University of East London

MSc. In Data Science

SPATIAL DATA ANALYSIS

DS7002

COURSEWORK

BY

DAMILOLA OLUWASENI OMISORE

2272153

ON

SPATIAL DATA ANALYSIS IN CRIME

ABSTRACT

Crime has been a reoccurring phenomenon affecting most if not all communities and societies worldwide. To subdue this, crime analysis has been adopted to provide timely and useful information on crime patterns and trends. For this work, I will be reviewing different spatial data analysis techniques that have been used for crime analysis such as spatial regression analysis, spatial clustering, and hot spot analysis. Will also be discussing, criticizing and pointing out the advantages and bottlenecks of these techniques. At the end of the paper, I will conclude with a summary of the other spatial data analysis techniques also used for crime analysis.

Table of Contents

INTRODUCTION

In plain English, a crime is an illegal act that is harmful to not just a specific person but also a group of people, a society, or the state ("a public wrong"). These actions are against the law and are punished(*Crime - Wikipedia*, no date). The issue of crime has to be thoroughly researched and addressed on a global scale. Throughout the past several generations, crime rates have increased as a result of the psychological, economical, and behavioural disorders of individuals(Mohammed and Baiee, 2020).

Spatial data analysis is a data analysis methodology that involves examining and interpreting data with a geographical or spatial component. Its importance has grown significantly in criminology in recent years as it provides a means to examine crime patterns and investigate the interplay between crime and the constructed environment (*Spatial analysis - Wikipedia*, no date). In criminology, the utilization of Geographic Information Systems (GIS) for spatial data analysis allows the mapping of crime incidents and the identification of patterns in crime data. This technique has demonstrated its effectiveness as a valuable tool for law enforcement agencies and researchers in identifying high-crime areas, optimizing resource allocation, and devising focused crime prevention strategies(Reid, Tita and Valasik, 2011).

In the context of crime, this study offers a comprehensive review of spatial data analysis. I'll go through the approaches taken to evaluate crime data as well as the advantages and drawbacks of employing spatial analysis in criminological studies. This paper's overall goal is to present a thorough review of spatial data analysis in crime and to show how useful it is for comprehending the nuanced connection between crime and the built environment.

METHODOLOGY

I'll start by giving a general review of GIS and spatial data analysis. The forms of crime data that are frequently employed in spatial analysis, such as crime incident data, offender data, and environmental data, will next be covered. The methodologies and approaches utilised in geographic data analysis, such as spatial clustering, hot spot analysis, and spatial regression analysis, will next be discussed. Lastly, I'll go through the advantages and drawbacks of using

spatial data analysis in crime research and give some examples of how this methodology has been applied in the real world.

RESEARCH AND DEVELOPMENT

SPATIAL REGRESSION ANALYSIS

One of the techniques that have been used in previous years is spatial regression analysis, which is a statistical technique used to model the relationship between spatially correlated independent variables and a dependent variable. When examining geographical patterns in data and determining the variables that impact those patterns, this method is frequently employed in spatial econometrics, geography, ecology and in the field of criminology to investigate the geographical patterns of crime and determine the variables that impact such patterns. The technique was employed to examine how crime and socioeconomic factors relate to one another in a spatial data analysis project on crime in Vancouver, British Columbia by Martin A. Andresen. Researchers can account for the geographic autocorrelation of crime, which states that the amount of crime in one location is impacted by the level of crime in other areas, using spatial regression analysis (Anselin, 1995). In particular, he examined the spatial association between crime rates and a variety of factors, such as social disorganisation variables (e.g., poverty, residential instability) and routine activity variables (e.g., population density, land use mix), using a spatial lag model. According to the results of the spatial regression study, both social disorganisation and regular activity factors were significant predictors of crime rates in Vancouver. More specifically, higher rates of poverty, housing instability, and ethnic heterogeneity were linked to higher crime rates, whereas higher rates of population density and land use diversity were linked to lower crime rates (Andresen, 2006).

Spatial regression analysis was also used to test a theoretical model of crime hotspot generation based on environmental criminology by Brantingham and Brantingham in 2008. They analysed the correlation between crime patterns and numerous environmental factors, such as the availability of criminal chances, the presence of guardians, and the visibility of crime locations, using data from the city of Vancouver, Canada, using a spatial error model. The findings of their study were consistent with their theoretical hypothesis,

which postulates that crime hotspots were caused by the concentration of criminal opportunities and motivated criminals in certain places. They discover that crime rates in hot places were positively correlated with the number of criminal possibilities, such as unattended premises and busy streets. In contrast, crime rates in high-crime areas were adversely correlated with the presence of guardians like police officers and security cameras. Their study also emphasised the significance of spatial autocorrelation in crime patterns, which implied that the crime rates in one place were affected by the crime rates in other places. They contend that spatial regression models can assist researchers in taking into account this geographical dependency and in creating more precise models of the creation of crime hotspots (Brantingham and Brantingham, 1999).

