

DOI:10.13203/j.whugis20170017



文章编号:1671-8860(2018)11-1719-06

Knox 时空交互检验空间阈值确定方法

岳瀚¹ 朱欣焰^{1,2} 芮维^{1,2,3} 余冰¹ 高超³

1 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

2 地球空间信息技术协同创新中心,湖北 武汉,430079

3 警用地理信息技术公安部重点实验室,江苏 常州,210046

摘要: Knox 检验是一种常用的城市犯罪时空交互模式分析方法,但其阈值需要人为指定,这种主观的阈值确定方法存在一定随意性,因此需探索更为合理的阈值确定方法。提出利用点对平均最近距离作为 Knox 检验空间阈值的确定方法,并通过城市入室盗窃、盗窃电动车和扒窃 3 类事件进行实验验证。结果表明,与常见的几种阈值确定方法相比,所提方法检测出了最大数量的显著性交互事件对数,能更加充分地了解事件的真实时空交互模式,为基于 Knox 检验的事件时空聚集模式分析提供了一种有效的空间阈值确定手段。

关键词: Knox 检验;时空交互;空间阈值;时空聚类;城市犯罪

中图分类号: P208

文献标志码: A

城市犯罪往往呈现出时空交互性,即在时间上邻近的事件,在空间上也邻近^[1]。常用的时空交互检验方法有 Knox 指数法^[2]、Mantel 指数法^[3]、Jacquez *K* 邻近法^[4]等。其中,Knox 指数法由于易于理解、运算方便等特性^[5],已被广泛运用于流行病学研究^[3]、公共卫生管理^[6]、犯罪分析^[1]、交通事故分析^[7]、城市管理^[8]等诸多领域。

然而,Knox 检验需要事先确定分析阈值。在缺乏相关知识与参考资料时,Knox 检验阈值常通过人为主观确定。例如,余冰等在研究城市管理事件时空分布时,人为指定时间阈值为 3 h,空间阈值为 500 m^[8];徐冲等选取 200 m 和 14 d 作为街头抢劫案件的时空聚集性检测分析阈值^[9];此外,100 m 是另一个常用的空间阈值,如 Ye 等^[10]、Johnson 等^[11]、Elmes 等^[12]在进行不同事件时空交互性检验时,均以 100 m 作为空间阈值。

在不同阈值下,事件表现出的时空交互性具有较大差异,而主观确定阈值的方法具有一定随意性,基于该阈值的 Knox 检验不一定能反映出事件真实的时空交互模式^[13]。因此需要寻找一种阈值确定方法,使得 Knox 检验在该阈值下能最大限度地体现事件的时空聚集模式。为此,一

些学者根据已有知识确定 Knox 检验阈值,例如刘巧兰等在利用 Knox 检验探测传染病时空聚集性时,根据流行性脑膜炎发病潜伏期及病例之间的平均空间距离,确定时间阈值为 13 d,空间分析阈值为 29.67 km^[14];Townsend 等利用平均街区长度作为 Knox 检验的空间阈值^[15]。

此外,已有研究提出若干空间阈值确定方法,如 Grubestic 等利用尝试法不断改变时空阈值并依次计算时空交互程度,据此得到最合适的时空阈值^[1],但该重复性实验过程会造成系统性偏差^[13];Kalantari 等比较了平均距离法、Ripley's *K* 函数临界值法、自然断点分级法这 3 类阈值确定方法的效果,结果表明自然断点分级法确定的空间阈值能更有效地探测出案件时空聚集性^[13],然而该方法需要人工指定分类数,且分类结果容易受异常值影响^[16]。

本文基于事件空间邻近性这一重要特征,提出以点对平均最近距离作为 Knox 检验的空间阈值,以武汉市江汉区唐家墩社区 2013 年 1 月—8 月期间的入室盗窃、盗窃电动车和扒窃事件为研究对象进行实验论证,并与常见的几种阈值确定方法进行对比,结果表明本文所提出的 Knox 阈值确定方法能检测出更多的显著性交互事件对

