

# 第四届“卿云杯”全国通识课程论文大赛

## 封面样张

学校	同济大学	院系	材料科学与工程学院 机械与能源工程学院 材料科学与工程学院
专业	高分子和复合材料 工业工程 高分子和复合材料	姓名	王佳音、胡修齐、杨玉冰
年级	大三	任课教师	夏焯
课程名称	数据量化推理		
论文题目	基于 CGT 模型的犯罪心理缓冲区计算方法研究		

基于 CGT 模型的犯罪心理缓冲区计算方法研究

王佳音<sup>1</sup>，胡修齐<sup>2</sup>，杨玉冰<sup>3</sup>

(1. 同济大学 材料科学与工程学院, 上海; 2. 同济大学 机械与能源工程学院, 上海; 3. 同济大学 材料科学与工程学院, 上海)

## 摘要:

在连环杀人案发生时, 犯罪者的作案地点通常呈现一定的规律, 反映了犯罪者的持续选择趋势。经过文献研究发现, 犯罪者对作案地点的选择反映了他们的“心理缓冲区”即在距离上犯罪风险与收益达成平衡的位置。通过比较不同心理缓冲区计算方法得到的犯罪者居住地点与实际地点的偏差, 可以分析出不同心理缓冲区计算方法的适用情况。本文还评估了模型在数据缺失情况下的稳健性, 不同的 B 值计算方法对模型稳定性的敏感度反映出它们对数据丢失的不同抵抗力。

**关键词:** 犯罪地理学、CGT 模型、心理缓冲区、模型稳健性、算法比较

## 1 引言

犯罪行为与地理环境之间的紧密关系表明犯罪不仅是一种社会现象, 更是一种地理行为。Brantingham 夫妇提出的犯罪模式理论深入阐释了这种关系, 指出犯罪活动最有可能发生在机会空间和认知空间相交的地方<sup>[1]</sup>。其中机会空间是指犯罪人感知的包含感兴趣目标的地点, 而认知空间则是犯罪人拥有具体环境信息的地点, 例如他们的住所附近。在这种理论框架下, 系列犯罪因其作案次数较多, 更容易通过案件发生地点的空间分布进行犯罪地理画像研究。

犯罪地理目标模型 (Crime Geographical Targeting, CGT) 由 Rossmo 博士提出, 是犯罪地理画像的方法之一, 具有较好的实战应用效果。CGT 模型通过分析连环犯罪案件的空间关联性, 确定犯罪者高概率的可能居住地区, 为犯罪防控提供有力的支持。

但是由于国内犯罪数据的敏感性与保密性, 我国犯罪地理画像研究仅仅限于理论介绍与探讨, 缺乏实例研究<sup>[2]</sup>, 更缺乏对于模型中涉及的具体参数的讨论<sup>[3]</sup>。CGT 模型中最为重要的参数即为 B 值, 而 B 值代表的是“心理缓冲区”, 其合理性和准确性直接关系到模拟结果的可信度。在未知犯罪人居住地点的情况下, 需要通过用案件点群内的综合距离算法拟合可能的缓冲区 B 值。目前, 研究者们普遍采用多种 B 值计算方法并得到大量计算结果。

本文致力于探究不同的 B 值计算方法对于 CGT 模型模拟结果精确度的影响。通过对多种 B 值计算方法的比较分析, 我们旨在揭示各方法在不同场景下的适用性, 为 CGT 模型的进一步发展和优化提供实际指导。

<sup>[1]</sup> 俸卫、高竟洪: 《预测连环犯罪区域的地理画像法研究》, 《内江师范学院学报》, 2017 年第 32 期, 第 38-42 页

<sup>[2]</sup> Kan Dai, “Criminal Geographic Target Model Optimal Integrated Distance between Spatial Points of Serial Burglaries,” *Geographical Science Research*, 2015, 4(1), pp. 16-26.

<sup>[3]</sup> 唐乙鑫, 饶亚君: 《国内犯罪地理学研究进展》, 《法制博览》, 2021 年第 33 期, 第 1-5 页。

## 2 犯罪地理目标模型

### 2.1 犯罪活动距离曲线

选取的两个案件的发生地点在城市，所以在计算犯罪地点之间的距离时，本文选择采用曼哈顿距离进行计算和分析。曼哈顿距离是两点之间的直角折线距离，特别适用于那些路网呈规则格网分布的城市。在这种距离度量下，路径沿着城市的道路网格形式移动，形成水平和垂直的直线组成的折线，相比于直线距离，能够更真实地反映了城市实际的道路布局。此外，本文假设犯罪者在犯罪活动中选择最短路径出行。

### 2.2 犯罪地理目标模型算法<sup>[4]</sup>

#### 2.2.1 确定计算范围

在模型的计算过程中，需要首先确定参与计算的地图范围，参与计算的地图面积大小也影响了格网大小的划分。CGT 模型将参与计算的地图范围确定为略大于包含所有犯罪地点的矩形区域。

$$y_{high} = y_{max} + (y_{max} - y_{min}) / 2(C - 1)$$

$$y_{low} = y_{min} - (y_{max} - y_{min}) / 2(C - 1)$$

$$x_{high} = x_{max} + (x_{max} - x_{min}) / 2(C - 1)$$

$$x_{low} = x_{min} - (x_{max} - x_{min}) / 2(C - 1)$$

式中， $y_{high}$  和  $y_{low}$  是矩形区域的最大和最小 y 坐标， $x_{high}$  和  $x_{low}$  是矩形区域的最大和最小 x 坐标； $y_{max}$  和  $y_{min}$  是所有案件点中的最大和最小 y 坐标， $x_{max}$  和  $x_{min}$  是所有案件点中的最大和最小 x 坐标；C 是案件点总数。

#### 2.2.2 计算网格点概率

依据案件点与犯罪人居住点之间的联系以及环境犯罪学的相关理论，将地图划分为大小合适的格网，通过 CGT 模型算法计算每个网格的犯罪人居住概率值，得到完整的概率表面，并基于概率表面预测犯罪人可能的居住地点。具体概率计算方法是：

对于地图上的任一格格点(i,j)，该点犯罪人居住的概率为：

$$P_{ij} = k \sum_n^c \left[ \frac{\Phi}{(|x_i - x_n| + |y_i - y_n|)^f} + \frac{(1 - \Phi)(B^{g-f})}{(2B - |x_i - x_n| - |y_i - y_n|)^g} \right]$$

式中， $P_{ij}$  是犯罪人居住在格网点(i,j)的概率值，是每一个犯罪地点计算得到的概率的累加概率。概率计算公式实际上是一个分段函数，用权重系数 $\Phi$ 来控制：

$$\phi = \begin{cases} 1, & |x_i - x_n| + |y_i - y_n| > B \\ 0, & |x_i - x_n| + |y_i - y_n| \leq B \end{cases}$$

式中，k 是一个常数，用来放大总体的概率数值；B 是缓冲区的半径，代表了

<sup>[4]</sup>李卫红、戴侃、闻磊：《顾及地理因素的犯罪地理目标模型改进方法》，《测绘科学》，2015年。

犯罪人“心理缓冲区”的距离，C 是犯罪地点的个数；f 和 g 均为经验值常数； $x_i$ ， $y_j$  是第 i 行，第 j 列对应格网点的坐标值； $x_n$ ， $y_n$  是第 n 个犯罪地点的坐标值。

### 2.2.3 形成概率表面

根据计算得到的格网点嫌疑人落脚概率，生成一幅连续的概率表面，其中高概率的区域被认为是嫌疑人最有可能居住的地方。

## 3 心理缓冲区的概念与计算方法

“心理缓冲区”指的是在距离上达成的犯罪风险与收益的平衡位置，环境犯罪学认为犯罪是犯罪人和他们所处环境相互作用的产物，系列案件的犯罪人具有一定的认知空间，这种认知空间随着作案次数的增多越来越成熟，犯罪人会从自己的认知空间中选择合适的目标<sup>[5]</sup>。

越过这个区域之后，随着距离的增加，犯罪人的出行成本以及对作案地区的陌生程度也逐渐增加，作案的概率也逐渐降低<sup>[6]</sup>。通过对出行距离与案件个数的关系研究，可以得到存在一个峰值的距离衰减曲线，峰值对应的距离范围即可认为是犯罪人“心理缓冲区”的距离范围，也就是 CGT 模型中缓冲区 B 的值。

