

Detecting Crime Hot Spots Using GAM and Local Moran's I

Jinseong Cheong

Department of Police Administration
Soon Chun Hyang University, Asan, Chungnam, 336-745, Korea

ABSTRACT

Scientific analysis of crime hot spots is essential in preventing and/or suppressing crime. However, results could be different depending on the analytic methods, which highlights the importance of choosing adequate tools. The purpose of this study was to introduce two advanced techniques for detecting crime hot spots, GAM and Local Moran's I, hoping for more police agencies to adopt better techniques. GAM controls for the number of population in study regions, but local Moran's I does not. That is, GAM detects high crime rate areas, whereas local Moran's I identifies high crime volume areas. For GAM, physical disorder was used as a proxy measure for population at risk based on the logic of the broken windows theory. Different regions were identified as hot spots. Although GAM is generally regarded as a more advanced method in that it controls for population, its usage is limited to only point data. Local Moran's I is adequate for zonal data, but suffers from the unavoidable MAUP (Modifiable Areal Unit Problem).

Keywords: GAM (Geographical Analysis Machine), Local Moran's I, Crime Hot Spots, Physical Disorder, Broken Windows Theory

1. INTRODUCTION

It is widely known that criminal justice practitioners and researchers have long realized the importance of identifying unusual crime clusters (hereinafter "hot spots") that may facilitate efficient distribution of police manpower and material resources and help theoreticians to explore the underlying socioeconomic conditions around the hot spot areas[1]. Unfortunately, however, such simple methods as plotting crime locations on a base map[2] and visually interpreting the incident distribution[3] still remain as the preferred technique for most law enforcement agencies in Korea. It would not be easy to exactly interpret the point pattern because, for example, some repeated incidents on the same or close location may be represented as one incident[1]. Although the Korean police have been trying to adopt more sophisticated computer-aided detection techniques, more advanced and data-relevant methods must be addressed to overcome the misleading interpretations of the simple visual techniques.

The Broken Windows theory, conceptual background of this study, posits that physical disorder (e.g., weed, trash, poorly maintained houses, etc.) increases serious crime if it remains unfixed for a while[4]. Such macrolevel criminological inquiry must consider spatial dynamics of crime and its related multivariate factors such as concentrated disadvantage, immigrant concentration, residential instability, and so on[5][6]. As a preliminary step, however, this study focused on exploring hot spots of serious violent crime using two advanced methods,

GAM (Geographical Analysis Machine)[7] and Local Moran's I[8]. The primary independent variable of interest, physical disorder, served as a proxy measure for the underlying variation of population at risk for the analysis of GAM.

In sum, the purpose of this study was to introduce the two advanced techniques for detecting crime hot spots, GAM and Local Moran's I, hoping for more police agencies to adopt better techniques. To this end, data collected in Lansing, Michigan, were analyzed. Since the two techniques use different types of data¹, the author further attempted to examine if there is any substantial location difference between the hot spots identified by the two methods. Prior to analyses, brief description of data, measure, and distribution is introduced with some visualizing maps.

2. DATA AND MEASURE

Violent crime data were obtained from the Lansing Police Department Incident Report for 2002. There were 1036 violent incidents including homicide, aggravated assault, rape, and robbery, which are all serious FBI index crimes. The reason the author aggregated only serious index crimes came from the logic of the Broken Windows metaphor that using non-serious crime as the response variable may end up with tautology because the physical and social disorder (another primary independent variable in the theory: e.g., panhandling, fighting, drug dealing, shouting, prostitution, etc.) appear to measure the same phenomena as the dependent variable does[9].

The violent crime data were stored in two different formats

* Corresponding author, Email: cjs2047@sch.ac.kr

Manuscript received Mar. 09, 2012; revised Jun 18, 2012;
accepted Jun 20, 2012

¹ GAM uses point pattern data having information on exact locations and Local Moran's I requires aggregated zonal data.

for the two techniques respectively: point data and areal data. The city of Lansing has 115 Census blockgroups and the 1036 violent crimes were aggregated into the corresponding zones. Fig. 1 represents the dot and choropleth maps of Lansing violent crime for the year 2002². Overall, several areas including the “Chestnut area” (northwestern area of the State Capitol) appeared to suffer from the concentrated violent crime problem. Although the choropleth map generally showed a similar pattern to the dot map, some blockgroups (particularly big ones) seemed to exaggerate or understate the actual concentration.

