

INVESTIGATING TEMPORAL AND SPATIAL EFFECTS OF URBAN PLANNING VARIABLES ON CRIME RATE: A GWR AND OLS BASED APPROACH

Z. Irandegani ^{1,*}, R. Mohammadi ¹, M. Taleai ¹

¹ Faculty of Geomatics, K. N. Toosi University of Technology, Tehran, Iran. (z.irandegani@gmail.com;
r.mohammadi0645@email.kntu.ac.ir; taleai@kntu.ac.ir)

Commission VI, WG VI/4

KEY WORDS: Crime mapping, Spatial-temporal Analysis, GWR, OLS, DBSCAN, Land Use Diversity

ABSTRACT:

Spatial, temporal, environmental and urban planning variables are some factors effected crime occurrence as a social undesirable phenomenon. In this paper, the effect of urban planning variables (including land use diversity and police station area) and temporal parameters (including daily and weekly time windows) on crime incidence were investigated by doing some spatial-temporal analysis. To tackle this, at first, in order to determine crime clusters, DBSCAN algorithm is utilized. Then by using the Ordinary Least Square (OLS) and Geographically Weighted Regression (GWR) the global effect of urban planning variables on crime spatial clusters are modelled and compared. Finally, the OLS method is used to investigate the effects of temporal variables on crime spatial clusters. The results demonstrate that there is a negative correlation between crime clusters and the police station coverage, and a positive relationship between crime clusters and land-use diversity. Furthermore, increasing the number of police stations in districts with high land-use diversity can result in reducing the crime rate. Also, the result of our temporal analysis shows that the high crime rate was happened at noon and on weekdays. Therefore, increasing the number of police officers at noon and on weekdays in the districts with high land-use diversity is recommended.

1. INTRODUCTION

Crimes, do not happen randomly (Nezami and Khoramshahi, 2016), and in order to manage the crime rate, determining the effective factors and assessing their impact on crime distribution is critical. The rate of crime incidents depends on some spatial, temporal and urban planning variables (Lee et al., 2009). Crime rate changes daily and weekly (time scale) and also varies from one location to another (spatial scale). Moreover, some urban planning variables including land-use diversity and distance to police stations, have a significant impact on changing the crime rate.

Previous studies have been conducted to investigate the effect of spatial and urban planning variables on crime distribution. Some researchers investigate the impact of spatial indicators on crime distribution. Cahill and Mulligan (2007) examined a mixed model based on spatial parameters using GWR and OLS to explore local crime patterns.

Wheeler and Waller (2009) compared a Bayesian model with a GWR model for the estimation of the spatially varying effects of alcohol outlets and illegal drug activity on violent crime in Houston, Texas. Yan et al. (2010) used the GWR to build a local model to analyze the effects of population density, road network density, and distance from police stations on the crime rate at the census tract level and analyze the relationships between the spatial distributions of crime and geographical factors to find the cause of crime spatial distributions. Analytical results demonstrate that the spatial relationship between crime and geographical factors is a spatially non-stationary process and the GWR model can help to improve the accuracy of parameter estimation. Ferreira et al. (2012) combined cluster analysis and GWR to identify areas with the highest crime rates. Catlett et al. (2018) defined an approach

based on spatial analysis and auto-regressive models for automatically detect high-risk crime regions in urban areas. Walker et al. (2014) studied the effect that natural disasters have on the space-time behaviour of crime patterns using local-level data and highlighted aspects of change in patterns of crime as a response to Hurricane Wilma in Miami, Florida, 2005. The results revealed that more accurate predictions of crime for specific crime types in specific cities with the use of GWR are possible. Stein et al. (2016) used spatial analysis techniques to examine how the effect of disorder and efficacy upon crime patterns itself varies across a neighborhood. The indices of disorder and efficacy are compared to crime data from police call logs using geographically weighted regression. Xu et al. (2017) used GWR and mapping techniques to model the correlations between Airbnb clusters and crime indices. The result revealed significant positive correlations by crime type (i.e. Robbery and motor vehicle theft), negative correlations were recognized in violent crime (i.e. murder/rape), and also, spatially varying relationships were found across the studied area. Boivin (2018) presented a simultaneous test of both propositions of routine activity theory (suggesting that an increased human presence in a given area is expected to be associated with both an increase and a decrease in criminal activity). GWR model is used to analyze data from the city of Toronto, at the census-tract level. Results suggest that, there is a positive relationship between crime and population in many tracts; however, empirical support is also found for the opposite proposition that larger populations are (sometimes) associated with less crime. This is especially true for areas that receive visits largely for shopping, school, and work.