以下是几种根据犯罪地点的空间点群特征计算 B 值的计算方法<sup>[7]</sup>。

1. SNND——平均最邻近距离的一半。
2. NND——平均最邻近距离，指所有点与其最邻近的点之间距离的平均值。
3. SFND——平均最远距离一半。
4. FND——平均最远距离，指所有点与其最远点之间距离的平均值。
5. IPD——点间平均距离，是指所有点间距离的平均值。
6. SD——标准距离， $SD = \sqrt{(\sum r_{cs}^2) / T}$ ， $r_{cs}$  表示第 c 个点与质心的距离，T 是点总数。
7. CD——质心距离，每个点与质心点之间距离的平均值。

## 4 案件信息收集

### 4.1 白银连环杀人案

白银杀人案的案发地点总共有九处，因为这个案件是一个连环入室杀人案，所以所有的案发地点都是受害者的家中，所以位置坐标比较准确。

这九处地点<sup>[8]</sup>分别为：

<sup>[5]</sup> 陈闻高：《犯罪心理画像论》，《犯罪研究》，2013 年第 6 期，第 2-15 页。

<sup>[6]</sup> 郑涵予、李明楚、韦东、陈俊、李晓临：《系列抢劫案中地理学犯罪心理画像的研究》，《法制博览》，2019 年第 21 期，第 26-30 页。

<sup>[7]</sup> Taylor, P.J., “Quantitative methods in geography: An introduction to spatial analysis,” Houghton Mifflin, Boston, 1977, pp.146-147.

<sup>[8]</sup> 央视新闻，《甘肃白银连环杀人案作案地点：图解高承勇作案手法》，<http://www.mnw.cn/news/shehui/1343706.html>

地点	坐标
甘肃省白银市白银区永丰街 177-1 号	104.17812,36.55590
甘肃省白银供电局	104.18376,36.55340
甘肃省白银区胜利街 88-6 号	104.18714,36.54744
甘肃省白银区水川路 6 号	104.17982,36.54371
甘肃省白银供电局计量所 4 楼 414 号	104.16807,36.55214
甘肃省白银区东山路 59-6-4 号	104.19150,36.54886
甘肃省白银棉纺厂平房家属区	104.17293,36.55640
甘肃省白银区水川路 28-1-12 号	104.18022,36.53831
犯罪者住址 104.17170, 36.55397	

## 4.2 马家坪案件

马家坪案的案发地点总共有十一处，因为这个案件是一个随机连环杀人案，犯罪者在大街上寻觅目标，所以所有的案发地点都是受害者去往某个目的地的路上，造成坐标的准确性无法保障。

这十一处地点<sup>[9]</sup>分别为：

事件	坐标
家在马家坪西居民区的钟蕊（化名）下班途中遇害	113.55730 ， 37.87769
女工张然(化名)在途经马家坪咸菜厂附近回家时， 被人卡住脖子用单刃刀连捅四刀	113.56056 ， 37.87734
女工时春(化名)途经马家坪西区石坡处时遇害	113.55799 ， 37.87848
女工李昕(化名)在矿区平坦新区居民楼道内，被尾 随而来的歹徒用单刃刀连捅八刀，当场死亡	113.56541 ， 37.87354
女工郝花(化名)下班回家行至马家坪西区运输二场 北侧土坡处被歹徒用匕首连捅七刀	113.55702 ， 37.87770
辽宁籍女青年张秀(化名)打着雨伞走进河神庙一拆 迁锅炉房废墟内时，被凶犯朝胸腹部连捅数刀	113.55805 ， 37.87682
女青年王花(化名)从河神庙华梦歌厅外出，当行至 河神庙口“之”字形台阶附近的厕所处时，被凶犯劫 持遇害	113.56094 ， 37.87618
女工毕平(化名)回到市委党校居民楼内时遇害	113.56967 ， 37.87536
女工张雪(化名)行至王岩沟口石板路段时遇害	113.55559 ， 37.87910
女工郭沁(化名)回家途经王岩沟居民区时失踪	113.55550 ， 37.87956

<sup>[9]</sup> 《深夜魔影——阳泉马家坪系列杀人案》，  
<https://baijiahao.baidu.com/s?id=1689692081691100035&wfr=spider&for=pc>

青年女子唐蕊(化名)与在矿区俱乐部遇害	113.55559 , 37.87910
犯罪者住址 113.55564 , 37.87722	

## 5 数据计算以及分析

使用 CGT 模型进行计算，并通过 python 实现。

### 5.1 白银市连环杀人案

表 1 展示了对白银市连环杀人案应用 CGT 模型后的计算数据，反映了不同计算方法对于预测地点的影响，以及它们对模拟结果的精确度的差异。较小的 RMSE 值通常表示计算方法更接近真实情况。

表 1 白银市连环杀人案计算数据

计算方式	预测的最可能地点经纬度	RMSE
SNND	104.18863, 36.54939	0.01240
NND	104.18146, 36.54865	0.00786
SFND	104.17524, 36.54680	0.00565
FND	104.16807, 36.54237	0.00859
IPD	104.17524, 36.54422	0.00734
SD	104.18002, 36.55197	0.00605
CD	104.17524, 36.54828	0.00474

图 1 将不同计算方法的 RMSE 值直观的体现在一张图中，能够清晰地比较不同计算方法的准确度，可以看到 CD 算法的偏差较小，精确度高。

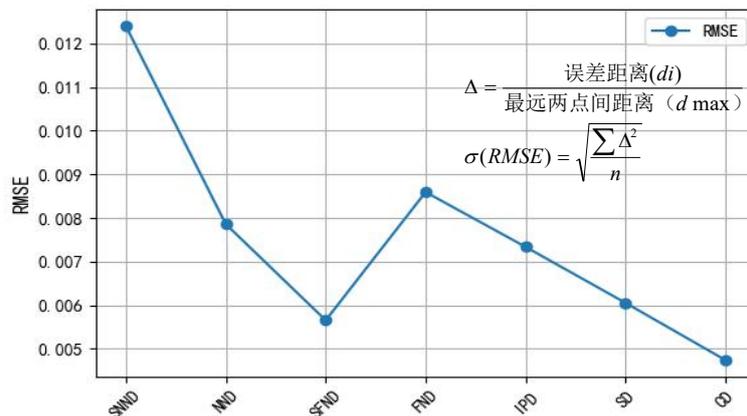


图 1 白银连环杀人案的不同计算方法的偏差

下文图 2-8 是不同 B 值计算方法对应的白银连环杀人案犯罪者居住地点概率表面分布图，按照顺序分别是 SNND、NND、SFND、FND、IPD、SD、CD。左边的图为二维概率表面图，图中红色的点代表案发地点，绿色的点代表模拟的犯罪者居住地点，蓝色的点代表的是犯罪者真实住址。右边的图是左图对应的三位概率表面图，从图中可以直观看出整个计算范围内犯罪者居住地点概率的分布，

除比较 RMSE 以外，搜索比率也是评价模拟结果的重要指标，观察三位概率表面的高度可以定性判断搜索比率<sup>[10]</sup>。在整个地图范围即略大于包含所有犯罪地点的矩形区域内，颜色越深代表这个地方是犯罪者居住地点的可能性越大，具体的概率通过右边的图例指示。