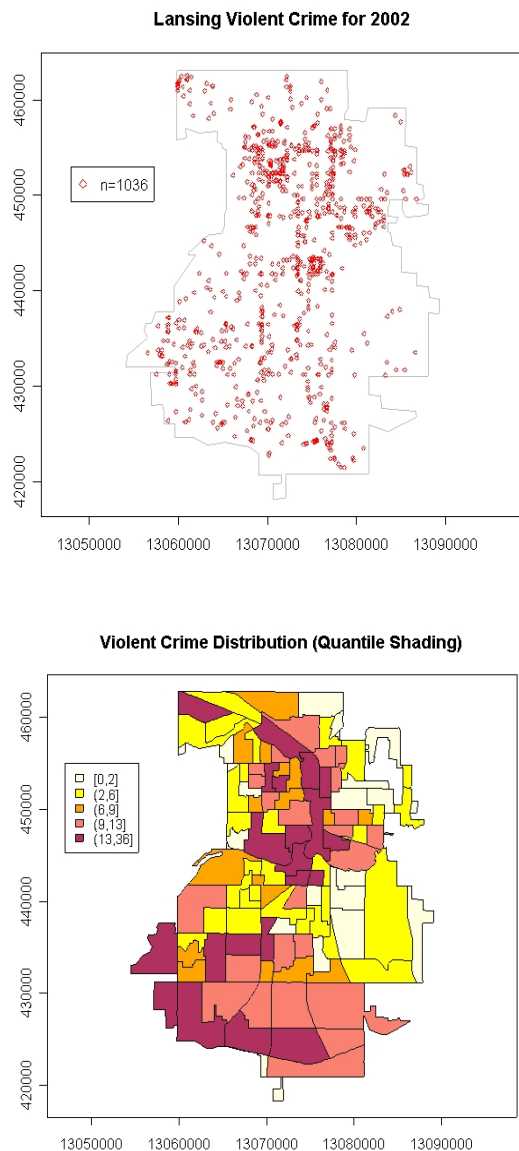


Fig. 1. Dot and choropleth maps of Lansing violent crime for 2002

² Free-distributed statistical program, R, and spatial statistical program, ArcView were used for this paper. Boundary of the first map (one polygon) was created by using “getpoly” function in splancs library and the second base map was transformed from a shape in ArcView to a polygon object by using “Map2poly” function in maptools library.

This could partly be related to the shading scheme³ or the misleading issue of the visualized dot map. Also, the infamous Modifiable Areal Unit Problem (MAUP) originated from the arbitrarily defined Census tract areas could be another factor[10]. These intrinsic problems of visualizing techniques made it necessary to adopt more sophisticated exploratory methods to identify more realistic hot spots in a more systematic and scientific way.