收稿日期:2017-08-04

项目资助:警用地理信息技术公安部重点实验室开放基金(2016LPGIT05)。

第一作者:岳瀚,博士,主要从事时空数据分析的理论与方法研究。hanygeo@163.com

通讯作者:芮维,副教授。guowei-lmars@whu.edu.cn

数,从而更有效地识别出事件的时空聚集模式。

1 Knox 时空交互检验模型

Knox 时空交互检验模型首先需要人工指定时间阈值 t 和空间阈值 s ,继而将所有 n 个事件点两两配对成 $n(n-1)/2$ 个时空事件对,并计算空间距离 s_{ij} 和时间距离 t_{ij} 。当 $s_{ij} \leq s$ 时,认为事件点空间邻近;当 $t_{ij} \leq t$ 时,认为事件点时间邻近。通过判断事件对是否时空邻近,可将事件对映射到一个 2×2 的矩阵中计数,如表 1 所示。表 1 中 Q_1 为时空邻近事件对数,也即 Knox 指数。

表 1 Knox 指数逻辑结构

Tab.1 Logical Structure of Knox Index

| 类别 | 时间邻近 | 时间非邻近 |
|-------|--------------|--------------|
| 距离邻近 | Q_1 (时空邻近) | Q_2 (空间邻近) |
| 距离非邻近 | Q_3 (时间邻近) | Q_4 (非邻近) |

得到事件时空邻近结构后,常用卡方检验对其统计显著性进行检验^[2]。卡方检验假设事件不存在时空交互性,在此零假设前提下计算 4 类邻近情况 Q_1 、 Q_2 、 Q_3 、 Q_4 的期望值,通过观测值 Q_i 和期望值 $E(Q_i)$ 的差异程度可以判断时空邻近的显著性。

如表 1 所示,当且仅当事件对空间邻近且时间邻近时才存在时空交互性。记事件对总数为 $N=Q_1+Q_2+Q_3+Q_4$,记 $S_1=Q_1+Q_2$, $S_2=Q_3+Q_4$, $S_3=Q_1+Q_3$, $S_4=Q_2+Q_4$ 。在零假设条件下,不存在时空交互作用,则 Q_1 至 Q_4 的期望值分别为: $E(Q_1)=S_1 \times S_3/N$, $E(Q_2)=S_1 \times S_4/N$, $E(Q_3)=S_2 \times S_3/N$, $E(Q_4)=S_2 \times S_4/N$ 。通过比较实际时空事件对数与期望事件对数的差异,构造卡方检验统计量:

$$\chi^2 = \sum_{i=1}^4 \frac{(Q_i - E(Q_i))^2}{E(Q_i)} \quad (1)$$

通过卡方值 χ^2 即可判断时空交互统计显著性。

2 Knox 检验空间阈值选择方法

在 Knox 检验中,时空阈值的选择对于实验结果具有决定性作用,因为它决定了在某一时空尺度下零假设将会被拒绝或接受。在已有研究中,空间阈值通常由经验确定或借鉴其他领域研究结果,例如 100 m 常作为各类事件时空交互检验的空间阈值;此外,一些学者提出了相关阈值确定方法。

2.1 平均距离法

对于 n 个事件点集,平均距离法首先分别计算点 i 与其他 $n-1$ 个点的距离,得到 $n-1$ 个距离值并对其求平均;对每个点均重复上述过程,得到每个点与其他点距离的平均值;对 n 个点的该平均值再次取平均,得到平均距离,该平均距离值即作为 Knox 检验的空间阈值^[13]。

2.2 平均 Ripley's K 函数临界值法

空间 Ripley's K 函数(简称 K 函数)是一种多距离空间聚类分析方法^[17],通过给定距离范围,K 函数能以一系列连续尺度确定点要素或与其相关联的属性值在该范围内随距离变化的聚集性或离散性变化情况。

