to the time sequence of the crime.

According to the geographical portrait model, simulate the minimum average value of the sum of the distances from the residence $G(C_{0j}, C_{0w})$ to each effective crime location, so as to calculate the residence location and the minimum average distance from the crime location.

$$S = \sum_{n=1}^M \sqrt{(C_{0j} - C_{nj})^2 + (C_{0w} - C_{nw})^2} \quad (4)$$

$$R = \frac{\min S}{M} \quad (5)$$

Where, $n = 1, 2, \dots, M$, M is the total number of crime locations, C_{0j} is the longitude value of the simulated residence G , C_{0w} is the latitude value of the simulated residence G , C_{nj} is the longitude value of a crime location, C_{nw} is the latitude value of a crime location, s is the sum of the distances from the simulated residence $G(C_{0j}, C_{0w})$ to each effective crime location, and R is the minimum average distance from the residence to the crime location.

When R is the minimum value, the corresponding $G(C_{0j}, C_{0w})$ is the criminal residence.

Assume $A = S_{jj}(C_{0j}, C_{0w})$; $B = S_{jw}(C_{0j}, C_{0w})$; $C = S_{ww}(C_{0j}, C_{0w})$. Function S has a second-order continuous partial derivative in a neighborhood of point $G(C_{0j}, C_{0w})$, and $G(C_{0j}, C_{0w})$ is a stationary point, so that $AC - B^2 = \begin{vmatrix} A & B \\ B & C \end{vmatrix} > 0, A > 0$, that is, at this time, point $G(C_{0j}, C_{0w})$ is the minimum point. Let $\frac{\partial S}{\partial C_{nj}} = 0$ and $\frac{\partial S}{\partial C_{nw}} = 0$, calculate the coordinate (29.8177112.025) corresponding to the minimum point through MATLAB operation, and calculate the value of the minimum average distance R as 7.4143km, where (29.8177112.025) is the point G where the criminal may live or stay.

Figure 1 shows the crime location and the predicted point G where the criminal may live or stay.

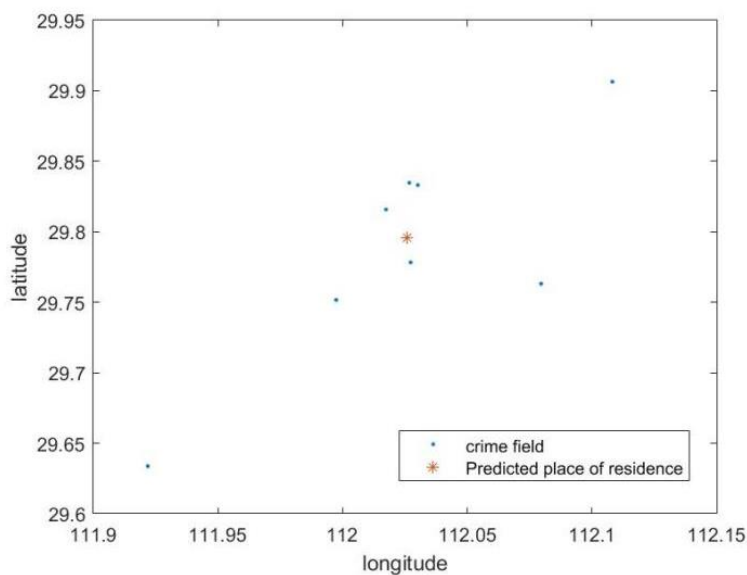


Figure 1. predicted map of residence

Calculate the residence G to any crime location C_p 's Distance r_p . Establish CGT model (Geographic target model of crime) [8] to determine the areas where criminals are likely to commit crimes. This area serves as the map range and participates in further probability calculation.

$$r = r_{p_{max}} + \frac{r_{p_{max}} - r_{p_{min}}}{2(M-1)}, p = 1, 2, \dots, M \quad (6)$$

Where r_p is the distance of residence G to any crime location C_p , and r is the radius of the crime area. Through matlab operation, the value of r is 24.8214km.

The predicted crime area with r as the radius is shown in Figure 2.