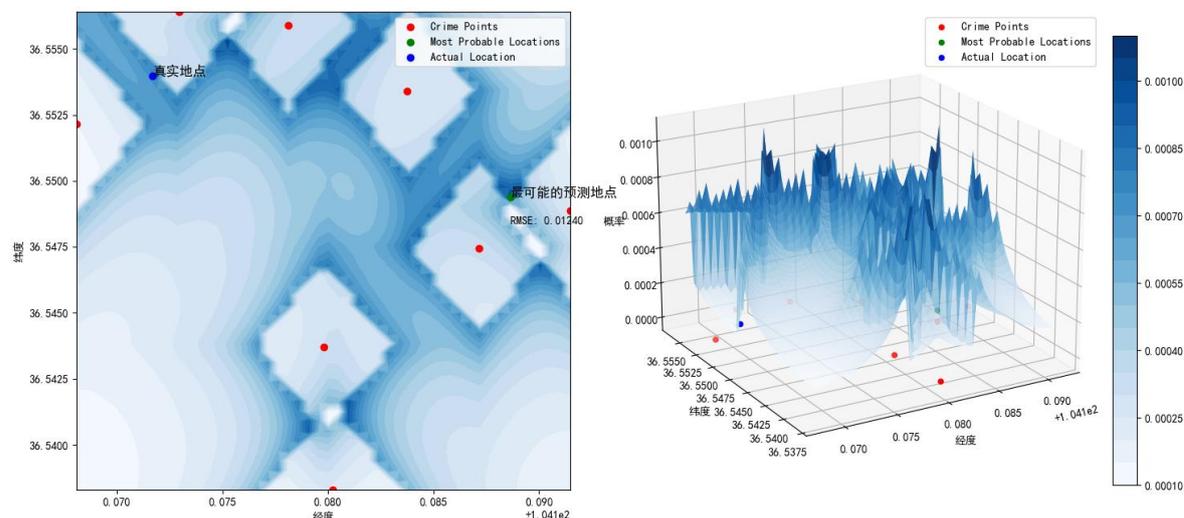


图2 SNND 计算出的 B 值的预测

图2可以看到SNND算法的概率表面上颜色较深的区域都分布在和红点有固定距离的地方，是一个以红点为中心的菱形的边缘，而绿点出现在两个菱形的交点处。

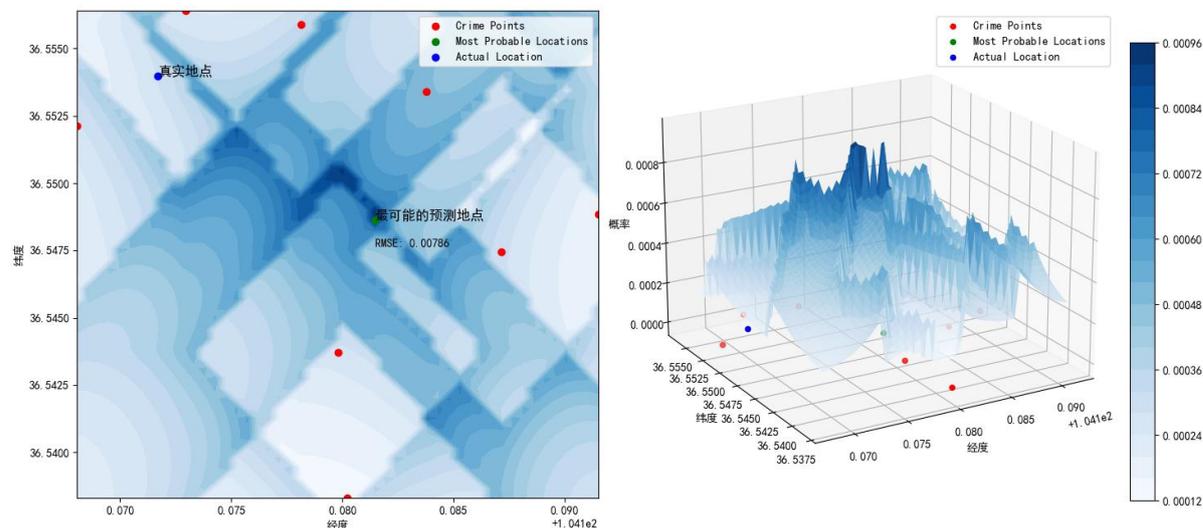


图3 NND 计算出的 B 值的预测

图3表示的NND算法概率表面的菱形边长为SNND算法的两倍，符合二者之间两倍的关系，绿点同样出现在菱形的交点处。

同样的规律也适用于图4的SFND概率表面和图5的FND概率表面，因为

[10] 俸卫、高竞淇：《预测连环犯罪区域的地理画像法研究》，《内江师范学院学报》，2017年第32期，第38-42页

这两种算法表示的是最远平均距离，所以它们的菱形要比计算最近平均距离的 SNND 算法和 NND 算法的菱形大，且 FND 算法的概率表面是 SFND 算法的概率表面的两倍。

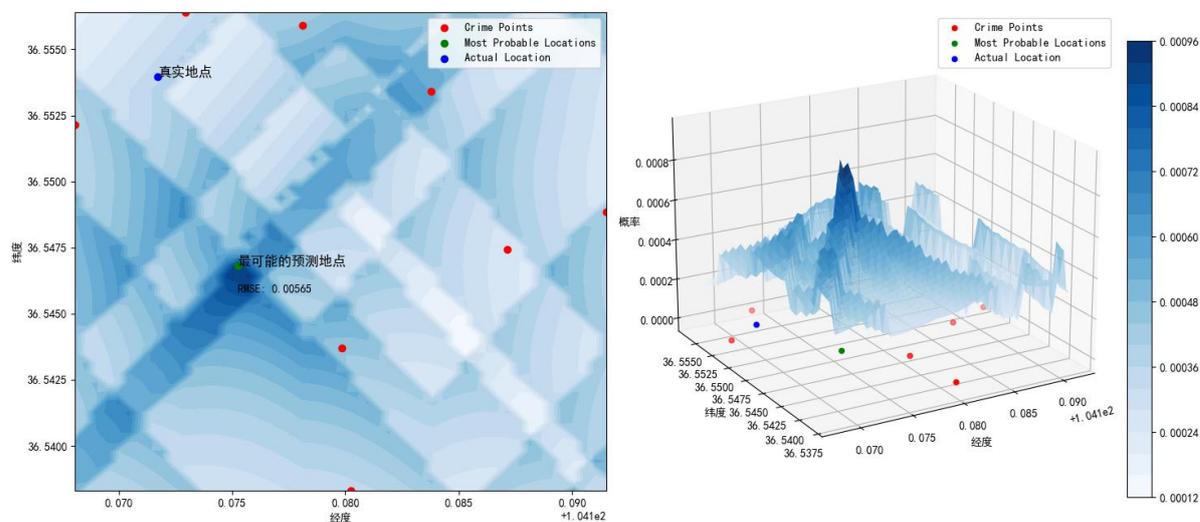


图 4 SFND 计算出的 B 值的预测

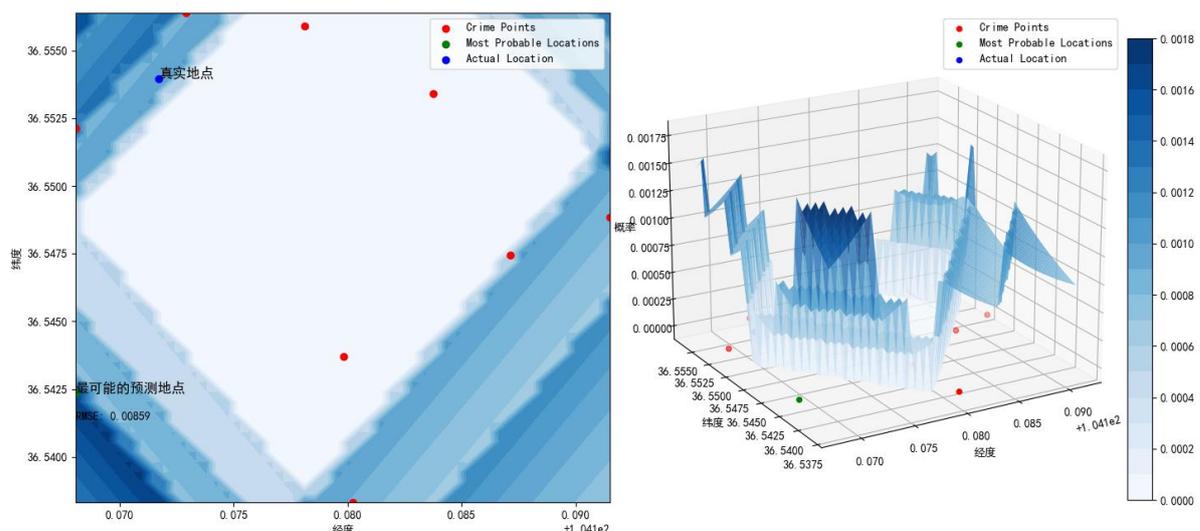


图 5 FND 计算出的 B 值的预测

图 6 的 IPD 表示的是点间平均距离，图 8 的 CD 表示质心距离即每个点与质心点之间距离的平均值，这两种算法和前四种算法一致的是它们的 B 值是确定的，而图 7 表示的 SD 算法的 B 值不是一个固定的值，所以它的概率表面也与其他不同，是一个围绕中心几何形状形成的嵌套形状，最可能的预测地点在中心几何形状中。