通过判断 K 函数观测曲线与零假设条件下期望值包络线的交点,可以确定聚集模式与离散模式对应距离的临界值,该临界值可以作为 Knox 检验的空间阈值。K 函数是一种累积性的计算过程,大尺度的分析结果包含了小尺度信息,因此混淆了大尺度与小尺度效应^[18],以 K 函数临界值作为 Knox 检验空间阈值可能无法得到较为显著性的交互模式。

2.3 自然断点分级法

自然断点分级法是一种单变量分类方法,分类原则为使分类对象类内方差最小、类间方差最大。这种分类方法利用了数据自身固有的自然分组,可对相似变量进行最恰当的归类,同时使得各类之间的差异最大,并在变量差异相对较大处设置分类界限^[16]。

在利用 Knox 检验分析 3 类犯罪案件的时空聚集性时,采用自然断点分级法将点要素最邻近距离集合分为 5 级,并以分级界限作为 Knox 检验的空间阈值^[13]。自然断点分级法的缺陷在于需要指定分类数 k ,分类结果受 k 影响较大,且容易产生异常结果^[16]。

2.4 平均最邻近距离法

地理学第一定律表明,邻近事件比疏远事件存在更为紧密的关系^[19],因此,两事件越邻近,其空间交互性越大。传统的阈值主观确定法存在一定的随意性,而上述两种阈值确定方法均存在不足。在没有其他参考资料的情况下,以事件最邻近距离为依据确定 Knox 检验的空间阈值是一种较为合理的方法。因此,本文提出基于平均最邻近距离的 Knox 检验空间阈值确定方法。首先计算每个点 i 与其最邻近点 j 的距离 $\min(d_{ij})$,进而对所有的最邻近距离求平均,得到平均最邻近距离^[20],并以该值作为 Knox 检验的空间阈值。

平均最邻近距离的计算公式为:

$$\bar{d} = \frac{\sum_{i=1}^n \min(d_{ij})}{n} \quad (2)$$

式中, d_{ij} 为点 i 与点 j 之间的距离; n 为事件点总数。

3 实验与分析

本文使用武汉市江宁区唐家墩社区 2013 年 1 月—8 月发生的入室盗窃、盗窃电动车和扒窃案件进行实验分析。通过指定不同的时空阈值, Knox 检验可以用来检测事件在不同时间尺度和不同空间尺度下的分布特征。在案件时空交互特征未知的情况下, 分析不同空间阈值确定方法在 30 d 内检测出的案件时空交互特征^[1,13]。 p 值常用来判断时空聚集性检验结果的统计显著性。本文研究中, 选取常用的 95% 置信度(即 $p < 0.05$) 作为显著性判断标准。

3.1 常用阈值法

在许多研究领域, 100 m 常用来作为 Knox 检验的空间阈值。图 1 为 3 类案件在 100 m 空间阈值下 30 d 内的 Knox 检验 p 值分布曲线。



图 1 100 m 空间阈值下 Knox 检验结果

Fig.1 Knox Test Results for 30 Days with 100 m Spatial Threshold

3.2 平均距离法

利用平均距离法确定 3 类案件的空间阈值, 基于该距离的 Knox 检验结果如图 2 所示。

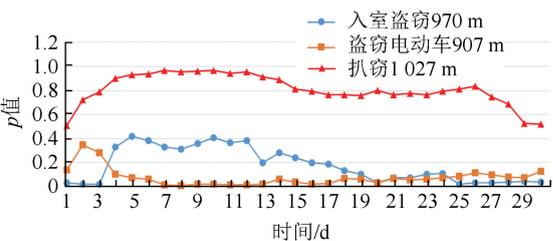


图 2 平均距离法 Knox 检验结果

Fig.2 Knox Test Results Based on Mean Distance

3.3 Ripley's K 函数临界值法

为了检验 K 函数结果的显著性, 本文利用蒙

特卡罗方法进行 999 次模拟实验, 得到置信度为 99% 的包络线。结果表明, 入室盗窃、盗窃电动车和扒窃案件分别在 375 m、438 m 和 430 m 空间范围内具有显著的空间聚集性。以 3 类案件的空间 K 函数临界值作为空间阈值得到 Knox 检验结果, 如图 3 所示。