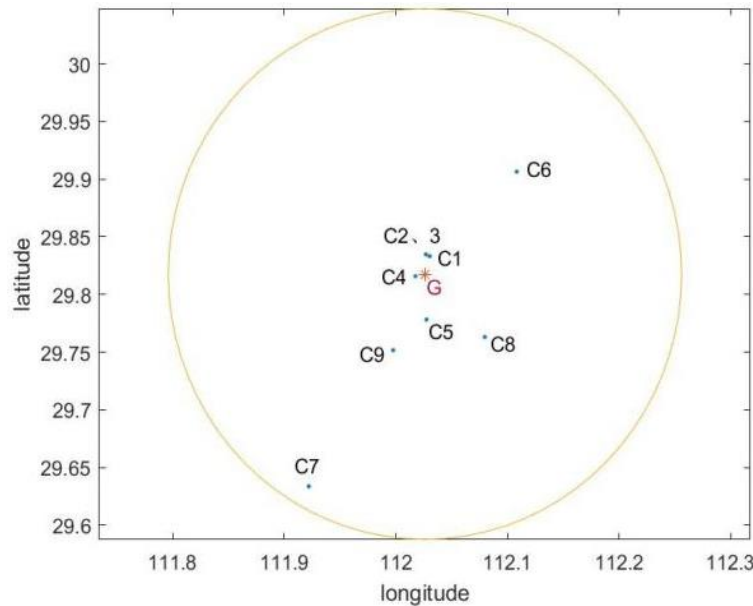


Figure 2. Predicted map of residence

3. Regional risk classification

3.1. The establishment of simulation model

Establish a regional coverage model for continuous crime prediction, analyze the known crime places in combination with statistics and crime graphics, find out the possible places of residence or stay of the criminals and crime hot areas, and determine crime zone 3, crime zone 2, crime zone 1 and crime zone 0 according to the level of crime risk [10].

Construct a crime distance matrix from any crime location to another crime location.

$$L = (L_{pq})_{M \times M} \quad (7)$$

Including: L_{pq} is the distance from C_p to C_q .

$$\begin{bmatrix} 0.0000 & 0.4202 & 0.4202 & 1.9622 & 6.0980 & 8.1391 & 25.1920 & 14.3151 & 9.7520 \\ 0.4202 & 0.0000 & 0.0000 & 2.3441 & 6.2675 & 12.0436 & 25.1698 & 9.8803 & 9.7837 \\ 0.4202 & 0.0000 & 0.0000 & 2.3441 & 6.2675 & 12.0436 & 25.1698 & 9.8803 & 9.7837 \\ 2.3905 & 2.3441 & 2.3441 & 0.0000 & 4.3070 & 14.2387 & 22.8263 & 9.0477 & 7.4612 \\ 6.0980 & 6.2675 & 6.2675 & 4.3070 & 0.0000 & 16.8207 & 19.8623 & 6.0460 & 4.4459 \\ 11.8812 & 12.0436 & 12.0436 & 14.2387 & 16.8207 & 0.0000 & 36.6627 & 16.2173 & 21.1297 \\ 25.1920 & 25.1698 & 25.1698 & 22.8263 & 19.8623 & 36.6627 & 0.0000 & 22.6501 & 15.5450 \\ 9.5134 & 9.8803 & 9.8803 & 9.0477 & 6.0460 & 16.2173 & 22.6501 & 0.0000 & 9.2153 \\ 9.7520 & 9.7837 & 9.7837 & 7.4612 & 4.4459 & 21.1297 & 15.5450 & 9.2153 & 0.0000 \end{bmatrix} \quad (8)$$

Find each crime location C_p average distance to other crime sites, in C_p is the center of the circle, with L_p is the radius to draw the crime extension area.

$$L_p = \frac{1}{M-1} \sum_{p \neq q} L_{pq}, p = 1, 2, \dots, M \quad (9)$$

Where L_p for each crime location C_p average distance to other crime sites.