De Maria André and Carvalho, (2019) developed a structural choice model on the risk of robbery at a large city in Brazil to estimate it by both maximum likelihood and geographically weighted regression using a dataset from the city of Fortaleza

* Corresponding author

that contains a precise measurement of individual's subjective probability of victimization among socioeconomic measurement, willingness to pay, and police presence variables. The resulted model, estimated by geographically weighted regression, suggests that there is a reasonable amount of spatial heterogeneity that follows the city's socioeconomic spatial distribution profile. Wang (2019) used a geographically weighted regression model to capture explicit local relationships between community opportunity and the physical setting of land uses, transportation infrastructure, and public facilities. Wang et al. (2019) used GWR, OLS and Local Moran's I for determining the spatial relation between crime rate and demographic characteristics.

Few studies have examined the impact of urban planning and spatial factors on crime distribution. Lee et al. (2009) applied a mixed GWR model in order to identify the effects of spatial and land use as urban planning factor on crime incidence. Nezami and Khoramshahi (2016) used GWR to develop a model to determine crime distribution based on some environmental impact and urban planning variables such as population, illiteracy, and distance to the police station. Wang and Chen (2017) used a geographically weighted regression model to better understand the spatially varied relationships between socioeconomic outcomes and policy investments. An ordinary regression and geographically weighted regression model were utilized to estimate and account the global and local effects of land uses and capital investments respectively, while controlling socio-demographical characteristics. The geographically weighted regression results revealed the existence of spatial mismatch that socially disadvantaged groups (e.g. black population, other minorities, single-parent families, and zero-vehicle households) tend to reside in vulnerable communities.

The limitation of previous studies is that the influence of urban planning variables and the temporal scale have not been investigated. To address the aforementioned limitations, this study explores a spatial model to investigate the impact of temporal (daily and weekly) and urban planning variables (land use diversity and police station coverage) on crime cluster distribution.

The structure of the article is as follows: In section 2, we explain the presented methodology including pre-processing, crime clustering and spatial modeling crime clusters. Section 3 introduces data sources as well as the experimental results. Section 4 concludes the article with an eye towards future works.

2. METHODOLOGY

In order to develop the spatial model of crime cluster and the relations among the location of crime occurrence, the impact of temporal variables in the form of daily and weekly distribution of crime, and land use diversity and police station coverage as urban planning criteria on crime distribution are investigated.

Figure 1 depicts the workflow of the presented spatial modeling of crime incidents distribution. The presented approach follows the process of three main steps: Pre-processing, Determining Crime Spatial Clusters and Spatial Modeling Crime Clusters. Basically, crime incidents data are pre-processed to eliminate noise and outliers of the data. Then, the data are preparing for the DBSCAN algorithm and the GWR model. The pre-processed crime data are used to determining the crime spatial clusters by the DBSCAN algorithm. Finally, OLS and GWR are used to investigating the relationship between extracted crime clusters and independent variables (land use diversity, police station coverage and daily/weekly temporal scales). The following sub-sections explain the proposed approach for

investigating effective parameters and determining spatial modeling for crime distribution.