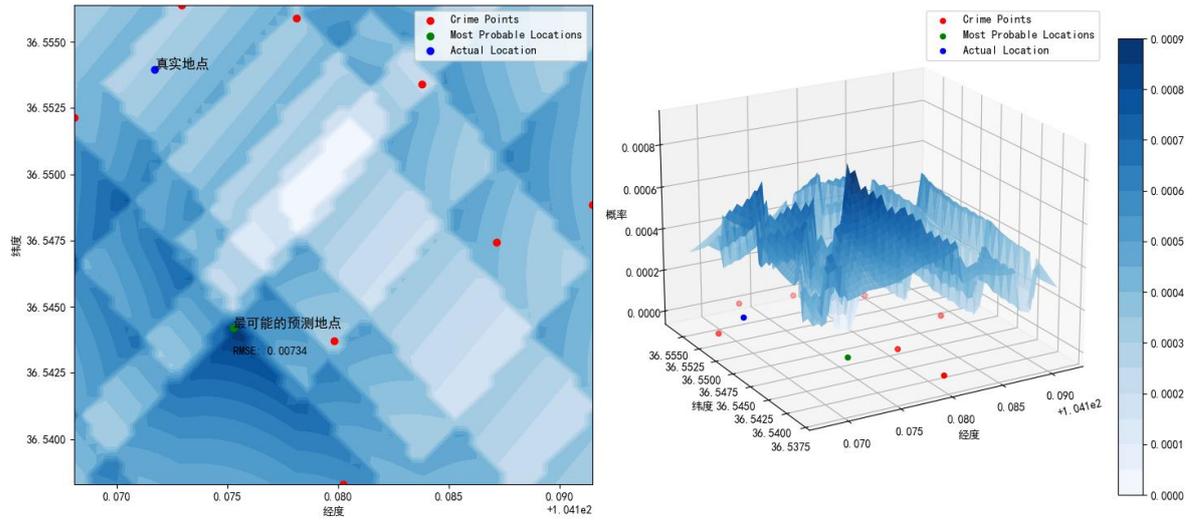


图 6 IPD 计算出的 B 值的预测

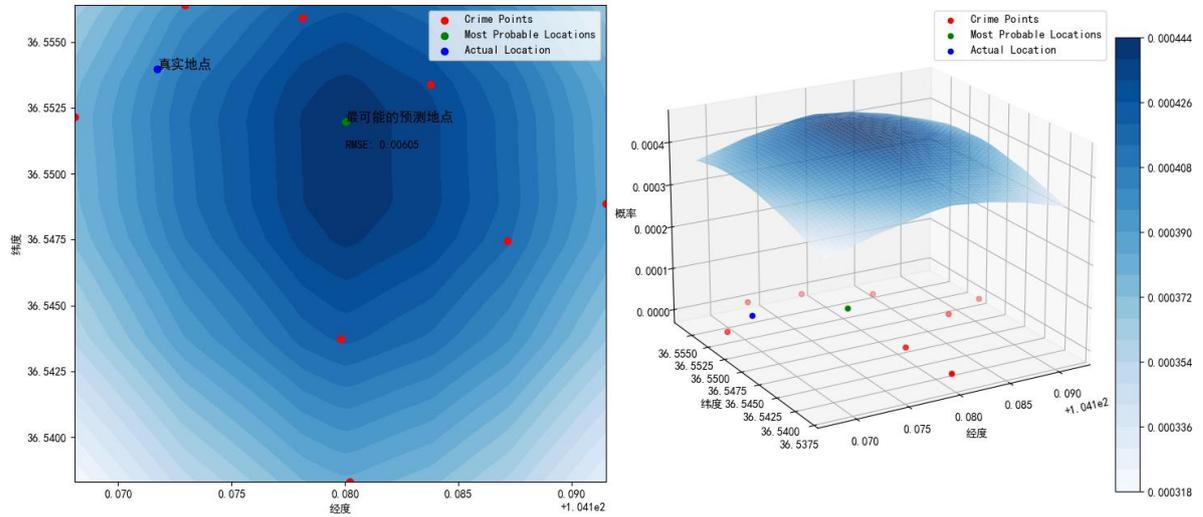


图 7 SD 计算出的 B 值的预测

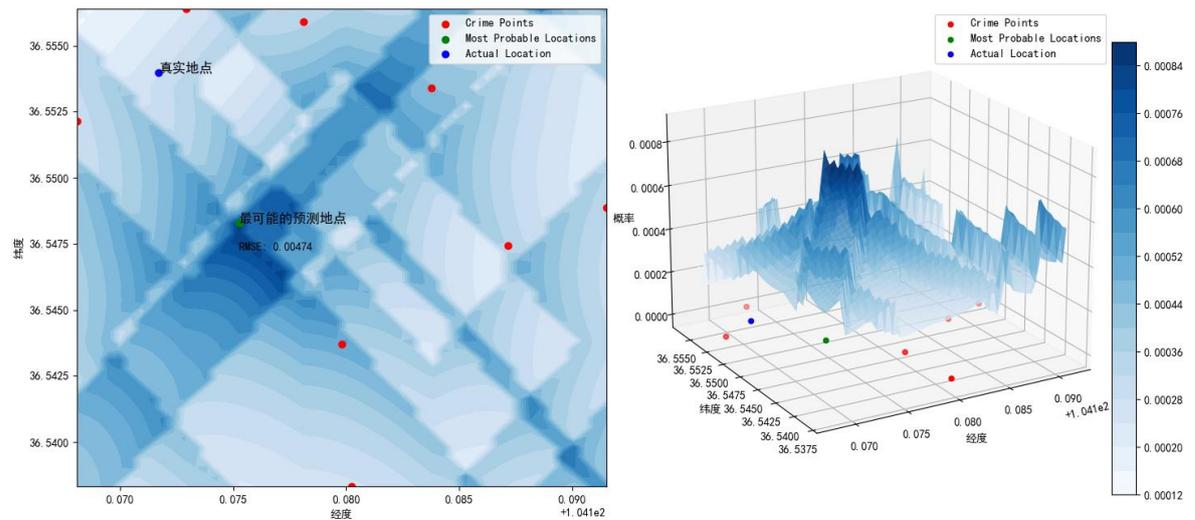


图 8 CD 计算出的 B 值的预测

## 5.2 马家坪案件

表 2 展示了对马家坪连环杀人案应用 CGT 模型后的计算数据，反映了不同

计算方法对于预测地点的影响，以及它们对模拟结果的精确度的差异。较小的 RMSE 值通常表示计算方法更接近真实情况。

表 2 马家坪案件的计算数据

计算方式	预测的最可能地点经纬度	RMSE
SNND	113.55608, 37.87870	0.00109
NND	113.55550, 37.87735	0.00013
SFND	113.56331, 37.87931	0.00562
FND	113.56880, 37.87403	0.00958
IPD	113.56475, 37.87956	0.00665
SD	113.55781, 37.87772	0.00158
CD	113.55550, 37.87428	0.00208