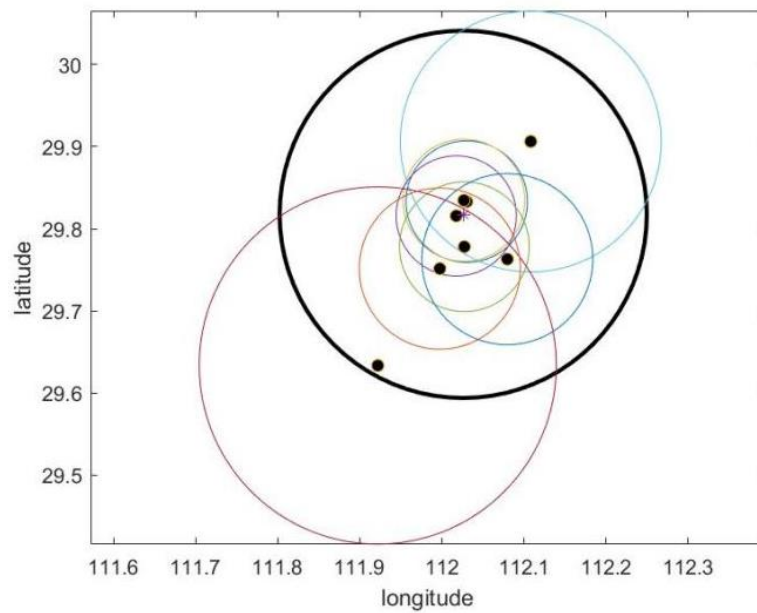


Figure 3. Schematic diagram of crime extension area

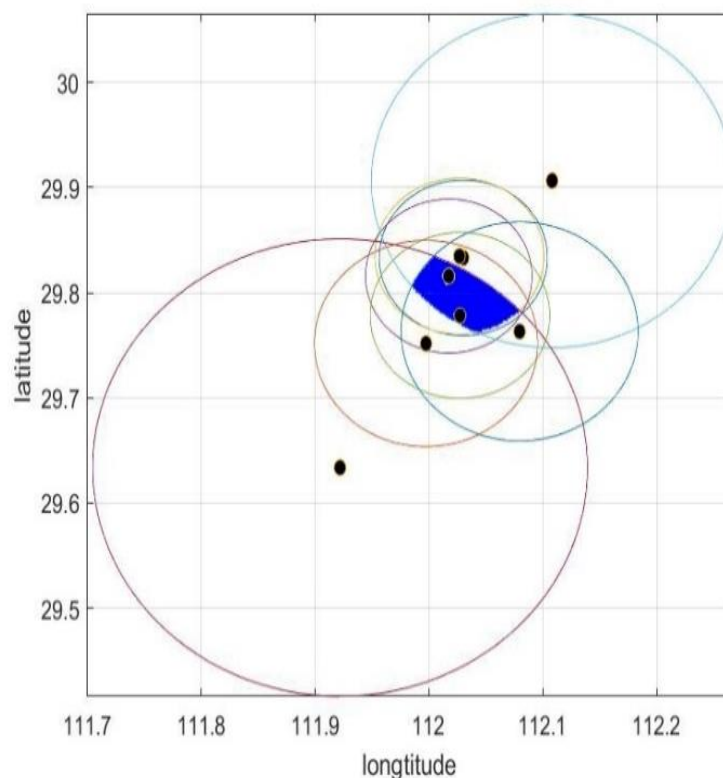


Figure 4. Schematic diagram of overlapping crime area expansion and residence

According to the characteristics of continuous crime, the area with the most overlapping times in all extended areas of crime is set as crime zone 0, and the G point should fall in crime zone 1, as shown in Figures 3 and 4, which is consistent with the above conclusion.

Taking point G as the common vertex, in this case, the crime zone 1 is divided into four triangles according to the shape, and the vertical center Z of each triangle is taken Z_1, Z_2, Z_3, Z_4 . Perpendicular to Z of the kth triangle Z_k is the vertex, connecting the two vertices of the bottom edge of the triangle except for the G point and extending to obtain four corner regions, using N_k represents the number of crime sites covered by each corner area.

The schematic diagram of the division of each corner of the crime area in this case is shown in Figure 5.