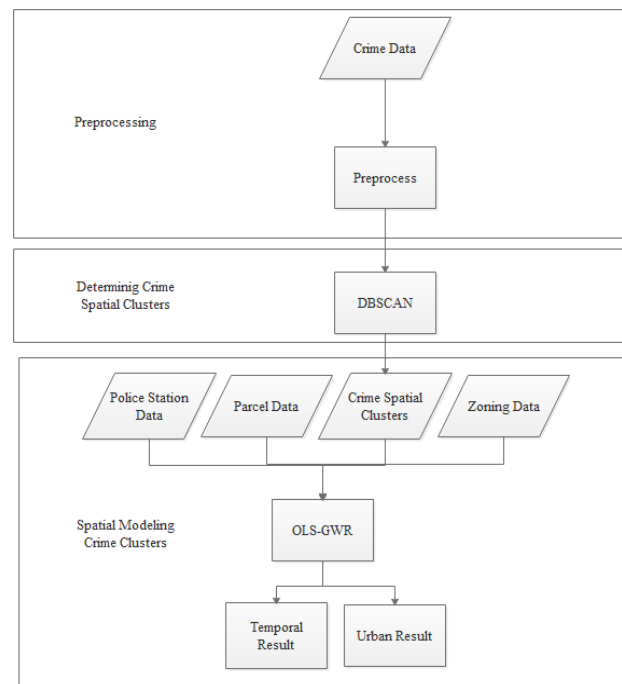


Figure 1. The workflow of spatial modeling for crime distribution

2.1. Pre-processing

The motivation behind the pre-processing step is to prepare data for the application. At first, missing values, noise and outliers are eliminated. Afterward, the local Moran's I is applied to proving the clustering distribution of crime data. The result of local Moran's I was Z-Score = 2.86627 that which demonstrates a high spatial correlation in crime incidents data and the data is not distributed randomly.

2.2. Determining crime spatial clusters

In the second step, to determine crime spatial clusters, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was applied to the pre-processed crime data. DBSCAN is a clustering algorithm that is designed to extract the clusters of a spatial database with arbitrary shape and noise (Ester et al., 1996). DBSCAN has two main variables namely minPts and ϵ , which must be specified by the user. The value of ϵ represents spatial threshold and given by the problem to solve (e.g. a physical distance), and minPts is the desired minimum cluster size (Ester et al., 1996).

2.3. Spatial modelling crime clusters

In the last step, to determine the coefficient values of effective factors that have an influence on crime spatial distribution, the OLS and GWR models are used.

Ordinary Least Square regression is applied to investigate the global relations between dependent and independent variables using equation (1) (Cramer et al., 2012).

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (1)$$

Where Y represents the dependent variable namely crime spatial clusters, X_i is an independent variable, and Alpha indicates the coefficient values of independent variables.

Simple linear regression is a frequently used model in the geographical analysis that modeled a dependent and independent variable as a linear function (Eq.2) (Brunsdon et al. 1996).

$$y_i = \beta_0 + \sum \beta_k x_{ik} + \varepsilon_i \quad k=1, m \quad (2)$$

Where y_i is the i^{th} observation of the dependent variable, x_{ik} is the i^{th} observation of the k^{th} independent variable, the ε_i are independent normally distributed error terms with zero means, and each β_k must be determined from a sample of n observations (Devkota ML et al. 2014).

Geographically weighted regression is a relatively simple technique that extends the traditional regression framework of equation (2) by allowing local variation in rates of change so that the coefficients in the model rather than being global estimates are specific to a location i . the GWR is then (Brunsdon et al. 1996):

$$y_i = \beta_{i0} + \sum \beta_{ik} x_{ik} + \varepsilon_i \quad k=1, m \quad (3)$$

GWR is an extension of OLS regression in which the parameters are allowed to vary spatially (Devkota ML et al. 2014). GWR analysis is a local regression model that can model, examine and explore spatial relations to better understand the factors behind spatial patterns (Cramer et al., 2012). In other words, GWR is mainly intended to prove the influence of independent variables (land use diversity and police station coverage) on crime spatial clusters as dependent variable (Y) using equation (3) (Fotheringham et al., 2003). Table 1 depicts the effective dependent and independent variables in crime clusters distribution. The independent variables are categorized into two group namely urban planning variables such as land-use diversity and police station coverage, and temporal variables including daily and weekly time period.

Variable description	Description	
Dependent variable		
Crime spatial cluster	Number of the cluster in each zone	
Independent variable		
Urban planning variable	Land-use diversity	Between 0 and 1
	Police station coverage	
Temporal variable	Daily	Number of clusters in each zone and each part of day and week
	Weekly	

Table 1. Effective variables in crime clusters distribution

The model for urban planning variables was set as:

$$\text{Crime spatial Cluster} = \beta_0 + \beta_1 (\text{land-use diversity}) + \beta_2 (\text{police station coverage})$$

3. URBAN PLANNING VARIABLE

In this sub-section, the aforementioned independent variable including land-use diversity index and police station coverage are explained.