图 9 将马家坪案件不同计算方法的 RMSE 值直观的体现在一张图中，能够清晰地比较不同计算方法的准确度，可以看到 NND 算法的偏差较小，精确度高。

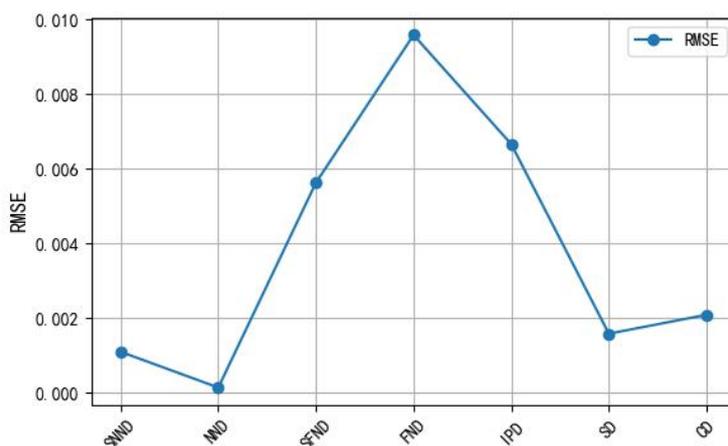


图 9 马家坪连环杀人案的不同计算方法的偏差

下文图 10-16 含义与上一节白银案相似。

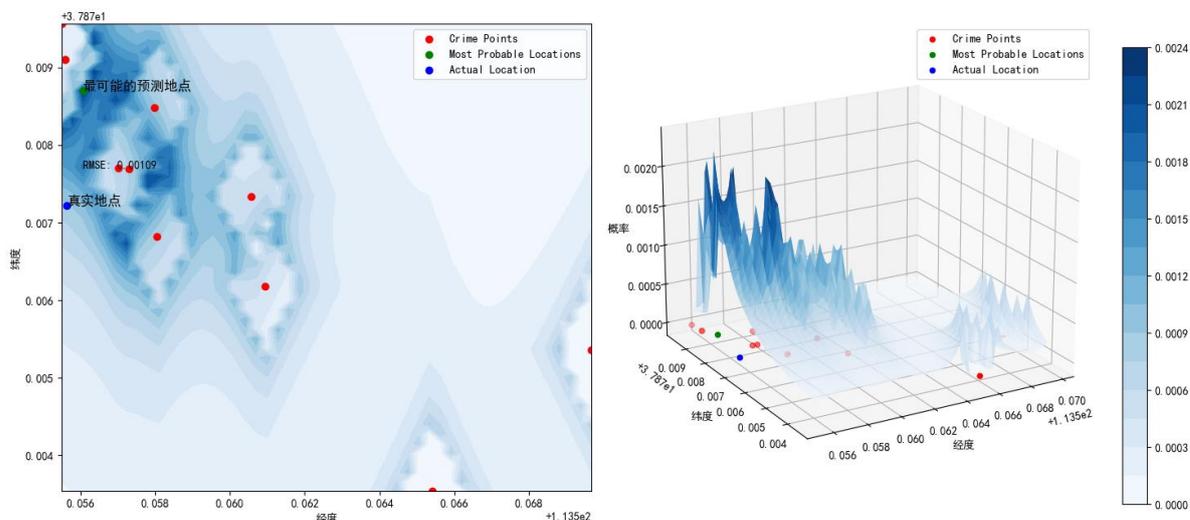


图 10 SNND 计算出的 B 值的预测

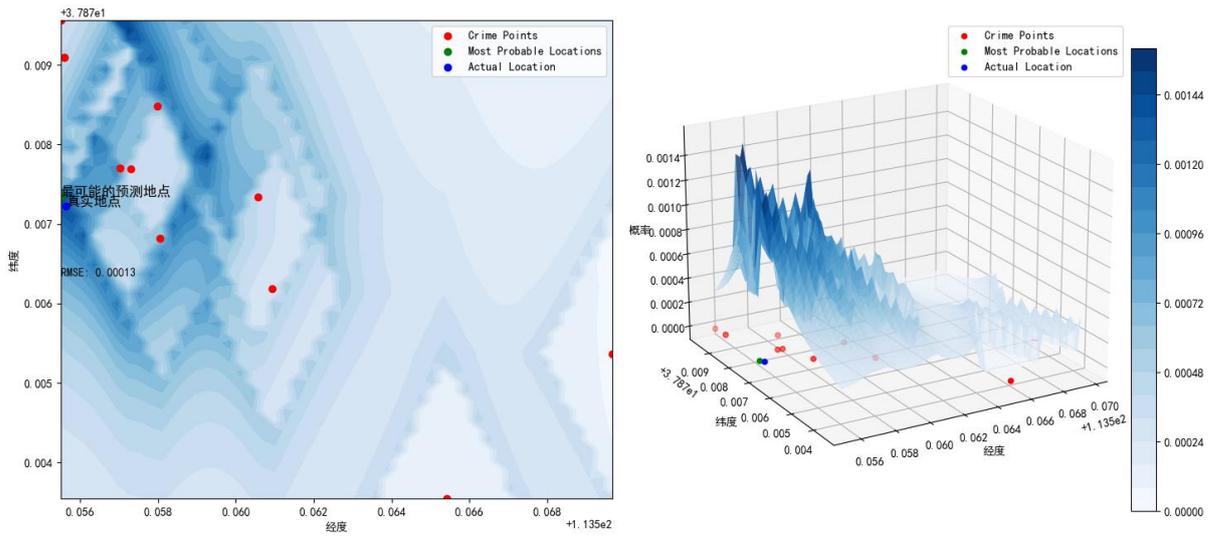


图 11 NND 计算出的 B 值的预测

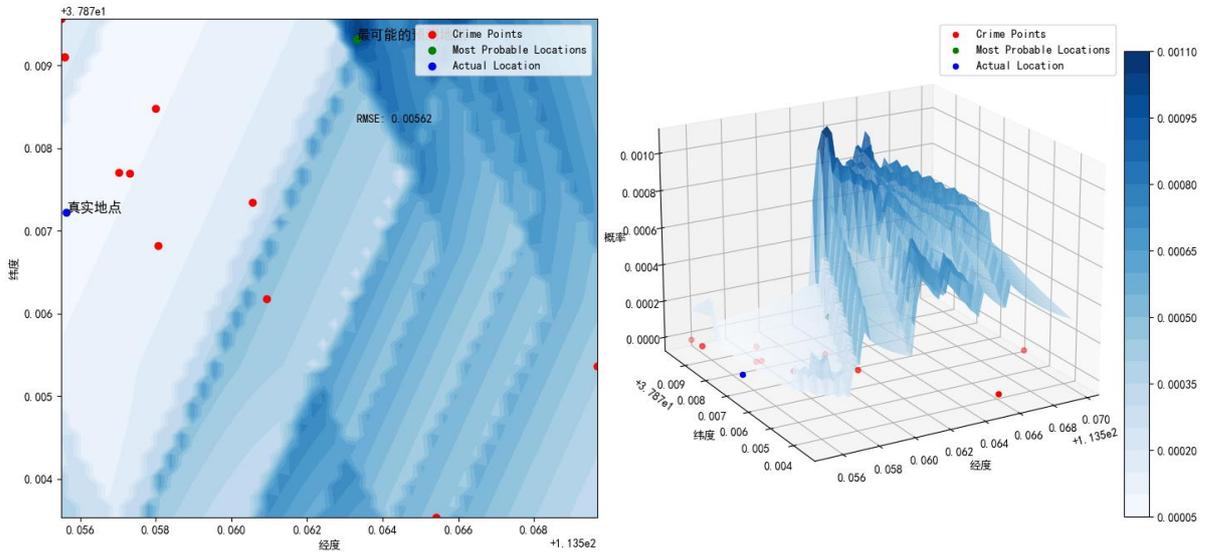


图 12 SFND 计算出的 B 值的预测

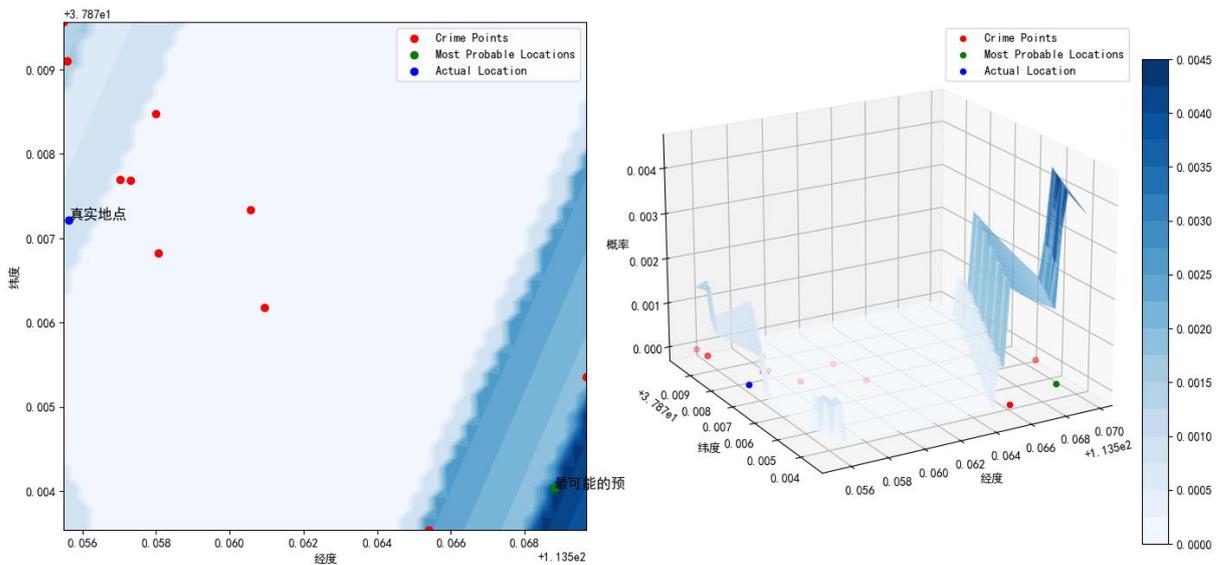


图 13 FND 计算出的 B 值的预测

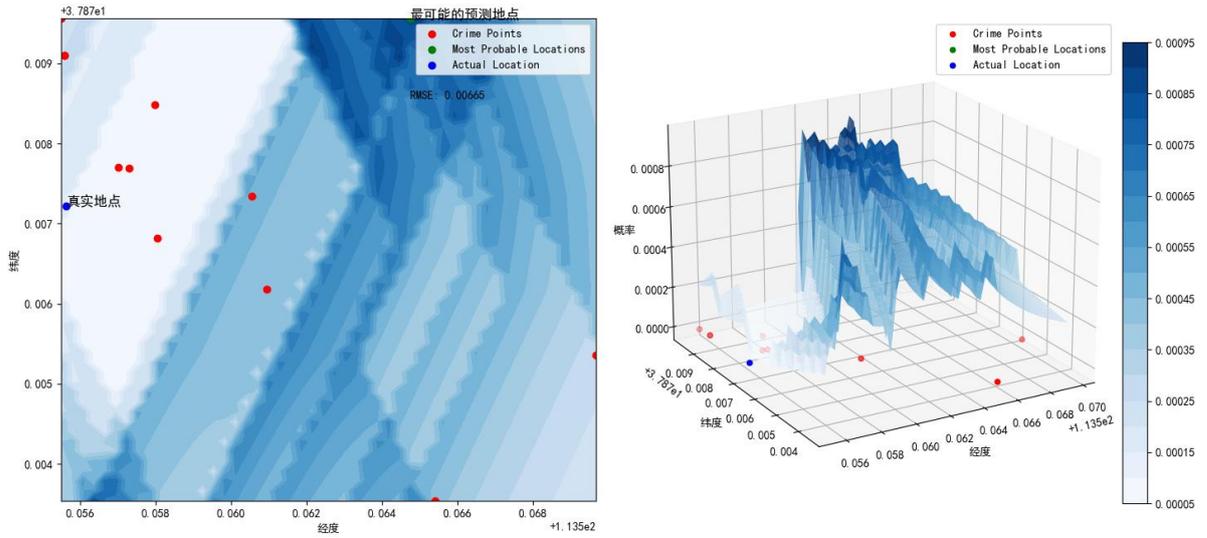


图 14 IPD 计算出的 B 值的预测

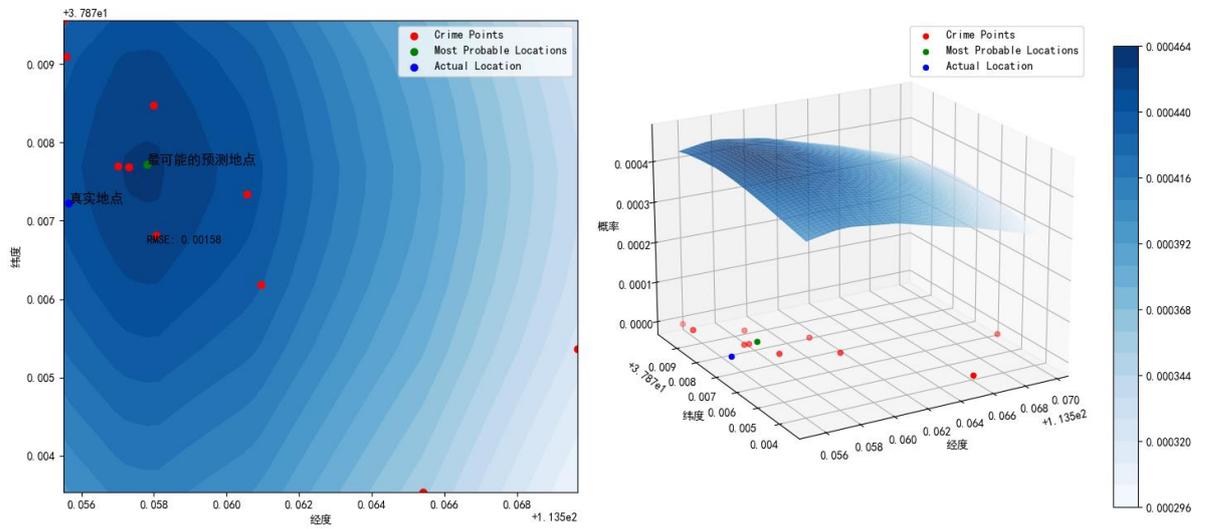


图 15 SD 计算出的 B 值的预测

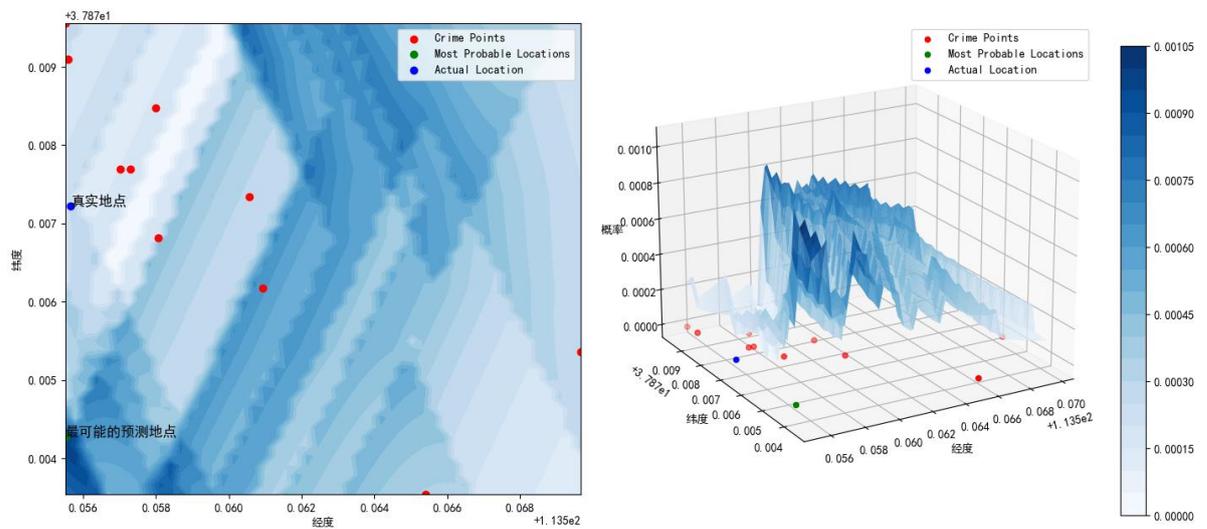


图 16 CD 计算出的 B 值的预测