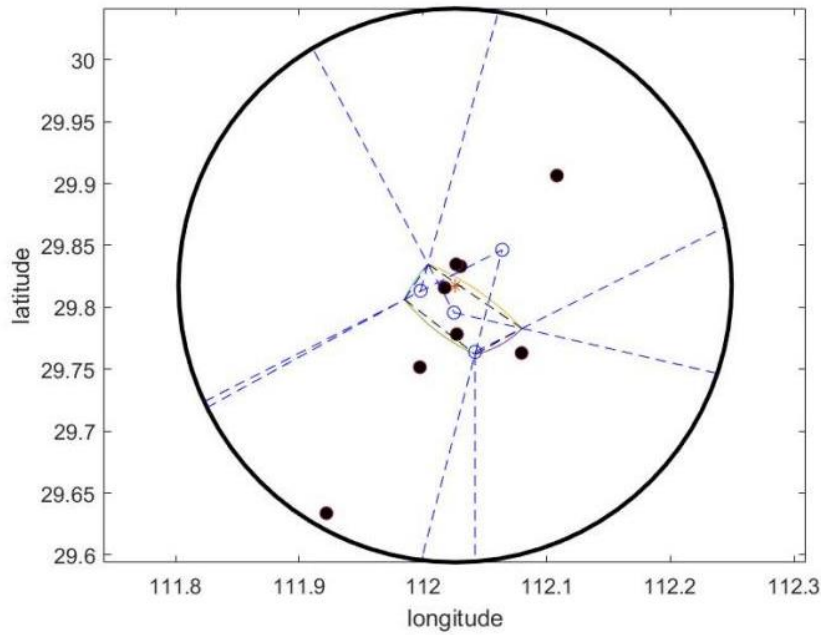


Figure 5. Schematic diagram of corner area

Calculate the average distance from the crime point covered by each corner area to point G.

$$R_k = \frac{1}{N_k} \sum_{1 \leq p \leq N_k} r_{kp} \quad (10)$$

Where r_{kp} is the distance from point G to arbitrary criminal place C_p covered by the k th triangle, R_k is the average distance from the crime point covered by each sector to point G, and $R_1 = 6.1273\text{km}$, $R_2 = 7.0568\text{km}$, $R_3 = 7.8555\text{km}$, the fourth triangle does not cover the crime area.

Then the crime radius coefficient of the k th sector is

$$\beta_k = \max \left\{ \sqrt{\frac{R_k}{R}}, \sqrt{\frac{R}{R_k}} \right\} \quad (11)$$

β_k is the crime radius coefficient of the k th sector, $R_{k_{\max}}$ is the maximum average distance from the sector area covering the crime point to the vertex Z_k , $R_{k_{\min}}$ is the minimum average distance from the sector area covering the crime point to the vertex Z_k .

Considering the development of criminal behavior and the comprehensiveness of prediction effect, if there is no crime place that can be covered in the P sector area, the maximum average distance of other areas will replace its maximum average distance, and the minimum average distance of other areas will replace its minimum average distance, that is

$$R_{p_{\max}} = \max(R_{k_{\max}}) \quad (12)$$

$$R_{p_{\min}} = \min(R_{k_{\min}}) \quad (13)$$

In Z_k is the center of the circle, while $R_{k_{\max}} = \beta_k \times R_k$ is the radius, with $R_{k_{\min}} = \frac{1}{\beta_k} \times R_k$ is the radius drawing sector, in which the red fan-shaped annular area between the two sectors is the crime hot area, which represents the higher possibility of crime. We call it crime zone 3, the white overlapping convex area in the center is crime zone 1, the other parts outside the yellow in the circular area are called crime zone 2, and the other parts are crime zone 0, as shown in Figure 6.