3.1. Land use diversity index

To calculating land-use diversity, at the first the dominant land use is calculated based on the frequency and the area of each land use at the district level. Afterward, to normalize the land use diversity value, the area of dominant land use is divided by the total area of districts using equation (3). Afterward, the

land-use diversity index is calculated in the district level using equation (4).

$$L = 1 - [(the\ area\ of\ dominant\ land\ -use) / (total\ area)] \quad (3)$$

$$\text{Land use diversity index} = L * (NL / TNL) \quad (4)$$

Where NL is the number of land uses in each district, TNL represents the number of land use classes which is equal to 8. Figure 2 shows the calculated land use diversity for each district. The highest values of land use diversity are related to Boston Proper (B1, B2, B3, and B4), South Boston Neighbourhood (C), Mission Hill Neighbourhood (M), Huntington Avenue/Prudential Centre (H), respectively. Also, the lowest value of land use diversity is related to Harborpark: Fort Point Waterfront (HP), Allston/Brighton Neighborhood (A), Midtown Cultural (MC), Government Center/Markets (G), and City Square Neighborhood (CS).

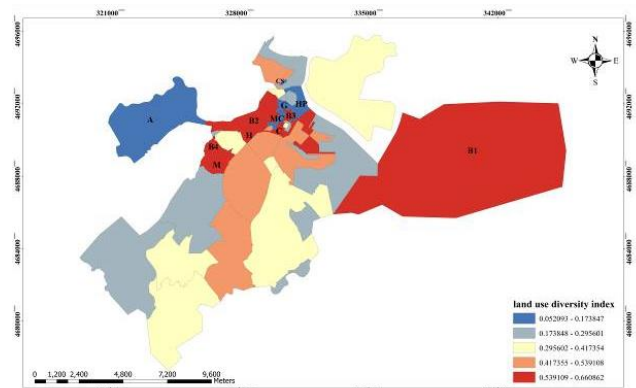


Figure 2. Land use diversity

3.2. Police station coverage

Figure 3 shows the service area of the Boston police station. The standard service area for police station was obtained by a buffer of 1500 meters based on (Sajadian et al., 2015). In order to normalize police station coverage values, the area covered by the police station was divided by the total area of each zone.

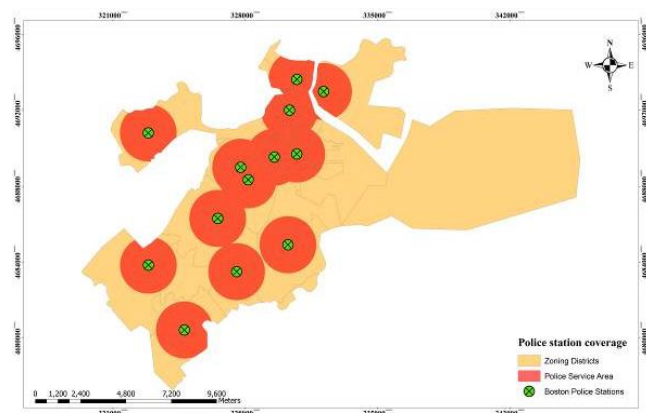


Figure 3. Police station coverage

3.3. Temporal variables

In order to determine the influence of time on crime clusters, the temporal analysis was investigated in two daily and weekly scale. As far as a daily scale is concerned, crime clusters are classified into four groups: Morning (6 am-12 pm), Noon (12

pm-6 pm), Night (6 pm-12 am), and Midnight (12 am-6 am). Furthermore, the crime cluster was categorized in weekday (i.e. Monday, Tuesday, Wednesday, Thursday and Friday) or weekend (Saturday, Sunday) to assess the impact of holiday on crime occurrence.

4. IMPLEMENTATION AND RESULT

4.1. Data set

The dataset used in this study is including crime data from 6/15/2015 to 12/30/2017, land use, districts (36 zones) and police stations of Boston, USA. The dataset obtained from Boston Maps Open Data¹. Figure 4 depicts the spatial distribution of crime incidents, police stations and land use in the district level and Figure 5 depicts the temporal distribution of crime incidents.