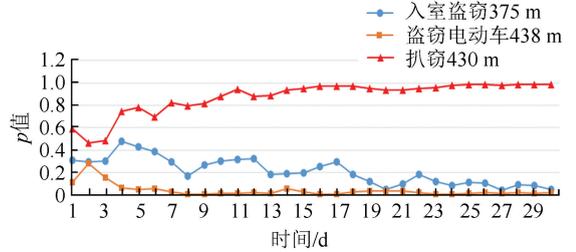
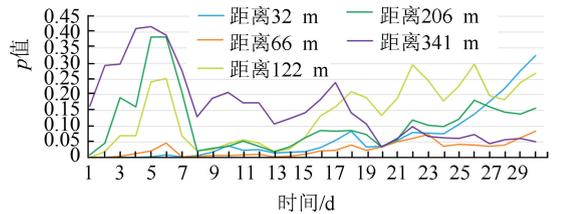


图 3 Ripley's K 函数临界值法 Knox 检验结果

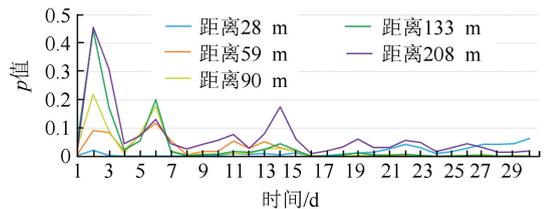
Fig.3 Knox Test Results Based on Ripley's K Function Threshold

3.4 自然断点分级法

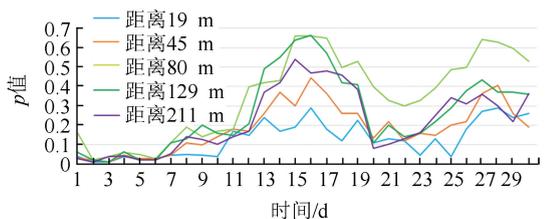
本文利用自然断点分级法将最邻近距离集合分为 5 级, 得到 3 类案件的分级界限分别为: 入室盗窃 32 m、66 m、122 m、206 m、341 m; 盗窃电动车 28 m、59 m、90 m、133 m、208 m; 扒窃 19 m、45 m、80 m、129 m、211 m。图 4 为 3 类案件在各自分级界限处的 Knox 时空交互检验结果。



(a) 入室盗窃



(b) 盗窃电动车



(c) 扒窃

图 4 自然断点分级法 Knox 检验结果

Fig.4 Knox Test Results Based on Natural Breaks Classification

3.5 平均最邻近距离法

通过式(2)计算得到3类案件的平均最邻近距离分别为:入室盗窃50 m,盗窃电动车43 m,扒窃36 m。图5为3类案件以平均最邻近距离为空间阈值的时空检验结果。结果表明,与上述3种常用阈值选择方法相比,在平均最邻近距离尺度下,3类案件均表现出更为显著的聚集性特征。



图5 平均最邻近距离法Knox检验结果

Fig.5 Knox Test Results Based on Average Nearest Neighbor Distances

3.6 结果分析

实验结果表明,在不同的时空阈值下,事件时空交互特征差异较大。为了比较5种阈值确定方法的效果,表2列出了3类案件在不同空间阈值下,1~30 d内具有显著交互性的事件对数及其平均数、总平均数;表3列出了不同空间阈值确定方法检验得到的3类案件的显著性时空交互对数总平均数。