Another use of spatial regression modelling was in 2010 by Groff E. R who used it to test the assumptions and predictions of routine activity theory concerning street robbery. In order to compare the simulation findings with actual crime data from Philadelphia, he created a simulation model based on routine activity theory and performed spatial regression analysis. Utilizing a spatial lag model to look at the spatial relationships between crime patterns and various environmental factors, such as the density of potential targets, the concentration of motivated offenders, and the accessibility of crime locations, is how spatial regression analysis was put into practice. Measures of geographical dependency and temporal dynamics were also included by Groff in his study. The findings of his study demonstrate that routine activity theory was a good framework for comprehending street robbery patterns and that the simulated results were congruent with the actual crime statistics. He discovered that while the concentration of motivated criminals had a smaller impact, the concentration of possible targets and the accessibility of crime areas were important determinants of street robbery rates. On the basis of their features and location, he also pinpointed particular Philadelphia neighbourhoods that were more susceptible to street robberies (Groff, 2007).

HOT SPOT ANALYSIS

Hotspot analysis is a spatial analysis and mapping technique interested in the identification of clustering of spatial occurrences. The locations of events or things are indicated by these spatial phenomena, which are shown as points on a map. In geographic data analysis, the hot spot analysis is frequently used to pinpoint locations with statistically significant high or

low rates of crime. A case study is the use of hot spot analysis to identify high-risk areas for gun violence in Boston by Braga et al. To locate geographical clusters of gun violence occurrences, the authors employed Getis-Ord Gi* statistics and kernel density estimation (Braga, Papachristos and Hureau, 2010). Kernel density estimation (KDE) is a method for estimating the probability density function of a random variable based on a set of geographical data points. It is frequently used to show how data is distributed spatially and to pinpoint locations with a lot or little activity (Silverman, no date). Getis-Ord Gi is a statistical method used in the study of geographical data to locate geographic hotspots with high or low values for a given variable of interest. It is a measurement of the degree to which the values of a variable are linked with their geographical neighbours, or spatial autocorrelation (Chainey Spencer and Ratcliffe Jerry, no date). The stability of these hot areas over time and potential contributing reasons to their persistence were then examined using spatial regression analysis. The analysis's findings demonstrated that Boston's gun violence was largely concentrated in a select few micro-locations and that these hotspots remained constant over time. The investigators also discovered that gang involvement, drug markets, and social poverty were important predictors of hot spot stability (Braga, Papachristos and Hureau, 2010).

Another case study is the analysis by Weisburd and Green in 1995. They examined using hot spot analysis, the effectiveness of a targeted policing intervention on drug-related crime in Jersey City. They made use of statistics on drug-related crimes that occurred over a six-month period in a particular Jersey City neighbourhood. Using kernel density estimation and closest neighbour analysis, they located drug activity hotspots. After that, the hot areas were allocated at random to either a treatment group where targeted policing was used, or a control group, where no intervention was used. The study's results demonstrated that the hotspot-specific targeted police strategy was successful in lowering drug-related crime. Comparing the treatment group to the control group, there was a statistically significant drop in drug-related criminal events. The study also discovered that hot regions discovered by hot spot analysis made better targets for the intervention than places picked at random (Weisburd and Green, 1995).

SPATIAL CLUSTERING

Spatial clustering is a technique used to find collections of observations that are geographically adjacent to one another and share characteristics. Spatial clustering may be used in crime analysis to locate crime hot spots, or regions with a high density of criminal activity (Pattnaik Anubhav, no date). Spatial clustering was used by J.H. Ratcliffe to analyse patterns of crime incidents in Philadelphia. The study's objective was to determine the minimal hit rate required for a geocoding system to be helpful in crime investigation. The spatial clustering method used in the study was the kernel density estimation (KDE) technique. Using a kernel function, the criminal incidence data was smoothed across a continuous surface in this manner. The resulting surface highlighted regions of high crime concentration and offers a visual depiction of crime density across the research area. The result was a precise identification of the research area's crime hotspots. These hotspots were discovered to be clustered in particular geographic regions, such as particular communities and commercial districts. The study also discovered that depending on the kind of crime being studied and the desired level of precision, different minimum hit rates were required for a geocoding system to be beneficial in crime analysis (Ratcliffe, 2004).

CRITICAL REVIEW AND DISCUSSION

On the spatial regression analysis technique used for the crime analysis in Vancouver by Martin A. Andresen, the use of a rigorous statistical methodology to pinpoint geographic trends in crime and analyse the connections between crime and numerous predictor factors is one of its strong points. The study makes use of a spatial lag regression model, which enables the examination of spatial autocorrelation and offers a more precise assessment of the influence of predictor variables on crime rates. However, the dependence on secondary data sources for the variables under study, such as census information and crime statistics, is one drawback. This might induce biases or measurement errors in the analysis. The study also doesn't take into account additional factors that can affect crime rates, such as police presence or neighbourhood demography. Another drawback is the model's potential for endogeneity is another drawback. Because the study only analyses cross-sectional data, estimations of the predictor factors' impact on crime rates are likely to be skewed due to the likelihood that crime rates might affect those variables.