### 5.3 数据分析

#### 5.3.1 不同 B 值计算方法对 CGT 模型的稳健性分析

敏感性测试 (Sensitivity Test) 可以用来评估模型的稳定性。为了全面考察模型的鲁棒性, 我们采用了多种距离计算类型, 包括 SNND、NND、SFND、FND、IPD、SD 和 CD 等。每一种距离计算类型代表了对模型稳定性的一种特定考察方式。在每次迭代中, 我们逐步删除数据点, 并计算相应的均方根误差 (RMSE)。这一方法为我们提供了深入了解模型鲁棒性和稳定性的机会, 从而更全面地评价其在实际应用中的可靠性。

表 3 白银市连环杀人案不同算法下删除数据点的模型响应数据

删除数量	SNND	NND	SFND	FND	IPD	SD	CD
0	0.0124	0.00786	0.00565	0.00859	0.00734	0.00605	0.00474
1	0.01234	0.00576	0.00581	0.00295	0.0077	0.00699	0.00564
2	0.01394	0.00774	0.00613	0.00223	0.00606	0.00744	0.00693
3	0.00186	0.00432	0.00792	0.01217	0.01029	0.00699	0.0076
4	0.00328	0.0072	0.00652	0.01263	0.00947	0.00163	0.00733
5	0.0069	0.01186	0.00704	0.00086	0.0074	0.01695	0.00733