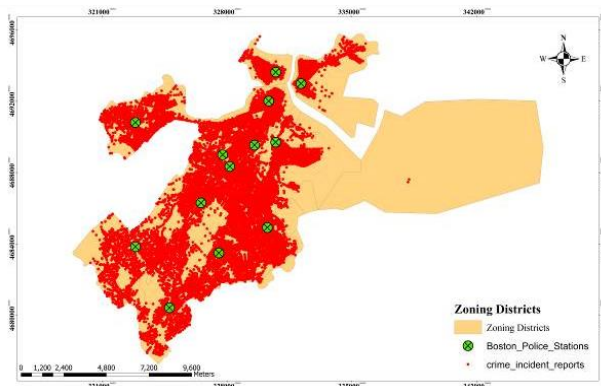


Figure 4. The location of the police station and crime data

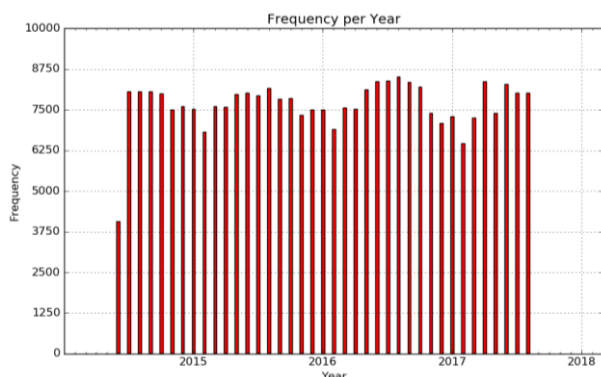


Figure 5. Crime temporal distribution (monthly), Boston, 06-2015 to 08-2017

5. RESULTS

5.1. Crime clusters result

Figure 6 shows crime clusters according to the setting of 20m as the spatial threshold (ϵ) and 10 as minimum crime cluster size (minPts). Spatial threshold and minimum crime cluster size chosen based on experiment and are adjustable.

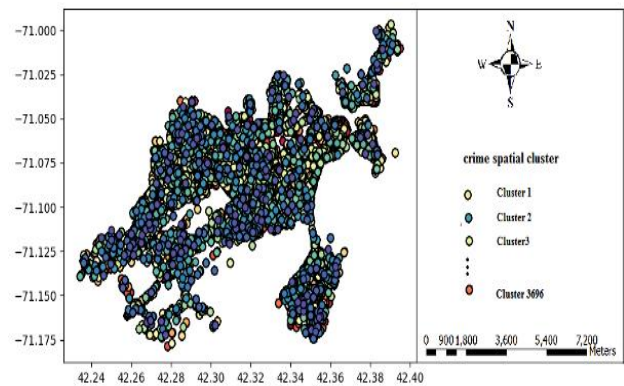


Figure 6. The result of crimes clusters by DBSCAN algorithm

5.2. Urban planning variables result

In order to determine the coefficient values of land-use diversity and police station coverage on crime distribution, the OLS model was applied on crime clusters as the dependent variable, and land use diversity and police station coverage as independent variables. The result of the OLS model were $\beta_1 = 0.384$ and $\beta_2 = -0.184$ for coefficients of land-use diversity and police station coverage, respectively (Table 2).

Independent Variables	Estimated Value	Standard Error	VIF
Intercept	0.174	0.119	
land-use diversity	0.384	0.242	1.002
police station coverage	-0.184	0.115	1.002
Adjusted R ²	0.085		
AIC	-3.563		

Table 2. Ordinary Least Squares (OLS) results

The result of the OLS model indicates that there is a negative relation between crime clusters and police station coverage. It also proves that there is a positive relationship between land use diversity and crime spatial clusters. To evaluating the confidence of spatial autocorrelation in residuals, Local Moran's I function is applied. The result of Moran's I function for urban planning variables was Moran's Index = 0.137571 (ZScore = 1.092, PValue = 0.275) that represents there is no autocorrelation between residuals.

Figure 7 indicates the result of the GWR model based on crime data, land use diversity and police station coverage in the district level. The results are in accordance with the standard deviation of residual values resulted from the difference between the model and actual values. The large and positive values of the standard deviation of residuals ($1.5 < \text{stdResid} < 2.5$) represent the high rate of crime spatial clusters (hot-spot) and the negative values ($\text{stdResid} > -2.5$) show the distinct with the lowest crime rate (cold-spot). Therefore, the most crime occurrence belongs to South Boston Neighbourhood (S), Newmarket Industrial Commercial Neighbourhood District (NI), North End Neighbourhood (NE), Bulfinch Triangle (B) & Government Centre (G) districts, respectively. Also, the lowest crime rate has happened in Charlestown Neighbourhood (C) district.