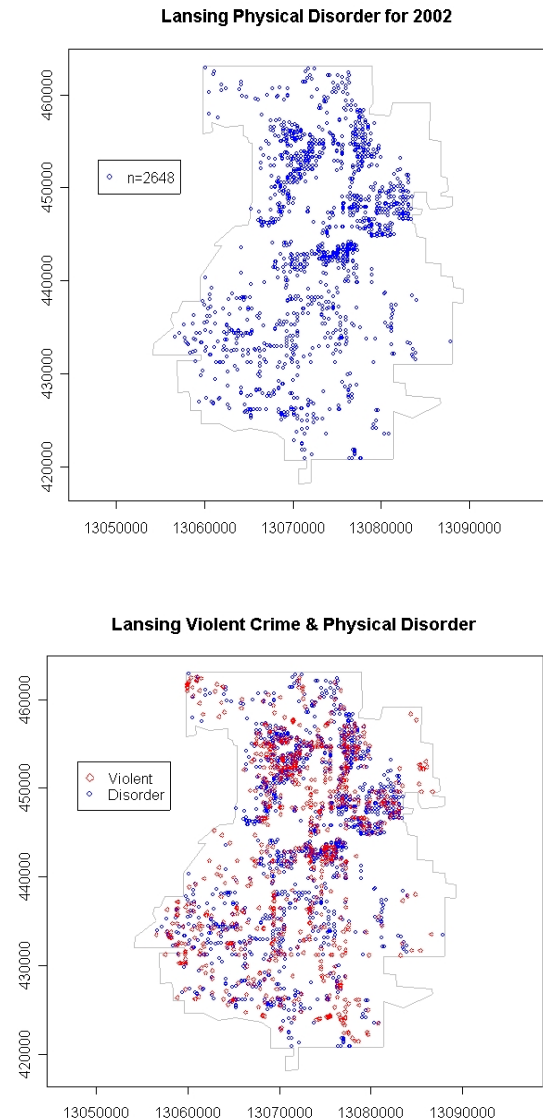


Fig. 2. Dot maps of Lansing physical disorder and violent crime for 2002

On the other hand, the physical disorder information was acquired from the Lansing Code Compliance Department. It contained weed, trash, damaged motor vehicles, and tagged houses, all of which were appropriate indicators of physical disorder according to the Broken Windows theory[4][9]. A total of 2648 cases were reported in 2002. The first map of Fig. 2

³ Quantile shading was used in this case.

shows the point distribution of physical disorder and the second one represents the distributions of both violent crime and physical disorder. Although the overall distributions of violent crime and physical disorder looked similar, the latter appeared to be more dispersed, and some areas seemed to have relatively higher level of crime than physical disorder.

Since the physical disorder was used as a proxy measure of underlying population at risk, the 2648 cases had to represent a random sample of the actual pattern of physical disorder. Unfortunately, however, it was more like a population, and the plausible critical issue was the “dark figure” problem that is prevalent in crime-related research area. That is, some cases could have been dropped out of the reporting and further the reporting could have been biased against poor neighborhoods. However, it would be much less problematic than that of crime, for the physical disorder is generally perceived as a less serious problem.

3. METHODS AND ANALYSIS

3.1 GAM

GAM has been developed by Openshaw and his colleagues since 1987 as an automated detector of local hot spots for point pattern data. It has a good level of academic respectability and also the visualized results can be easily interpreted by practitioners without having much expert knowledge. Basically, it looks for hot spots centered at points on a dense grid mesh across the study region, controlling for the underlying distribution of population at risk. Except the “dark figure” problem, as shown above, the physical disorder could well serve as an appropriate proxy measure of population distribution at risk for it is more common than violent crime. In a sense, further, assuming that population values are centered at specific points and treating them as a point pattern would be problematic, which provides another flexible reason to using the spatial process of physical disorder as a surrogate for the population distribution at risk[10].

First, the author calculated the overall background rate for the events (violent crime) by dividing the number of events by the total sum of number of events and controls (population at risk = physical disorder). The background rate for the violent crime was .281⁴. Next, this background rate was used to calculate the expected number of events for each circle by multiplying the total sum of events and controls by the background rate⁵. Finally, the expected number of events was compared to the observed number of events within each circle using the Poisson distribution⁶, which produced a probability level (p-value) indicating how much the two numbers are different. When the p-value for a circle was less than .05 (as a general rule), the circle was retained as an anomaly that have anomalously high rates of violent crime. Meanwhile, using the kernel smoothing method⁷, the results could be visualized. The

kernel technique allowed the author to change the radius (bandwidth) of the kernel disc, which heavily influenced the resulting number of anomalies and maps. Fig. 3 shows six different maps that have different radius from 500 meters to 3000 meters. Readers must note that each map has different number of anomalies⁸. It was clear that as the bandwidth gets larger, the kernel map gets more smoothed. The first and second maps seemed to be too picky and the fifth and sixth maps appeared to be too smoothed. Overall, the third and fourth map appeared to best reflect the real hot spots of violent crime. Several areas including the Chestnut and State Capitol areas were consistently shown to be hot spots, which were compared to their counterparts produced by the Local Moran's I later in this section.

This method is certainly a useful tool for identifying potential hot spots of crime, but has several problems to be addressed. First, GAM is an exploratory method that cannot statistically confirm whether the detected locations are real hot spots or not. Also, in relation to the first issue, since only one variable, or physical disorder, is controlled for, the result is more like correlation than causation. Further, varying background rates may produce different results. That is, while the above results were produced using the citywide rate as control, it would also have been possible to use the rate for the corresponding Census blockgroup to the grid points. Finally, the important underlying assumption of independence is violated for the circles are overlapped, which causes the multiple testing problem.