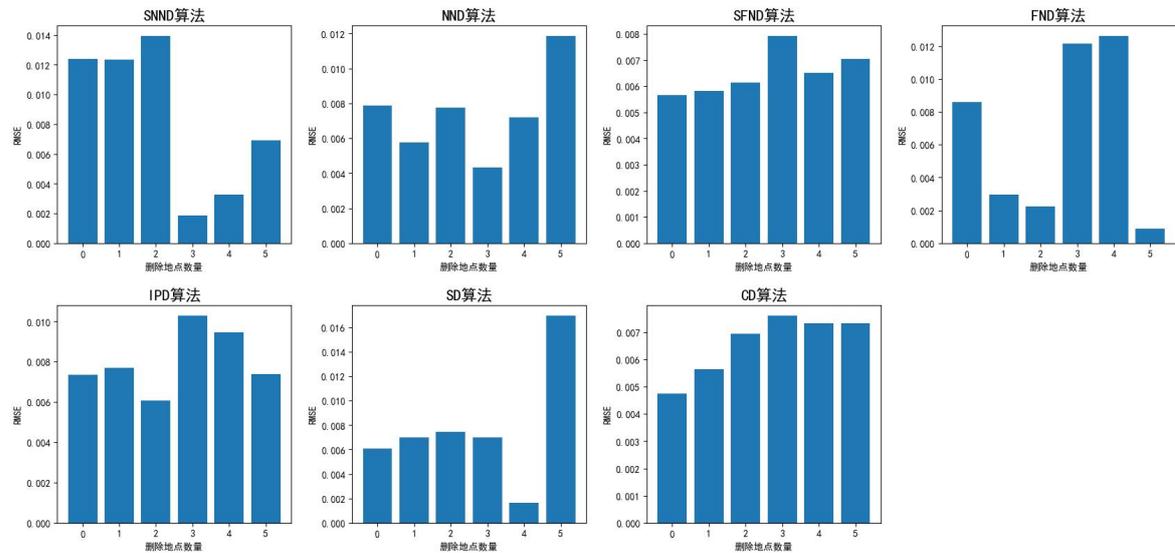


图 17 白银市连环杀人案不同算法下删除数据点的模型响应图

表 3 和图 17 是白银市连环杀人案不同算法下删除数据点的模型响应。从中可以看到, SNND 算法当删除的数据点增加时, RMSE 值波动较大, 这意味着 SNND 算法对数据丢失的敏感性较高; NND 算法在数据删除时显示出一定的波动, 尤其是在删除 5 个数据点时, RMSE 值显著增加, 表明该算法对数据完整性较为敏感; SFND 算法的 RMSE 值在删除 0 至 3 个数据点时相对稳定, 但在删除更多数据点时有所上升, 这表明算法对较大程度的数据丢失较为敏感; FND 算

法的 RMSE 在初始阶段下降，这可能是因为删除的异常值导致预测更准确；然而，当删除更多数据点时，RMSE 值急剧下降，表明算法可能对数据丢失不敏感，或者在数据量较少时表现更好；IPD 算法在删除数据点时表现出一定的波动性，但整体趋势没有显著上升或下降，这表明 IPD 算法在数据不完整时相对稳定；SD 算法在删除 4 个数据点之前相对稳定，但在删除 5 个数据点时，RMSE 值显著上升，这表明 SD 算法对大量数据丢失非常敏感；CD 算法的 RMSE 值在不同的数据删除量中变化不大，表明该算法对数据的丢失具有一定的鲁棒性。

表 4 白银市连环杀人案不同算法下删除数据点的模型响应数据

删除数量	SNND	NND	SFND	FND	IPD	SD	CD
1	0.00109	0.00013	0.00562	0.00958	0.00665	0.00158	0.00208
2	0.00124	0.00199	0.0056	0.00942	0.00643	0.00178	0.00226
3	0.00142	0.00036	0.00564	0.0101	0.00506	0.00158	0.00252
4	0.00084	0.00044	0.00564	0.00994	0.00556	0.00118	0.00445
5	0.00084	0.00044	0.00564	0.00978	0.00464	0.00099	0.00217
6	0.00167	0.0025	0.0058	0.00797	0.00501	0.00131	0.0026
7	0.00186	0.00089	0.00582	0.00866	0.00698	0.00131	0.00253
8	0.00207	0.00347	0.00641	0.0094	0.00273	0.00131	0.00484
9	0.00131	0.00131	0.0014	0.00143	0.00148	0.00126	0.00148
10	0.00166	0.00168	0.00166	0.00168	0.00166	0.00168	0.00166

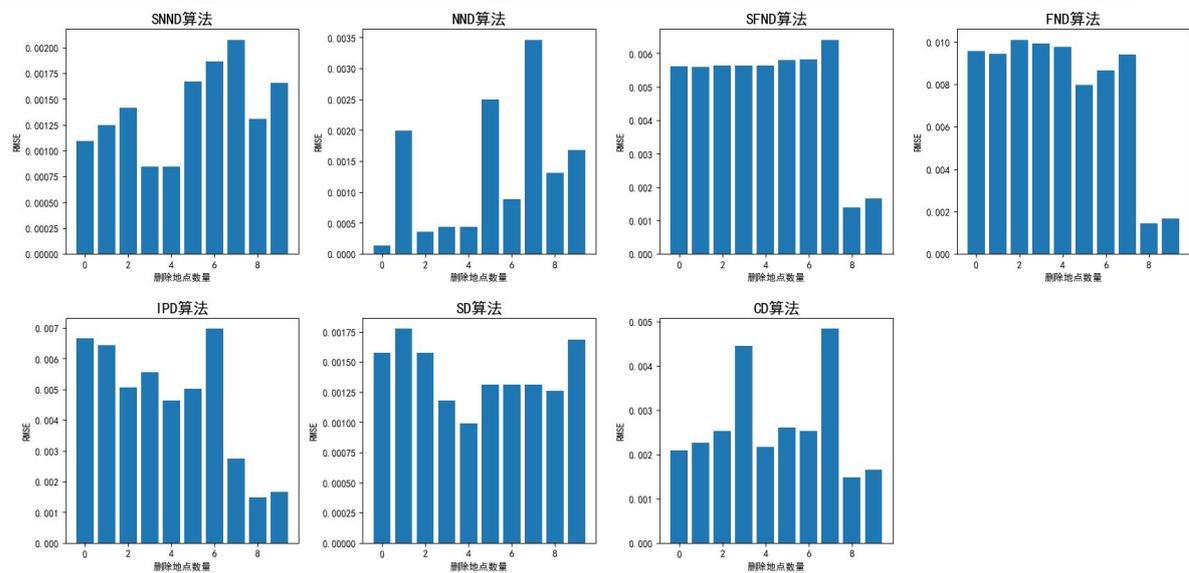


图 18 马家坪案不同算法下删除数据点的模型响应图

表 4 和图 18 是马家坪案不同算法下删除数据点的模型响应。从中可以看到，SNND 算法的 RMSE 值随着数据点的删除呈现波动性增加，但在删除 9 个数据点后，准确度突然提高，这可能是由于模型在特定的数据配置下偶然表现得较好；

NND 算法的准确度在删除数据点时表现出不同程度的波动；SFND 算法和的 RMSE 值在删除 1 到 8 个数据点时稳定性很好；FND 算法在删除 1 到 8 个数据点时稳定性较好；IPD 算法的 RMSE 值在删除数据点时波动较大；SD 算法在删除不同数量的数据点时准确度波动较为缓和，后趋于稳定；CD 算法的 RMSE 值随着数据点的删除呈现上升趋势，但在删除 8 个数据点时准确度显著提高。

综合两组数据，从波动性上看，SNND、NND 算法对数据删除敏感，波动性较高；SFND、FND 算法在删除较少数据点时相对稳定；IPD 算法对数据删除敏感，但波动性不高；SD 算法在第二组中的波动性不同；CD 算法在两组数据中都显示出波动性，但相对其他算法而言，在数据删除较多时保持了更好的稳定性。所以若数据经常丢失或不完整（在收集犯罪地点信息时经常出现），选择 FND 或 CD 算法可能更为合适。

### 5.3.2 现象一以及分析

对于两个案件进行总体比较，发现马家坪案件的预测偏差总体上小于白银杀人案的偏差，如图 19 所示。

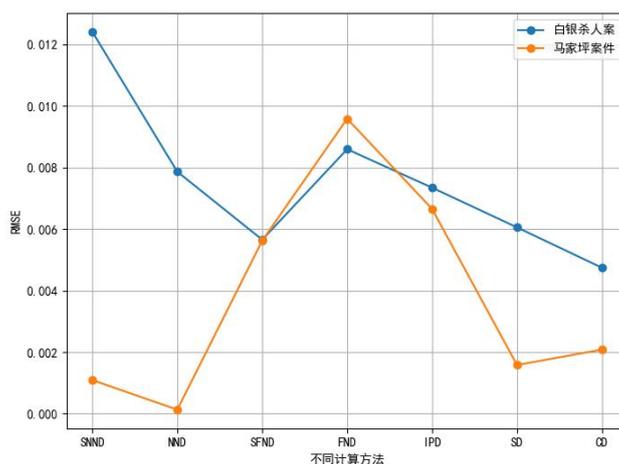


图 19 两个案件预测的 RMSE 比较

首先，我们计算了两个案件参与计算的地图范围（按照 1 经度和 1 纬度的长度都是 111km 来计算），得出白银杀人案的地图面积是 5.222 平方千米，而马家坪案件的地图面积是 1.051 平方千米，可观察到马家坪案件的地图面积小于白银案的地图面积，马家坪案犯罪者的活动区域比白银案的小，作案地点比较集中。