¹ <http://bostonopendata-boston.opendata.arcgis.com/>

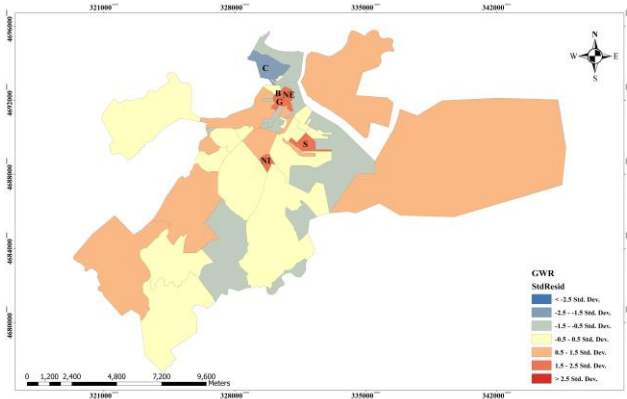


Figure 7. GWR results for crime clusters

Figure 8 demonstrates that the highest rate of crime clusters has happened in districts with low police station coverage.

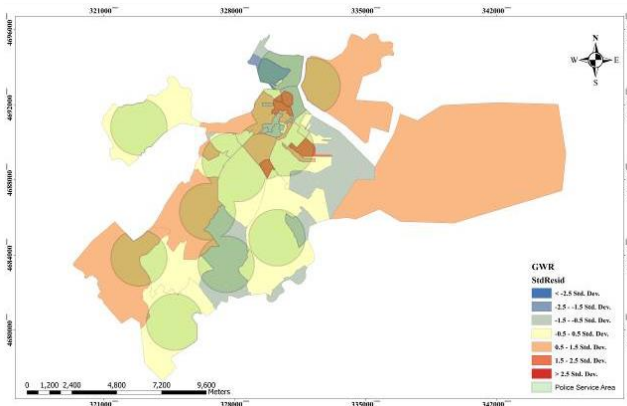


Figure 8. The compared result of GWR and Police station coverage

According to the results, we can conclude that land use diversity has a positive relation with crime spatial clusters in the study area. However, the relation of police station coverage variable with crime was a negative relationship and the crime will be reduced in the region covered by police stations.

5.3. Temporal variables result

Table 3 shows the results of the crime distribution based on a daily scale. According to the results, the coefficients of daily parameters are close to each other. In order to clarify the effect of daily parameters on crime clusters, the OLS method applied in the Morning, Noon, Night, and Midnight temporal scales. The results show that the most crime clusters occurred at Noon, Night, Morning and Midnight, respectively.

Parameter	α_{morning}	α_{noon}	α_{night}	α_{midnight}
Confidence value	4.321494	2.898603	3.333472	7.789458

Table 3: The results of the crime distribution model based on daily parameters

Due to autocorrelation between weekdays and weekend, the OLS applied in Weekday and Weekend. The results of the OLS

method were 0.968 and 1.082 for Weekday and Weekend, respectively. The results prove that most crime clusters happened on Weekdays.

6. CONCLUSION

In this paper, the effect of land-use diversity and police station coverage as some urban planning variables and daily/weekly time windows as temporal parameters on crime incidence were investigated. For this purpose, Boston crime occurrence data are clustered by utilizing DBSCAN clustering algorithm and then the clustered data are modelled by OLS and GWR. By modelling the crime distribution with the GWR and OLS model, it is possible to predict the crime distribution in an area through the change in urban planning and temporal variables. For example, the position of police stations has a significant impact on crime reduction. Generally, the achieved model may be efficient in taking suitable decisions toward reducing crimes by the positioning of new police stations, increasing the number of police in times with overpopulation, or policy-making based on the land-use diversity. Also, some variables such as demographics (i.e., age, income, and gender), which are relevant to the crime rate and not included in this study should be included in future studies.