3.2 Moran's I

Often times it is impossible to perform the above point pattern analysis due to the unavailability of point data. In particular, macrolevel research in social science almost always uses aggregated data to a certain unit of analysis (e.g., street block, Census tract, city, county, state, etc.). Moran's I is one of the most commonly used exploratory methods for those aggregated zonal data to identify the second-order spatial process. More specifically, while global Moran's I provides a general notion of spatial autocorrelation, local Moran's I shows where within a region such association happens⁹[8]. Below are the formula for global I and local I.

$$I = \frac{N \sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j w_{ij}) \sum_i (X_i - \bar{X})^2}$$

$$I_i = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1, j \neq i}^N w_{ij} (x_j - \bar{x})$$

provides a useful visual indication of the overall trend of the anomalous circles[10]. However, it is different from the kernel ratio map that simply indicates violent crime spots that are not explained by the physical disorder, which does not necessarily tell actual hot spots of crime.

⁸ 222 for R=500, 330 for R=1000, 346 for R=1500, 392 for R=2000, 386 for R=2500, 388 for R=3000

⁹ According to Anselin[8], the local I has two functions: it may be used to identify local clusters and also outliers that heavily contribute to the global I. However, this study focused on identifying local clusters.

⁴ 1036 / (1036 + 2648) = 1036 / 3684 = .2812

⁵ Expected number of violent crime for each circle = (violent crime + physical disorder) * .2812

⁶ “ppois” function in spatstat library was used.

⁷ The function of “kernel2d” in splancs library in R was used to estimate kernel intensity in this study. It employs edge correction and