此外，马家坪案件发生地区还有一条较大的河流流经参与计算的区域内，这条河流的出现使得马家坪案件的犯罪者的活动区域更加减少<sup>[11]</sup>。

从这里我们可观察到：CGT 模型不同 B 值预测的结果总体上在较小的区域内具有更好的准确性，所以当我们面对一个已知一系列作案地点的连环案件时，可以通过参与计算的面积的大小来定性看出结果的准确性。

[11] 齐杰：《忻州市刑事案件的犯罪地理学分析》，山西师范大学,2018。

### 5.4.2 现象二以及分析

马家坪案通过 NND 算法计算所得 B 值预测更为准确。

通过对 NND 算法预测的图像与其他算法的预测图像进行对比，发现 NND 算法预测概率分布图像密度集中度更高。NND 算法采用平均最邻近距离，即考虑了相邻犯罪地点之间的距离。这使得模型更贴近实际地理环境，反映了犯罪者在选择作案地点时通常倾向于靠近之前的犯罪地点。当犯罪地点集中形成犯罪热点区域时，NND 算法能更好地捕捉到这些犯罪热点，并提供相对准确的心理缓冲区预测。结合上文提到的马家坪案作案地点较为集中的特点，我们可以分析在 B 值算法选择上，如果多个犯罪地点区域上较为集中，总体地理环境上也相对集中时，选择 NND 算法计算 B 值，可以使预测准确度更高。

### 5.4.3 现象三以及分析

在白银杀人案中，使用 CD 和 SD 算法计算出的 B 值在预测上表现更为准确，如图 19 所示。

白银连环案件属于入室作案类型，参与计算的范围较广，犯罪者的作案时间跨度较长且相对独立。在这种情况下，SNND 等方法计算的 B 值的预测结果不理想。然而，CD 和 SD 算法以犯罪地点构成的点群质心为计算标准点，更适应这一场景，这两种算法的优越性在于更好地捕捉了犯罪者心理地图和归属点的空间特征。

犯罪心理地图是指犯罪者在犯罪活动中与周围环境互动在大脑中形成的对环境的认知心像。这种心理地图反映了犯罪者对熟悉环境的感知、理解以及与该环境的情感联系<sup>[12]</sup>。对于连环案件的犯罪者而言，这种心理地图可能与他们的作案地点和潜藏地点有着紧密的关系。质心是犯罪地点构成的点群的中心点，它代表了这些地点的集中趋势。由于心理地图通常表现为犯罪者在某一区域的集中活动，CD 和 SD 算法通过质心的计算更能反映这种集中趋势，进而更准确地预测犯罪者的心理缓冲区。这使得这两种算法在考虑心理地图的情境下能够提供更为精确的结果。

CD 和 SD 算法的优越性还可归因于对归属点<sup>[13]</sup>的考虑。犯罪者通常有一个或多个固定的归属点，是其可能寻求庇护的地方。这两种算法通过质心的计算，在犯罪者有固定的家庭住址的情况下，CD 和 SD 算法能够更准确地捕捉到犯罪者的心理缓冲区，提高预测结果的可靠性。

因此，当处理犯罪间隔时间较长、犯罪者可能有相对稳定家庭的犯罪情况预测时，CD 或 SD 算法都是较为合适的选择，因为它们更有效地综合了犯罪者的心理地图和归属点的影响，提高了模型的预测精度。

---

<sup>[12]</sup> 虞宗麟，马皓：《论犯罪地理画像的可行性：以白银高承勇连环杀人案为例》，《心理学进展》，2019 年第 9(6)期，第 963-970 页。

## 6 结论

本研究通过应用地理学和犯罪学原理,对白银市连环杀人案和马家坪案的犯罪地点数据进行了深入分析。利用不同的 B 值计算方法(包括 SNND、NND、SFND、FND、IPD、SD 和 CD),研究了 CGT 模型在这两个案件中的稳健性和预测准确性。通过敏感性测试,即逐步删除数据点并计算均方根误差(RMSE),评估了模型对数据完整性的依赖程度。

研究发现,当面对数据缺失时,不同算法的稳健性和对结果影响的敏感度存在显著差异。例如,SNND 和 NND 算法对数据删除较为敏感,其预测准确度随着数据点的缺失而波动较大。而 SFND、FND 在删除较少数据点时,显示出更好的稳定性。特别地,IPD 算法虽对数据缺失敏感,但其波动性较小,说明算法有一定的抗干扰能力。SD 算法在两组数据中的波动性表现不一,可能因案件具体情况而异。CD 算法在数据点缺失较多时,相较于其他算法仍保持较好的稳定性,显示出其对数据缺失的高鲁棒性。

案件的地理环境是影响预测准确性的另一个关键因素。通过对白银连环杀人案和马家坪案件的预测结果进行比较分析,我们发现在白银市连环杀人案中,CD 和 SD 算法表现出色,而在马家坪案件中,NND 算法最准确。造成差异的原因涉及到地理环境。马家坪案参与计算的地理范围相对较小,犯罪者的活动区域更为集中,适合使用 NND 算法计算 B 值。相反,白银案的地图面积较大,CD 和 SD 算法通过质心的计算更能反映犯罪者心理地图和归属点的空间特征,适用于犯罪间隔时间较长、犯罪者可能有相对稳定家庭的情况。在这种情况下,CD 和 SD 算法能更准确地捕捉到犯罪者的心理缓冲区,提高预测结果的可靠性。

使用 CGT 模型模拟犯罪地理分布时,不同 B 值计算方法在不同案件环境下具有一定的适用性:

- 1) 对于入室连环案件,如白银案,建议综合考虑地理环境、犯罪者心理地图和归属点的影响,选择 CD 和 SD 算法进行心理缓冲区计算,以提高准确性。
- 2) 对于随机连环案件,如马家坪案,NND 算法更为适用,尤其当犯罪者的活动区域较为集中时,可以优先考虑使用该方法。

综上所述,不同的犯罪案件环境可能需要选择不同的 B 值计算方法,以提高 CGT 模型的准确性。然而,本研究仍存在一些不足之处,由于犯罪数据难以获取,仅选取了两个案例进行分析,无法涵盖所有可能情况。对于 B 值计算方法的选择,本文只考虑了常见的几种,而未对其他可能的方法进行深入探讨。未来的研究可以通过更多案例的验证和对计算方法的进一步研究,提高犯罪地理学模型的实际应用效果。

## 【参考文献】

- [1]俸卫、高竟淇：《预测连环犯罪区域的地理画像法研究》，《内江师范学院学报》，2017年第32期，第38-42页
- [2]Kan Dai,“Criminal Geographic Target Model Optimal Integrated Distance between Spatial Points of Serial Burglaries,” Geographical Science Research, 2015,4(1), pp. 16-26.
- [3]唐乙鑫,饶亚君：《国内犯罪地理学研究进展》，《法制博览》,2021年第33期，第1-5页。
- [4]李卫红、戴侃、闻磊：《顾及地理因素的犯罪地理目标模型改进方法》，《测绘科学》，2015年。
- [5]陈闻高：《犯罪心理画像论》，《犯罪研究》,2013年第6期，第2-15页。
- [6]郑涵予、李明楚、韦东、陈俊、李晓临：《系列抢劫案中地理学犯罪心理画像的研究》，《法制博览》，2019年第21期，第26-30页。
- [7]Taylor, “P.J,Quantitative methods in geography: An introduction to spatial analysis,” Houghton Mifflin, Boston, 1977,pp.146-147.
- [8]央视新闻，《甘肃白银连环杀人案作案地点：图解高承勇作案手法》，<http://www.mnw.cn/news/shehui/1343706.html>
- [9]《深夜魔影——阳泉马家坪系列杀人案》，<https://baijiahao.baidu.com/s?id=1689692081691100035&wfr=spider&for=pc>
- [10]俸卫、高竟淇：《预测连环犯罪区域的地理画像法研究》，《内江师范学院学报》，2017年第32期，第38-42页。
- [11]齐杰：《忻州市刑事案件的犯罪地理学分析》，山西师范大学，2018。
- [12]虞宗麟，马皓：《论犯罪地理画像的可行性：以白银高承勇连环杀人案为例》，《心理学进展》，2019年第9(6)期，第963-970页。