MODELING URBAN CRIME PATTERNS USING SPATIAL SPACE TIME AND REGRESSION ANALYSIS

H. Hashim 1, W.M.N.W. Mohd 1, E.S.S.M. Sadek 2, K.M. Dimyat 1

1 Centre of Studies for Surveying Science & Geomatics, Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, UiTM Shah Alam 40450 Shah Alam, Malaysia
2014994239@isiswa.uitm.edu.my; wmn@salam.uitm.edu.my; eran@salam.uitm.edu.my; khoiri_1852@uitm.edu.my

KEY WORDS: Crime Mapping, Urban Crime, Emerging Hot Spot Analysis, Space Time Analysis, Spatial Regression, GIS

ABSTRACT:

The population size, population density and rate of urbanization are often credited to contributing increasing a crime pattern specially in city. Urbanism model stating that the rise in urban crime and social problems is based on three population indicators namely, size, density and heterogeneity. The objective of this paper is to identify crime patterns of the hot spot urban crime location and the factors influencing the crime pattern relationship with population size, population density and rate of urbanization population. This study employed the ArcGIS Pro 2.4 tool such as Emerging Hot Spot Analysis (Space Time) to determine a crime pattern and Ordinary Least Squares (OLS) Regression to determine the factors influencing the crime patterns. By using these analyses tools, this study found that 54 (53%) out of 102 total neighbourhood locations (2011-2017 years) had a 99 percent significance confidence level where z-score exceeded +2.58 with a small p-value (p <0.01) as the hot spot crime location. The result of data analysis using OLS regression explains that combination of exploratory variable model rate of urbanization and population size contributes 56 percent ($R^2 = 0.559$) variance in crime index rate incident [$F (3.39) = 18.779, p <0.01$]. While the population density ($\beta = 0.045$, $t = 0.700$, $p> 0.10$) is not a significance contributor to the change in crime index rate in Petaling and Klang district. The importance of the study is useful information for encouraging government and law enforcement agencies to promote safety and reduce risk of urban population crime areas.

1. INTRODUCTION

United Nation (UNDESA, 2017, p.3) stated “crime is studied in order to prevent it”. The population size, population density and rate of urbanization are often crediting to contributing increasing a crime pattern especially in a city. Crime is omitted by people and thus a study of relationship of crime-based population is important and always significant relevant in modern society for a sustainability city. The study of urban population and crime almost 190 years old and it based on a noble first study about mapping crime and population in city by Guerry dan Quetelet (1833) in early 19th century at France (Eck & Weisburd, 1995; Paulsen & Robinson, 2004; Weisburd et al, 2012; Chainey & Ratcliffe, 2013; Wortley and Mazerolle, 2017). They found urban population composition is among size of city, population urban density and population size of city were related to criminality. The study of population and crime has attractive United Nation (1995) published white paper about concerning an urban crime phenomenon in world and provide a guideline that describe a population-based urbanism and socio-economic factors must be focus in country for preventing urban crime. Crime problem as describes by United Nation (2007) is affected by a series factors including population growth, urban density, the rate of urbanization, poor urban planning, design and management, poverty, inequality and political transitions.

Urban is defined a gazette area and its built-up area adjacent to it and has a population of 10,000 people or more, or special development area, or district administrative centre even though the population is less than 10,000 and at least 60% population age with 15 years and above engaged in non-agricultural activities (Second National Urbanization Policy Malaysia, 2016). The population provides a variety of positive and negative effects to everyday life, and among its negative side effects is the crime incidents daily are higher in urban than in rural areas. In Malaysia, high crime areas are in states with major urban hierarchy status and local cities such as in Selangor, Kuala Lumpur, Penang and Johor and are recognized by the government official report (GTP Annual Report, 2010; GTP Annual Report, 2011; GTP Annual Report, 2012; GTP Annual Report, 2013; GTP Annual Report, 2014; NTP Annual Report, 2015; NTP Annual Report, 2016; NTP Annual Report, 2017). Interestingly, in Malaysia, high-risk urban areas have experienced a significant reduction in crime - from 486 crimes that occur daily in 2010 year decreased to 272 crimes each day in 2017 year when the government launched a transformation plan 2010-2020 years to reduce the national crime rate, while Malaysia urban population is expected will increase from 74.8% in 2015 year to 83.3% in 2025 year (Second National Urbanization Policy Malaysia, 2016). This issue raises the question that, is its true crime reduction succeed while population increases every year or are there a crime pattern in certain urban areas always still high concentration (hotspots).