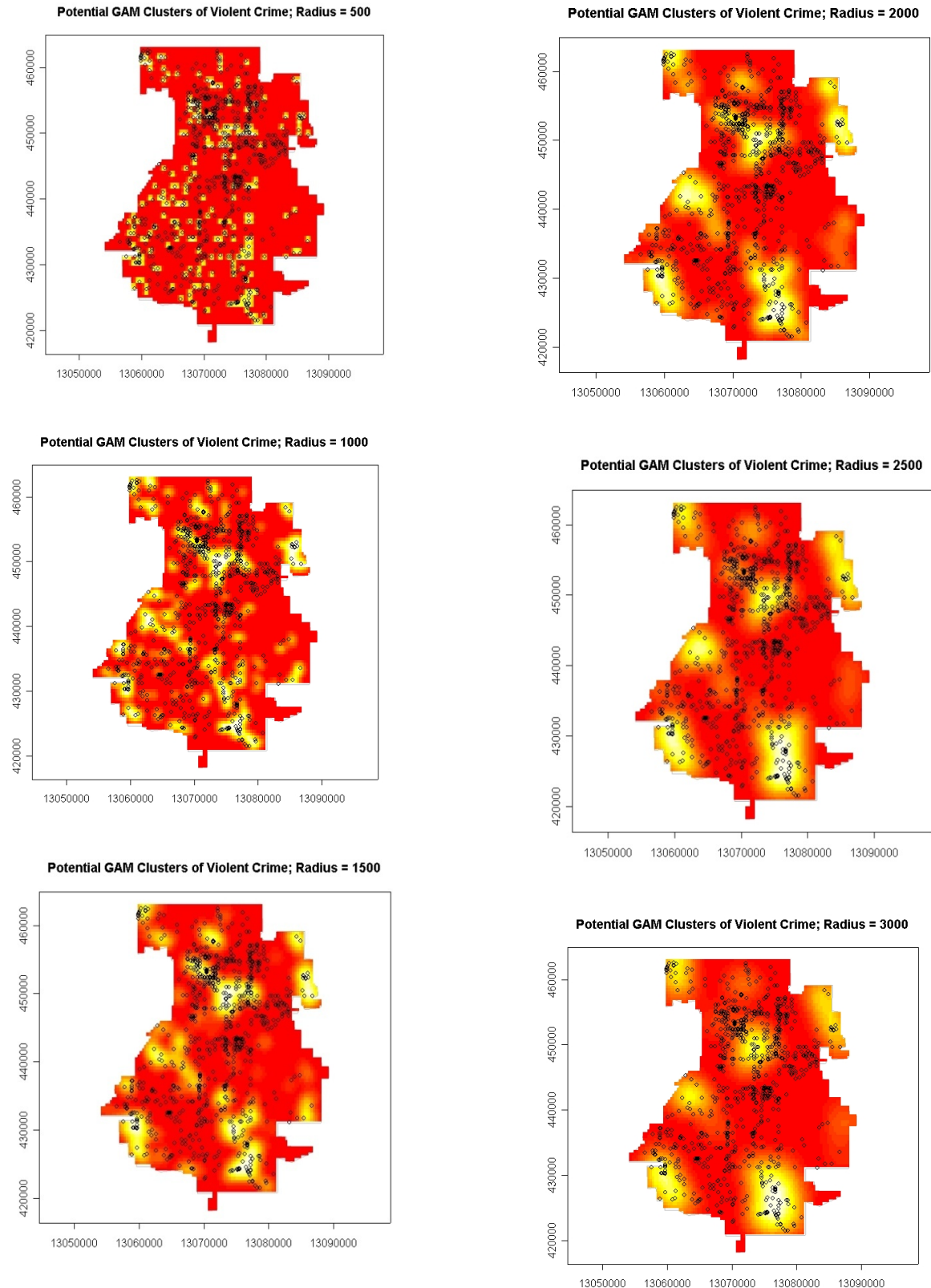


Fig. 3. Potential GAM hot spots of violent crime

The local I statistic decomposes the global I into contributions for each location, I_i . The sum of I_i for all observations is proportional to the global I [8]. Once the global I

indicates there is a spatial autocorrelation in the study region, the local values of I must be examined to identify the potential hot spots.

To perform this technique, the author first aggregated the point data of violent crime into 115 Census blockgroups. The second map in Fig. 1 shows the quantile-shaded distribution of violent crime. Then, the author calculated the global I using 3, 5, and 15 nearest neighbors respectively, giving equal weights for all nearest neighbors¹⁰. Both randomization and permutation test were performed to check if the statistics were significant or not¹¹[8][10]. Table 1 contains the result. As the number of nearest neighbors increased, the global I statistics became smaller. However, all the test statistics were significant for both randomization and permutation tests, which suggested that there is a global autocorrelation of violent crime in Lansing.

Table 1. Global I statistics using equal weights for all nearest neighbors

Nearest neighbors	Moran's I randomization	z	p	Moran's I permutation	p
3	.249	3.70	.00	.249	.01
5	.188	3.65	.00	.188	.01
15	.118	4.30	.00	.118	.01

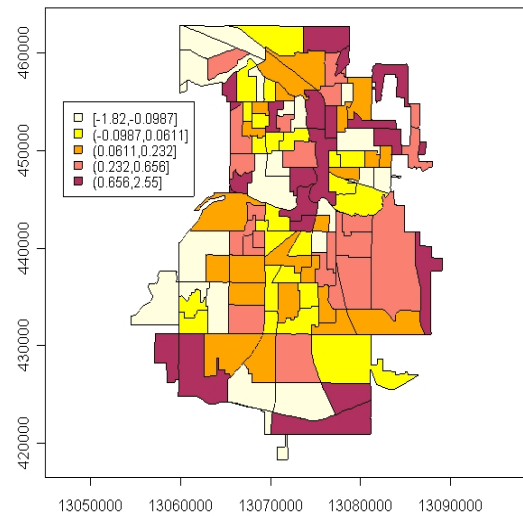
Thus, the author further examined where the local hot spots exist by analyzing the local I for the 115 Lansing blockgroups. Attached are the results for 3 and 5 nearest neighbors respectively that contain local I values, mean, variance, and their corresponding values of z-score, and probability. Quantile shaded maps are also provided in Fig. 4. The local I for 3 nearest neighbors ranged from -1.82 to 2.55 and that for 5 nearest neighbors ranged from -2.4 to 2.05. Although the two maps seemed to be somewhat different, the overall distribution looked very similar. Moreover, since there was no clear criterion to judge which one is better, further analysis was based on the proximity measure of 3 nearest neighbors. Readers must note that, like the global I statistics, the negative local I indicates spatial clustering of dissimilar values¹² and the positive local I represents a clustering of similar values of either high or low[8]. Thus, the red-shaded areas are likely to be the potential clusters.

¹⁰ Other weighting schemes such as "inverse distance weighting" could be given, but the author decided not to use them simply because there was no strict rule to follow. The same weighing scheme was used for the local I.

¹¹ "moran.test" function in spdep library was used to perform randomization test and "moran.mc" was used for permutation test. Details of the significance tests were above the scope of this paper for it focused on detecting crime hot spots.

¹² For example, a location with high values surrounded by neighbors with low values[8].