Kalantari等^[13]在利用Knox检验分析汽车配件盗窃、商店盗窃、摩托车盗窃这3类常见盗窃事件的时空交互模式时指出,某阈值下得到的时空显著性交互事件对数越多,表明该阈值对应的事件时空交互模式越显著,以该尺度进行进一步分析,能更充分地了解事件的真实分布模式,掌握不同事件的发生和分布规律,为有效地对事件进行监管、最大程度地减少有害事件的发生提供决策依据。因此,利用某方法确定的空间阈值检验

表2 各空间阈值确定方法在1~30天内检验出的显著性时空交互事件对数

Tab.2 Frequencies of Significant Interactions of Five Methods at Different Intervals for 30 Days

| 空间阈值确定方法 | 入室盗窃 | | | | 盗窃电动车 | | | | 扒窃 | | | |
|-------------------|--------|---------------------------|------|-------|--------|---------------------------|------|-------|--------|---------------------------|-----|------|
| | 空间阈值/m | 1~30 d内显著性交互事件对数 (p<0.05) | 平均数 | 总平均数 | 空间阈值/m | 1~30 d内显著性交互事件对数 (p<0.05) | 平均数 | 总平均数 | 空间阈值/m | 1~30 d内显著性交互事件对数 (p<0.05) | 平均数 | 总平均数 |
| 常用阈值法 | 100 | 13 | 13 | | 100 | 26 | 26 | | 100 | 4 | 4 | |
| 平均距离法 | 970 | 10 | 10 | | 907 | 11 | 11 | | 1 027 | 0 | 0 | |
| Ripley's K 函数临界值法 | 375 | 3 | 3 | | 438 | 23 | 23 | | 430 | 0 | 0 | |
| 自然断点分级法 | 32 | 18 | | 13.08 | 28 | 29 | | 22.88 | 19 | 10 | | |
| | 66 | 25 | | | 59 | 23 | | | 45 | 8 | | 6.12 |
| | 122 | 8 | 12.4 | | 90 | 26 | 24.4 | | 80 | 5 | 6.6 | |
| 平均最邻近距离法 | 206 | 9 | | | 133 | 25 | | | 129 | 6 | | |
| | 341 | 2 | | | 208 | 19 | | | 211 | 9 | | |
| 常用阈值法 | 50 | 27 | 27 | | 43 | 30 | 30 | | 36 | 20 | 20 | |

表3 5种空间阈值确定方法检验出的3类案件显著性时空交互对数总平均数

Tab.3 Mean Values of Significant Interactions of Five Methods for Three Types of Crimes

| 空间阈值确定方法 | 3类案件显著交互对数平均数 | 3类案件显著交互对数方差 |
|-------------------|---------------|--------------|
| 常用阈值法 | 14.3 | 122.33 |
| 平均距离法 | 7 | 37 |
| Ripley's K 函数临界值法 | 8.67 | 156.33 |
| 自然断点分级法 | 14.47 | 82.41 |
| 平均最邻近距离法 | 25.67 | 26.33 |

出的显著性时空交互对数越大,即表明该空间阈值确定方法能越有效地识别出案件的时空交互特

征^[13]。表2表明,盗窃电动车案件的时空邻近事件对数总平均数最大(22.88),即该案件具有较强的时空交互程度,其次是入室盗窃(13.08)和扒窃案件(6.12)。表3表明,本文提出的平均最邻近距离阈值确定方法检验得到的3类案件的时空邻近事件对数平均数最大(25.67),其次是自然断点法(14.47)、常用阈值法(14.3)、Ripley's K 函数临界值法(8.67)和平均距离法(7);另外,3类案件的显著交互对数方差体现了该阈值确定方法在探测不同类型的案件时空交互性的稳定性,结果表明本文所提方法具有更好的稳定性(方差为26.33,最小)。因此,本文所提方法能更有效地检测出案件的时空交互模式,且具有较好的稳定性。