From a criminology approach, Brantingham and Brantingham (1984) defines the city “are generally considered bad places, places filled with crime, disease, and strife. It is often implied that crime, disease and strife are inevitable consequences of cities” (p.15). The rise in crime in the American cities in the early 20th century has attracted many sociologists from the University of Chicago or known by the Chicago School which is the pioneer of conducting criminal and urban population research and draws many studies on urban crime and well known in criminology science as Ecology of Crime Theory (Baldwin dan Bottom, 1976; Wilson and Schulz, 1978; Herbert, 1982; Davidson, 1981; Sampson et al, 1997; Harries , 1999;
Hayward, 2004; Chamlin and Cochran, 2004; Paulsen and Robinson, 2004; Crutchfield et al, 2007) and is the fundamental of the Environmental Criminology Theory, introduced by the end of the 20th century (Brantingham and Brantingham, 1984; Wortley and Mazerolle, 2017). Louis Wirth (1938) is a sociological figure from the Chicago School has put an urbanism model stating that the rise in crime, social problems and the way of life in the city, is based on three dimensions population indicators namely; size, density and heterogeneity. Size refers to the number of residents within a city area. Density refers to population density within a city area and heterogeneity refers to the number of racial diversities within a city area.

Based on previous researcher discoveries founds that refused and supported the urbanism model proposed by Louis Wirth (1938). Several studies (Wolfgang, 1968; Gale, 1973; Wilson and Boland, 1976; Skogan, 1977; Blau, 1977; Tittle, 1980; Conkline, 1981; Land et al, 1990; McCall et al, 1992; Ackerman, 1998; Ousey, 2000; Harries, 2006) support positive correlation relationship between crime rate with population size and density. While another researcher (Pressman and Carol, 1971; Kavalseth, 1977; Shichor et al, 1979; Chamlin and Cochran, 2004) reported unsupported the model, found insignificant and negative relationship between crime rate with population size and density. In Malaysia, a study by Sidhu (2005) based on index crimes from 1990 to 2002 years found an increase in crime rates affected by population growth, unemployment, races problem, illegal workers and narcotics.

It is worth noting, previous studies used correlation methods in examining the significant indicator of urbanism model without paying attention to spatial aspects of the location urban areas affected by crime incident. Crime pattern based hot spot that is critical to crime pattern-based urban risk population location should be noted as it impacts the effectiveness of policing policy and sustainable city program such as Safe City Initiative and Omnipresence Initiative applies recent by the Malaysia policing policy. Hence, this study applies the crime hot spot analysis along with the correlation method for testing the variable urbanism model pattern against the risk involved in urban areas for provide policing policy to make a better decision on preventing crime.

Therefore, the aim of study to model the spatial crime pattern of the study area. The objectives of this study are to map the crime pattern and to determine the population indicators factors that influencing the crime pattern.

2. METHODOLOGY

The study using GIS-based crime mapping methodology and tool use for analysis is Emerging Hot Spot Analysis (EHSA) based in Getis-Ord Gi* tool statistics provided in ArcGIS Pro 2.4 software to cluster a crime location distribution-based urban risk settlement for determine crime pattern for hot spot areas. Getis-Ord Gi* tool statistics is common techniques use in crime analysis and mapping (Gorr and Kurtland, 2012; Chaaney & Ratcliffe, 2013) and become the most popular amongst crime analysts (Chainey, 2015) because Gi* result are z scores used extensively in determining confidence thresholds and in assessing statistical significance of hot and cold spots location.

For this study, Emerging Hot Spot Analysis tool is used to identifies trends clustering of point densities (counts) and summary fields by Getis-Ord Gi* tool statistics in a space-time analysis created using Create Space Time Cube. The eight (8) categories hot and cold spots crime pattern result will be automated determine include new, consecutive, intensifying, persistent, diminishing, sporadic, oscillating, and historical. Hot spots are determined with red colours and cold spots are determined with blue colours in map. The focus of this study analysis is to identify four main output results for the crime pattern category namely;

i) New hot spot and a new cold spot (a location that is a statistically significant hot spot or cold spot and the most recent time step interval is hot for the first time),

ii) Intensifying hot spot and cold spot (a location that has been a statistically significant hot spot at least 90% of the time step intervals are hot, and becoming hotter over time)

iii) Persistent hot spot and cold spot (a location that has been a statistically significant hot spot at least 90% of the time step intervals are hot, with no trend up or down) and

iv) Diminishing hot spot and cold spot (a location that has been a statistically significant hot spot or cold spot at least 90% of the time step intervals are hot and becoming less hot over time) with a polygon-based output-based study location.

For parameter neighbourhood distance, the scale of analysis study is 43 urban boundaries for uniformity unit area analysis with standard distance interval is 400 meters with interval time for entire study for 84 months (2011-2017 years) and 1 month for each year for 7 years.

2.1 Study area and data preparation

This study area focuses on Petaling and Klang District, in the state of Selangor, Malaysia. Justification for the study area is due to the highest crime hot spot in Selangor based on official government report (GTP, 2011; NTP, 2015, NTP, 2017). The report indicates that Selangor State has the highest crime hot spot and from that, mostly crime hot spot is within Petaling and Klang District which are covered by 43 urban police station boundary with 4 City Council (Petaling Jaya, Subang Jaya, Shah Alam and Klang) as shown in Figure 1. Data set used is crime data index containing a set of x, y coordinates of 93,462 crime points 13 type index incidents with WGS 1984 World Mercator projection system from police department and web portal i-selamat.my from 2011-2017 years.