Local I for Violent Crime (Quantile Shading - 3nn)



Local I for Violent Crime (Quantile Shading - 5nn)

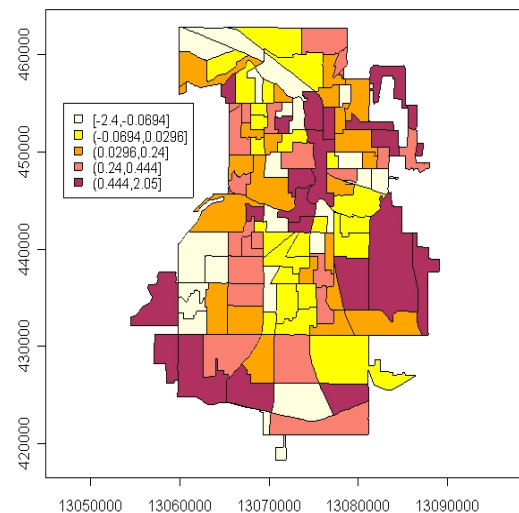


Fig. 4. Local I for violent crime (3 and 5 nearest neighbors)

Meanwhile, there were 11 blockgroups that have significantly high local I values at the .05 level¹³. The conventional significance level of .05 had to be adjusted because there was the multiple comparisons problem[8]. Employing a Bonferroni bounds procedure¹⁴ produced a rather conservative significance level of .000435, and only 4 blockgroups exceeded this level¹⁵. Fig. 5 shows two maps for both significance levels. Interestingly, only one blockgroup around the Chestnut area had significant local I at the .05 level, and no blockgroup in that area was significant at the .000435

¹³ blockgroup id = 11, 12, 46, 55, 62, 70, 77, 84, 89, 94, 111

¹⁴ Bonferroni adjustment = $\alpha / n = .05 / 115 = .000435$

¹⁵ blockgroup id = 12, 46, 62, 77

level. The author further examined 5% of the highest and lowest local I values¹⁶.

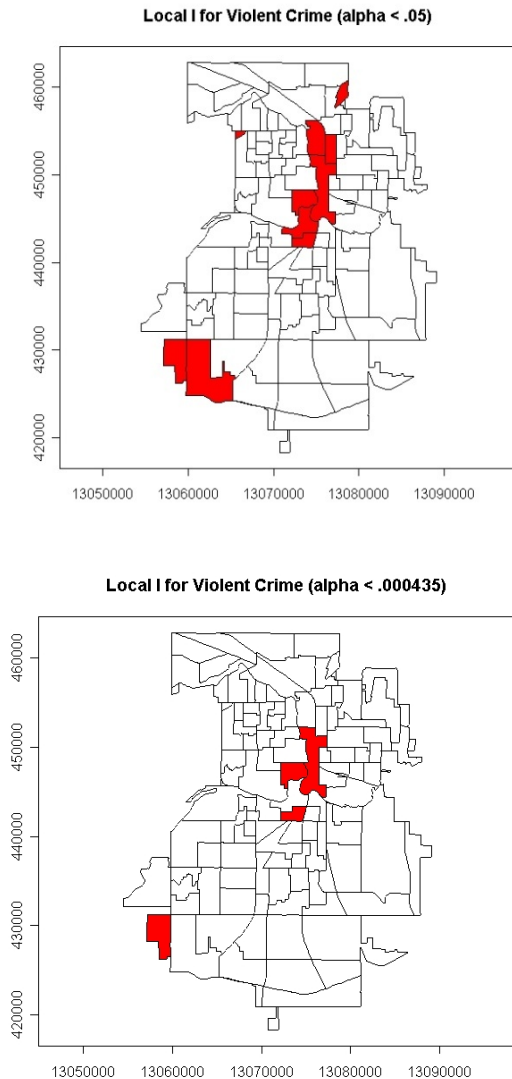


Fig. 5. Local I for violent crime (significant blockgroups)

Fig. 6 shows the result. There was not much difference from the overall pattern shown in the dot and quantile map except for the Chestnut area. Overall, the various maps of local I appeared to suggest that the State Capitol area and southwestern region are the most serious hot spots, whereas, the Chestnut area is not much problematic. However, one plausible reason of such unexpected result could be because the Chestnut area has many small blockgroups (due to high concentration of population) and some of them having relatively small number of violent crimes¹⁷ might have affected the local I that have used 3 nearest neighbors as its proximity measure;

¹⁶ highest 5% id = 11, 12, 46, 62, 77, 84 / lowest 5% id = 1, 56, 80, 81, 105, 111

¹⁷ Refer to the quantile shaded choropleth map of violent crime in Fig. 1.