For the hot spot analysis by Braga et al, the study's use of a hot spot analysis tool, notably the Getis-Ord Gi* statistic, a commonly used and reliable technique for locating hot spots in geographical data, is one of its strong points. The authors' use of this method enables them to pinpoint high-risk gun violence hotspots and evaluate the geographical clustering of these episodes. However, there are some drawbacks. The analysis's arbitrary choice of the distance band parameter is one possible problem. Although there isn't any clear advice on how to choose this parameter, different distance bands could produce distinct hot spot patterns. The authors picked a distance band of 150 metres based on prior research. Another drawback is that the hot spot study did not illuminate the reasons why some places have a higher propensity for gun violence. Despite identifying high-risk regions, the approach did not shed light on the underlying factors or mechanisms causing the geographic clustering of gun violence and the technique is also limited by the quality and completeness of the data.

Lastly, the spatial clustering used by Jerry H. Ratcliffe has one strength. The employment of a spatial clustering methodology in the study specifically, the K-function, which is a commonly used and reliable way to detect spatial grouping in point data is one of its strong points. The author can evaluate the geocoding's correctness and the distribution of criminal occurrences thanks to the K-function. However, there are some limitations to the study's use of the spatial clustering technique. The K-function analysis's selection of the distance band might be a problem. Based on prior study, the author used a distance band of 500 metres; however, alternative distance bands may produce different spatial clustering patterns, which can have an impact on the results.

CONCLUSION

The study reviews the spatial data analysis techniques used in crime highlighting the strengths and limitations. The study shows how effective spatial regression analysis approaches are in revealing the nuanced relationships between crime and environmental and social elements in urban settings. Also, the study emphasises how hot spot analysis methods may be used to locate places with high levels of crime, comprehend their temporal stability and provide insightful information about the geography of crime. According to the

research, spatial clustering of crime incidents can provide important information about crime trends and the efficacy of policing tactics. Finally, I believe the use of spatial analysis techniques should be carried out in comparison that is implementing multiple techniques that person similar tasks and comparing the outcomes. This will help to identify possible loopholes in the analysis.

REFERENCES

Andresen, M.A. (2006) 'A spatial analysis of crime in Vancouver, British Columbia: A synthesis of social disorganization and routine activity theory', *Canadian Geographer*, 50(4), pp. 487–502. Available at: https://doi.org/10.1111/j.1541-0064.2006.00159.x.

Anselin, L. (1995) 'Local Indicators of Spatial Association—LISA', *Geographical Analysis*, 27(2), pp. 93–115. Available at: https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

Braga, A.A., Papachristos, A. V. and Hureau, D.M. (2010) 'The Concentration and Stability of Gun Violence at Micro Places in Boston, 1980–2008', *Journal of Quantitative Criminology*, 26(1), pp. 33–53. Available at: https://doi.org/10.1007/s10940-009-9082-x.

Brantingham, P.L. and Brantingham, P.J. (1999) 'A theoretical model of crime hot spot generation.', *Studies on Crime & Crime Prevention*, 8, pp. 7–26.

Chainey Spencer and Ratcliffe Jerry (no date) 'GIS_and_Crime_Mapping'.

Crime - Wikipedia (no date). Available at: https://en.wikipedia.org/wiki/Crime#cite_note-oed-1 (Accessed: 5 April 2023).

Groff, E.R. (2007) 'Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery', *Journal of Quantitative Criminology*, 23(2), pp. 75–103. Available at: https://doi.org/10.1007/s10940-006-9021-z.

Mohammed, A.F. and Baiee, W.R. (2020) 'Analysis of Criminal Spatial Events in GIS for predicting hotspots', in *IOP Conference Series: Materials Science and Engineering*. IOP Publishing Ltd. Available at: https://doi.org/10.1088/1757-899X/928/3/032071.

Pattnaik Anubhav (no date) *Geospatial Clustering: Kinds and Uses | by Anubhav Pattnaik | Towards Data Science*. Available at: https://towardsdatascience.com/geospatial-clustering-kinds-and-uses-9aef7601f386 (Accessed: 9 April 2023).

Ratcliffe, J.H. (2004) 'Geocoding crime and a first estimate of a minimum acceptable hit rate', *International Journal of Geographical Information Science*, 18(1), pp. 61–72. Available at: https://doi.org/10.1080/13658810310001596076.

Reid, S.E., Tita, G. and Valasik, M. (2011) 'The Mapping and Spatial Analysis of Crime', in *Criminology*. Oxford University Press. Available at: https://doi.org/10.1093/obo/9780195396607-0123.

Silverman, B.W. (no date) 'Density_Estimation_for_Statistics_and_Da'.

Spatial analysis - Wikipedia (no date). Available at: https://en.wikipedia.org/wiki/Spatial_analysis (Accessed: 5 April 2023).

Weisburd, D. and Green, L. (1995) 'Policing drug hot spots: The Jersey City drug market analysis experiment', *Justice Quarterly*, 12(4), pp. 711–735. Available at: https://doi.org/10.1080/07418829500096261.