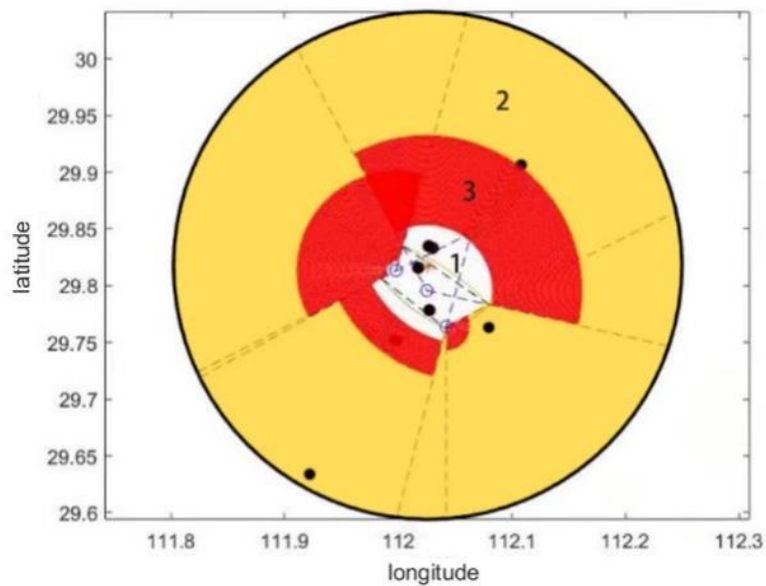


Figure 6. Prediction map of crime hotspots

According to the analysis of this model, we can get the basic distribution map of the crime area and the possible place of residence or stay of the offender.

Crime zone 3, that is, the high-risk area of crime, is the most likely place for the offender to commit the next crime; Then there are the second high-risk crime zone 2, the middle risk crime zone 1 and the low-risk crime zone 0.

4. Model checking

In this case data, we take the first seven of the nine crime cases as an example to test the model to verify whether the eighth crime place falls within the high-risk crime area predicted by the first seven crimes.

The crime prediction area generated based on the first seven data is shown in Figure 7, in which the green area is the actual next place of occurrence.

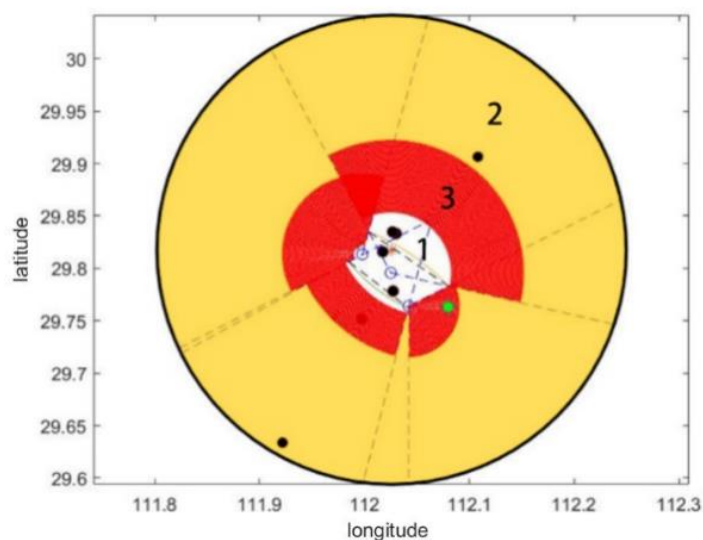


Figure 7. Generate the eighth crime prediction area based on the last seven data

Conduct a secondary test, and take the first eight data of the nine crimes as an example to verify whether the ninth crime falls within the high-risk crime area predicted by the first eight crimes.

The schematic diagram of the crime prediction area generated based on the first eight data is shown in Figure 8, The green area is the actual place of the next case.

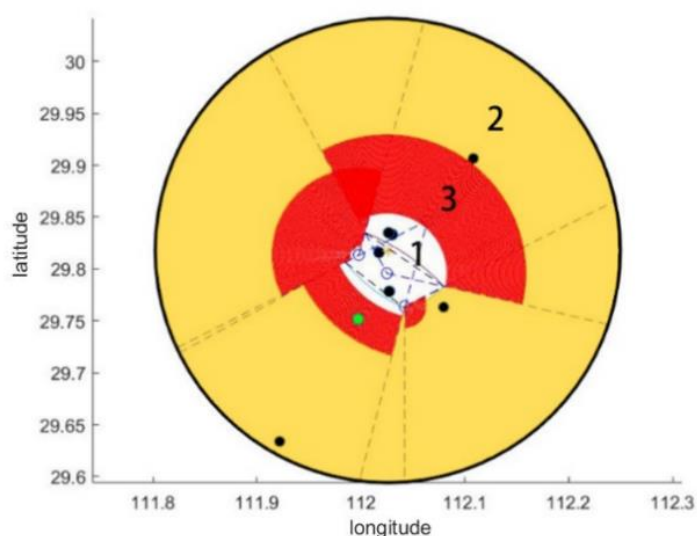


Figure 8. Generate the ninth crime prediction area based on the last eight data

In Figure 8, the red area is the crime area 3 predicted for the next crime location based on our crime prediction model, that is, the highest risk prediction area, and the green area is the actual next crime location. After testing, the eighth and ninth places of crime are highly consistent with the prediction of high-risk crime, which proves that the accuracy of the model is high, and can provide effective theoretical guidance for the actual crime prediction.