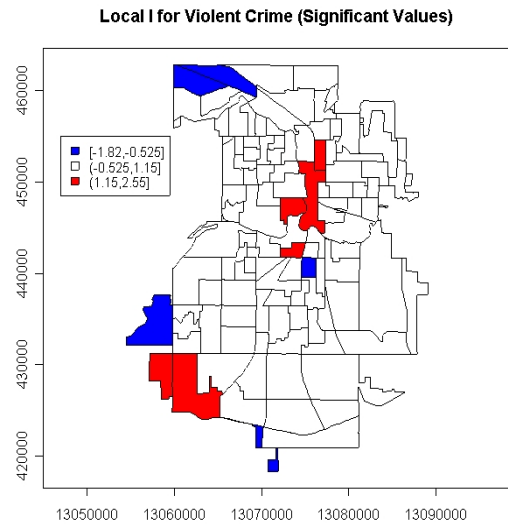


Fig. 6. Highest and lowest 5% local I

3.3 GAM v. Local Moran's I

Fig. 7 represents the two resulting maps of GAM and local Moran's I¹⁸. It is notable that the standard deviation map of local I is fairly liberal because the significance level of some of the red-colored blockgroups is less than .05. Nevertheless, readers may be able to easily detect the difference between hot spots identified by the two techniques. In particular, while the GAM consistently highlights the Chestnut area (even after controlling for physical disorder), the local I does not identify it as crime-concentrated zones. In addition, the northwestern and midwestern areas have very opposite results. Although they use different types of data and have distinct approaches, the results are still not explicable in an easy way. The author's personal knowledge on the phenomenon of violent crime recommends that GAM be applied if the point pattern data are available.

With regard to the second purpose of this research, however, the underlying differences of methodological characteristics between the two techniques need to be explored to correctly understand the different results. Table 2 summarizes the comparison. (1) Both are exploratory techniques that cannot be used for hypothesis testing or modeling to explain the observed spatial patterns. (2) While GAM uses point pattern data, local Moran's I is appropriate for aggregated areal data. Once aggregated, the point data lose much valuable information. (3) GAM controls for underlying distribution of population at risk, but local I does not. Thus, local I may simply reflect the population distribution. (4) Choosing background rate heavily influences the result of GAM. Whereas, the number of neighbors, weighting scheme, and MAUP may produce different results for local I. (5) Poisson distribution is used to find anomalous circles in GAM. Whereas, local I uses normal distribution for significance testing. (6) Although both methods suffer from multiple comparisons problem, only local I can adjust the issue by using a Bonferroni bounds procedure. (7) Poor visualizing sometimes causes misleading conclusion. The

¹⁸ GAM map is the fourth one in Fig. 3, and local I map is shaded according to standard deviation.

GAM map gets smoothed as the radius of kernel disc increases. Meanwhile, large areas tend to exaggerate the degree of spatial association in local I maps.