REFERENCES

- Boivin, R., 2018. Routine activity, population (s) and crime: Spatial heterogeneity and conflicting propositions about the neighborhood crime-population link. *Applied Geography*, 95, pp.79-87.
- Brunsdon, C., Fotheringham, A.S. and Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), pp.281-298.
- Cahill, M. & Mulligan, G. 2007. Using Geographically Weighted Regression to Explore Local Crime Patterns. *Social Science Computer Review*, 25(2): 174-193. Doi.Org/10.1177/0894439307298925
- Catlett, C., Cesario, E., Talia, D. & Vinci, A. 2018. A Data-Driven Approach for Spatio-Temporal Crime Predictions in Smart Cities. *IEEE International Conference On Smart Computing (Smartcomp)*, 1(2018): 17-24. 10.1109/SMARTCOMP.2018.00069
- Cramer, D., Brown, A.A. & Hu, G. 2011. Predicting 911 Calls Using Spatial Analysis. *Software Engineering Research, Management and Applications, SCI 377:15-26*. Springer-Verlag Berlin Heidelberg.
- de Maria André, D. and Carvalho, J.R., 2019. Assessing the Benefits of Robbery Reduction: The Case of a Large Brazilian City. *Economi A* 20(1), pp.44-72. DOI:10.1016/j.econ.2019.02.001
- Devkota, M.L., Hatfield, G. and Chintala, R., 2014. Effect of sample size on the performance of ordinary least squares and geographically weighted regression. *Br. J. Math. Comput. Sci*, 4, pp.1-21.
- Ester, M., Kriegel, H.P., Sander, J. & Xu, X. 1996. A Density-Based Algorithm for Discovering Clusters In Large Spatial

Databases With Noise. Institute For Computer Science, University Of Munich, 226-231.

Ferreira, J., João, P. & Martins, J. 2012. GIS for Crime Analysis: Geography For Predictive Models. *Electronic Journal Of Information Systems Evaluation*, 15: 36-49.

Fotheringham, A.S., Brunson, C. & Charlton, M. 2003. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons.

Lee, S., Kang, D. & Kim, M. 2009. Determinants Of Crime Incidence In Korea: A Mixed GWR Approach. World Conference of the Spatial Econometrics Association, Barcelona, Spain, July 8-10.

Nezami, S. & Khoramshahi, E. 2016. Spatial Modeling Of Crime by Using of GWR Method. *Baltic Geodetic Congress (Bgc Geomatics)*, Poland. 222-227. DOI:10.1109/BGC.Geomatics.2016.47.

Stein, R.E., Conley, J.F. and Davis, C., 2016. The differential impact of physical disorder and collective efficacy: A geographically weighted regression on violent crime. *GeoJournal*, 81(3), pp.351-365.

Sajadian, N., Alipour, K., Keshtgar, L. & Moridi, V. 2015. Review and Analyze the Spatial Distribution of Switching the Location of Police Stations in Ahvaz Using Fuzzy Hierarchical Analysis (FAHP). *Journal of Police Geography*, 3 (9): 1-24.

Walker, W.C., Sim, S. and Keys-Mathews, L., 2014. Use of geographically weighted regression on ecology of crime, response to hurricane in Miami, Florida. In *Forensic GIS* (pp. 245-262). Springer, Dordrecht.

Wang, C.H., 2019. A land-use and capital-investment allocation optimization model to develop a fair community opportunity framework for Columbus, Ohio. *Computers, Environment and Urban Systems*, 74, pp.151-160.

Wang, C.H. and Chen, N., 2017. A geographically weighte regression approach to investigating the spatially varied built-environment effects on community opportunity. *Journal of transport geography*, 62, pp.136-147.

Wang, L., Lee, G. & Williams, I. 2019. The Spatial and Social Patterning of Property and Violent Crime in Toronto Neighbourhoods: A Spatial-Quantitative Approach. *ISPRS International Journal Of Geo-Information*. 8(1): 51. DOI.Org/10.3390/Ijgi8010051

Wheeler, D.C. and Waller, L.A., 2009. Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems*, 11(1), pp.1-22.

Xu, Y.H. and Pennington-Gray, L., 2017. Explore the spatial relationship between Airbnb rental and crime.

Yan, J., Shu, X. and Yuan, H., 2010. Relationship between spatial distribution of thief crime and geographical factors. *Journal of Tsinghua University: Science and Technology*, 50(2), pp.174-176.