4 结 语

在利用 Knox 检验进行事件时空交互模式分析时,阈值的选择对于实验结果具有决定性作用。本文在分析几种常见阈值选择方法的缺陷的基础上,提出了利用平均最邻近距离作为 Knox 检验空间阈值确定的方法。实验通过对城市入室盗窃、盗窃电动车和扒窃 3 类案件的时空交互模式进行分析,结果表明,由于顾及到了事件对间的空间邻近性,基于该阈值确定方法的 Knox 检验检测出了最大数量的显著性交互事件对数,能够有效识别出城市犯罪时空交互模式及邻近重复发生事件对的分布区域。本文所提方法能够为基于 Knox 检验的事件时空聚集模式分析提供一种有效的空间阈值确定手段,进而为城市犯罪的重点时段、重点区域的针对性防控提供参考依据^[21],以便优化警力资源部署,减少或抑制犯罪的发生^[22]。未来将研究本文所提方法在一维路网空间约束下的运用效果,并进一步探讨其他类型城市犯罪的时空交互模式及其形成原因。

参 考 文 献

- [1] Grubestic T H, Mack E A. Spatio-Temporal Interaction of Urban Crime [J]. *Journal of Quantitative Criminology*, 2008, 24(3):285-306
- [2] Knox E G, Bartlett M S. The Detection of Space-Time Interaction [J]. *Applied Statistics*, 1964, 13(1):25-29
- [3] Mantel N. The Detection of Disease Clustering and a Generalized Regression Approach [J]. *Cancer Research*, 1967, 27(2):209-220
- [4] Jacquez G M. A K Nearest Neighbor Test for Space-Time Interaction [J]. *Statistics in Medicine*, 1996, 15(18):1 935-1 949
- [5] Kulldorff M, Hjalmars U. The Knox Method and Other Tests for Space-Time Interaction [J]. *Biometrics*, 1999, 55(2):544-552
- [6] Sonesson C, Bock D. A Review and Discussion of Prospective Statistical Surveillance in Public Health [J]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2003, 166(1):5-21
- [7] Black W R. Transportation: A Geographical Analysis [M]. New York: Guilford Press, 2003
- [8] She Bing, Zhu Xinyan, Guo Wei, et al. Spatial Distribution and Evolution of City Management Events Based on the Spatial Point Pattern Analysis: A Case Study of Jiangnan District, Wuhan City [J]. *Progress in Geography*, 2013, 32(6):924-931(余冰,朱欣焰,芮维,等.基于空间点模式分析的城市管理事件空间分布及演化——以武汉市江汉区为例[J].地理科学进展,2013,32(6):924-931)
- [9] Xu Chong, Liu Lin, Zhou Suhong. Patterns of Near-Repeat Street Robbery in DP Peninsula [J]. *Geographical Research*, 2015, 34(2):384-394(徐冲,柳林,周素红. DP 半岛街头抢劫案件的临近重复发生模式[J].地理研究,2015,34(2):384-394)
- [10] Ye X, Xu X, Lee J, et al. Space-Time Interaction of Residential Burglaries in Wuhan, China [J]. *Applied Geography*, 2015, 60:210-216
- [11] Johnson S D, Bowers K J. The Stability of Space-Time Clusters of Burglary [J]. *The British Journal of Criminology*, 2004, 44(1):55-65
- [12] Elmes G, Roedl G. The Use of Geospatial Information Technology to Advance Safer College Campuses and Communities [M]//Leitner M. Crime Modeling and Mapping Using Geospatial Technologies. Netherlands: Springer, 2013
- [13] Kalantari M, Yaghmaei B, Ghezelbash S. Spatio-Temporal Analysis of Crime by Developing a Method to Detect Critical Distances for the Knox Test [J]. *International Journal of Geographical Information Science*, 2016,30(11):1-19
- [14] Liu Qiaolan, Li Xiaosong, Feng Zijian, et al. Study on the Application of Knox Method to Temporal-Spatial Cluster for Infectious Diseases [J]. *Chinese Journal of Epidemiology*, 2007, 28(8):802-805(刘巧兰,李晓松,冯子健,等. Knox 方法在传染病时空聚集性探测中的应用[J].中华流行病学杂志,2007,28(8):802-805)
- [15] Townsley M, Homel R, Chaseling J. Infectious Burglaries: A Test of the Near Repeat Hypothesis [J]. *The British Journal of Criminology*, 2003, 43(3):615-633
- [16] de Smith M J, Goodchild M F, Longley P. Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools [M]. 2nd ed. Leicester, United Kingdom: The Winchelsea Press, 2007
- [17] Ripley B D. The Second-Order Analysis of Stationary Point Processes [J]. *Journal of Applied Probability*, 1976, 13(2):255-266
- [18] Zhang Jian, Hao Zhanqing, Song Bo, et al. Spatial Distribution Patterns and Associations of Pinus Koraiensis and Tilia Amurensis in Broad-Leaved Korean Pine Mixed Forest in Changbai Mountains [J]. *Chinese Journal of Applied Ecology*, 2007, 18(8):1 681-1 687(张健,郝占庆,宋波,等.长白山阔叶红松林中红松与紫椴的空间分布格局及其关