5. Conclusion

This paper comprehensively studies and judges the serious impact of a series of crimes on people's stability and social stability in today's era, and analyzes the difficulties encountered in the detection and capture of a series of crimes. This paper gets rid of the traditional police mode of "handling cases only by experience", excavates the psychological fluctuation factors, residence influencing factors, and the relevance of the place of crime in the continuous crime process of the perpetrator, starts from the rear, innovatively adopts the method of data modeling, and explores the practical prediction method in the process of cracking a series of criminal cases in the grass-roots police.

The model established in this paper combines the advantages of the two original crime prediction models, avoids its singleness and limitations, and solves the problems of low accuracy and unclear risk division in the crime prediction of serial crimes. On the basis of accurately predicting the location of criminals, it comprehensively studies and judges the location and crime location, as well as the differences, links and inherent laws between various crime locations. Rely on the combination of data modeling and geometric analysis to accurately draw the schematic diagram of the crime area, find the data correlation degree in many aspects, and establish a crime prediction model with high prediction accuracy and clear risk division through data fitting and geographical risk probability map weighting.

After testing, the model is simple and easy to use, has strong field operability, the required data is neat and easy to collect, and the prediction accuracy is high. This model can provide strong theoretical support for the prediction and early warning of serial crime cases, police force deployment and arrest action in practical work, and to a certain extent, it solves the current situation of the separation of scientific and technological innovation technology from grass-roots actual combat in today's grass-roots police work.

Acknowledgements

The authors gratefully acknowledge Ministry of Public Security Theory and Soft Science Project "Research on Application Evaluation System of Public Security Informatization and Intelligent Construction of Big Data" (Project No. : 2022LL27) funds.

References

- [1] LIU Moubin. A Preliminary Exploration of the Regularity of Series Crimes [J]. Journal of Political Science and Law, 2009,26 (6):29-33.
- [2] LÜ Yunping, NI Xiaofeng et al. The Criminal's Behavior Characteristics in a Series of Crimes and the Investigating Countermeasures for the Series Crimes [J]. Journal of Zhe Jiang Police College. 2001 (5):24-27.
- [3] WANG Lanxiang, CHEN Youfei, LI Minqiang, LIN Guangfa. Spatial Analysis for Crime Hot Spots [J]. Journal of Fujian Police College. 2012,26 (1):16-20.
- [4] MIAO Lijun, LIU Qiang, WANG Yizhou, CHEN Changchun. Space-time prediction of serial crimes [J]. Journal of Nanjing University of Information Science and Technology: Natural Science Edition. 2010,2 (6): 514-518.
- [5] HOU Zhenting, LIU Zaiming, ZOU Jiezhong, LI Xuwei. Markov Skeleton Processes [J]. Chinese Science Bulletin. 1998,43 (5):457-467.
- [6] [EB/OL].http://www.0719house.com/news/news_72513_68108_1.html,2012-04-12.
- [7] [EB/OL].<http://www.deyi.com/forum-viewthread-tid-4459347-highlight.html>, 2013-5-15.
- [8] Joshua Kent; Michael Leitner. Efficacy of standard deviational ellipses in the application of criminal geographic profiling[J]. Journal of Investigative Psychology and Offender Profiling, 2007, 4 (3): 147-165.
- [9] LI Xinguang, ZHANG Jiayan, FENG Xugang, ZHOU Kewu. An Optimization of Geographic Profiling Model Based on Fuzzy Algorithm [J]. Journal of Changzhou Institute of Technology. 2017,30 (4-5):28-32.
- [10] XUE Zhong, QIAO Liang, WANG Feng, GAO Qi. The Application of Area Overlap Weighted Model (AOWM) in Series-Crime Prediction [J]. Mathematics in Practice and Theory. 2010,40(15):212-217.