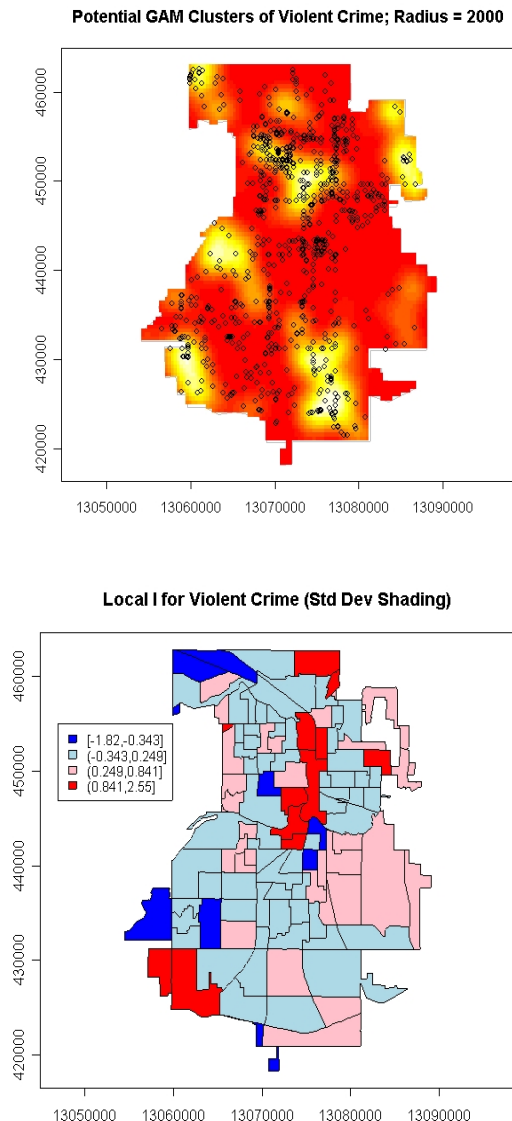


Fig. 7. GAM v. Local Moran's I

Table 2. Methodological comparison between GAM and local I

	GAM	Local Moran's I
Analysis	exploratory	exploratory
Data type	point pattern	aggregate zonal data
Number of variables	bivariate	univariate
Deciding factor	background rate	number of neighbors, weighting, MAUP
Significance testing	Poisson	normal
Multiple comparisons	yes – no adjustment	yes – Bonferroni adjustment
Visualizing	radius of kernel disc	large areas

4. CONCLUSION

Most Korean law enforcement agencies are still relying on the simple visual techniques in detecting crime hot spots. Considering the importance of identifying exact hot spot areas in efficiently distributing police resources, this study attempted to introduce the two sophisticated techniques, GAM and Local Moran's I. Utilizing data of Lansing, Michigan, the author showed how to perform them. The results showed that there is a substantial difference in the locations found by the two methods. Some areas that were identified as hot spots in GAM were indicated as the opposite in local Moran's I and vice versa. However, the difference does not necessarily indicate that the local Moran's I is a poor technique. To the contrary, it is one of the most sophisticated exploratory methods for cluster detection for zonal data. The difference is rather to be sought in the distinct methodological characteristics of those techniques. Nevertheless, this study suggests that more efforts to obtain point data including exact information on the event location would be valuable because GAM can avoid the MAUP problem intrinsic in the analysis of zonal data. That said, the law enforcement agencies need to collect more exact point data of crime and collaborate with researchers to make better analyses of crime hot spots.

REFERENCES

- [1] E. Jefferys, "A multi-method exploration of crime hot spots," *Proc. Academy of Criminal Justice Sciences*, 1998.
- [2] J. Eck, J. Gersh, and C. Taylor, "Mapping hot spots of crime and related events," *Proc. Spatial Analysis of Crime*, 1997.
- [3] C. Mamalian and N. La Vigne, *The Use of Computerized Crime Mapping by Law Enforcement: Survey Results*, National Institute of Justice, Washington, D.C., 1999.
- [4] J.Q. Wilson and G.L. Kelling, "The Police and Neighborhood Safety: Broken Windows," *Atlantic Monthly*, vol. 127, 1982, pp. 29-38.
- [5] R.D. Baller, L. Anselin, S.F. Messner, G. Deane, and D.F. Hawkins, "Structural Covariates of U.S. County Homicide Rates: Incorporating Spatial Effects," *Criminology*, vol. 39, no. 3, 2001, pp. 561-590.
- [6] R.J. Sampson, S.W. Raudenbush, and F. Earls, "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy," *Science*, vol. 277, 1997, pp. 918-924.
- [7] S. Openshaw, M. Charlton, C. Wymer, and A.W. Craft, "A Mark I Geographical Analysis Machine for the Automated Analysis of Point Data Sets," *International Journal of Geographical Information Systems*, vol. 1, 1987, pp. 335-358.
- [8] L. Anselin, "Local Indicators of Spatial Association-LISA," *Geographical Analysis*, vol. 27, no. 2, 1995, pp. 93-115.
- [9] W.G. Skogan, *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*, Free Press, New York, 1990.

- [10] T.C. Bailey and A.C. Gatrell, *Interactive Spatial Data Analysis*, Prentice Hall, Addison-Wesley, 1995.



Jinseong Cheong

He received the B.A. in public administration from Korea National Police university in 1996 and received M.S. in criminology from the Florida State university, USA in 2003. He earned Ph.D. in criminal justice from Michigan State university, USA in 2008. Since 2009, he has been working

as a professor at Soon Chun Hyang university. His main research interests include geographic profiling, crime prevention, and scientific crime investigation.