- 联性[J]. 应用生态学报, 2007, 18(8):1 681-1 687)
- [19] Tobler W R. A Computer Movie Simulating Urban Growth in the Detroit Region [J]. *Economic Geography*, 1970, 46(sup1): 234-240
- [20] Clark P J, Evans F C. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations [J]. *Ecology*, 1954, 35(4):445-453
- [21] Zhu Xinyan, Zhou Chenghu, Guo Wei, et al. Preliminary Study on Conception and Key Technologies of the Location-Based Pan-Information Map [J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(3): 285-295 (朱欣焰, 周成虎, 芮维, 等. 全息位置地图概念内涵及其关键技术初探[J]. 武汉大学学报·信息科学版, 2015, 40(3): 285-295)
- [22] Duan Lian, Ye Xinyue, Hu Tao, et al. Prediction of Suspect Location Based on Spatiotemporal Semantics [J]. *ISPRS International Journal of Geo-Information*, 2017, 6(7): 185-1-185-27

A Method for Determining the Critical Spatial Threshold of Spatio-Temporal Interaction for the Knox Test

YUE Han¹ ZHU Xinyan^{1,2} GUO Wei^{1,2,3} SHE Bing¹ GAO Chao³

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

2 Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China

3 Key Laboratory of Police Geographic Information Technology, Ministry of Public Security, Changzhou 210046, China

Abstract: City crimes tend to be clustered spatially and temporally. Knox test is an effective way to detect such interactions, however, this method has been criticized as being subjective because the determination of critical distances is arbitrary. This paper begins with an introduction of the Knox test and some common methods to select the spatial thresholds. Then a new approach is proposed to detect critical distances based on the average nearest neighborhood distance. Burglary, electric vehicle theft and pick pocketing events in Wuhan city are used as experimental data to validate the method. Results show that, compared with four common criteria, i.e., empirical distance, mean distance, Ripley's K function threshold and natural breaks classification, the method we proposed is able to detect spatio-temporal interaction patterns of different events more effectively. The analysis results provide a reference for selecting critical distances for the Knox test.

Key words: Knox test; spatio-temporal interaction; spatial threshold; spatio-temporal clustering; city crimes

First author: YUE Han, PhD, specializes in the theories and methods of spatio-temporal data analysis. E-mail: hanygeo@163.com

Corresponding author: GUO Wei, associate professor. E-mail: guowei-lmars@whu.edu.cn

Foundation support: The Open Research Fund Program of Key Laboratory of Police Geographic Information Technology, Ministry of Public Security, No.2016LPGIT05.