![Figure 1: The study area](image-url)
and cold spot result in attribute data analysis. The z-score is standard deviations and p-value is a probability. The z-scores (critical value) and p-values (significance value) are associated with the standard normal distribution. Very high or very low (negative) z-scores, associated with very small p-values, are found in the tails of the normal distribution. A confidence level for EHSA as standard by ESRI (2018) is at least 90% and above, is confidence accepted for determines a pattern and trend for features data analysis and to reject the null hypothesis that there is a clusters pattern with hot and cold spot crime with polygon-based fishnet result. Workflow method for EHSA as showing in Table 1.

Table 1: EHSA workflow using Model Builder ArcGIS Pro 2.4

The OLS regression were calculated using standard equation;

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + (\beta_n X_n) + \epsilon \]  

(1)

where

- \( Y \) = dependent variable (crime incident)
- \( X \) = independent/explanatory variables (size population, density population and urbanization rate)
- \( \beta \) = regression coefficients. Weights reflecting the relationship between the explanatory and dependent variable.
- \( \beta_0 \) = the regression intercept. It represents the expected value for the dependent variable if all the independent (explanatory) variables are zero.
- \( \epsilon \) = residuals represented in the regression equation as the random error term and the value not explained by the model.

Therefore, for modelling urban crime, OLS regression equation will be:

\[ \text{Crime Incident} = \beta_0 + \beta_1 \times \text{size population} + \beta_2 \times \text{density population} + \beta_3 \times \text{urbanization rate} + \epsilon \]

The model must meet all six assumptions required for validation (Mitchell, 1999, ESRI, 2018) and known as passing model’s where Exploratory Regression must run firstly:

1. Model Performance. The adjusted \( R^2 \) value must >.30
2. Model Coefficient has the reflecting the expected with positive or negative coefficient relationships
3. Model Significance. Access Joint F-Statistic and Joint Wald Statistic for a 95 percent confidence level, a p-value (probability) smaller than 0.05 indicates a statistically significant model.
4. Model Stationarity. Access the Koenker (BP) Statistic (for a 95 percent confidence level, a p-value (probability) smaller than 0.05 indicates statistically significant heteroscedasticity and/or nonstationary). The result must stationary.
5. Model Bias. The Jarque-Bera test is not statistically significant (larger than > 0.1) to show normally distributed.
6. Model Assess. Residual spatial autocorrelation (Moran’s I) must be random (p-value range between -1.65 to 1.65).

3. RESULT AND DISCUSSIONS

3.1 Urban Crime Pattern

In 2011 year, the space time cube has aggregated 7,479 points into 6,460 fishnet grid locations over 12-time step intervals. Each location is 400 meters by 400 meters square. The entire space time cube spans an area 38,000 meters west to east and 27,200 meters north to south. Each of the time step intervals is 1 month in duration so the entire time period covered by the space time cube is 12 months. Of the 6,460 total locations, 1,142 (17.68%) contain at least one point for at least one-time step interval. These 1142 locations comprise 21,698 space time bins of which 4,632 (21.35%) have point counts greater than zero. There is a statistically significant crime increase in point counts over time. The result as show in Figure 2 for EHSA and Figure 3 for space time cube.
In 2014 year, the space time cube has aggregated 14,816 points into 7,956 fishnet grid locations over 12-time step intervals. Each location is 400 meters by 400 meters square. The entire space time cube spans an area 46,800 meters west to east and 27,200 meters north to south. Each of the time step intervals is 1 month in duration so the entire time period covered by the space time cube is 12 months. Of the 7,956 total locations, 2,367 (29.75%) contain at least one point for at least one-time step interval. These 2,367 locations comprise 28,404 space time bins of which 8,409 (29.60%) have point counts greater than zero. There is a statistically significant crime decrease in point counts over time. The result as show in Figure 4 for EHSA and Figure 5 for space time cube.

![Figure 4: Crime pattern result using EHSA in 2014](image)

![Figure 5: Crime hot spots pattern by 3D visualization in 2014](image)

In 2017 year, the space time cube has aggregated 13,655 points into 9,744 fishnet grid locations over 12-time step intervals. Each location is 400 meters by 400 meters square. The entire space time cube spans an area 46,400 meters west to east and 33,600 meters north to south. Each of the time step intervals is 1 month in duration so the entire time period covered by the space time cube is 12 months. Of the 9,744 total locations, 2,704 (27.75%) contain at least one point for at least one-time step interval. These 2,704 locations comprise 32448 space time bins of which 8,909 (27.46%) have point counts greater than zero. There is a statistically significant crime decrease in point counts over time. The result as show in Figure 6 for EHSA and Figure 7 for space time cube.

![Figure 6: Crime pattern result using EHSA in 2017](image)

![Figure 7: Crime hot spots pattern by 3D visualization in 2017](image)

By overall in 2011 to 2017 years, the space time cube has aggregated 93,462 crime points incidents into 12,194 fishnet grid locations over 318-time step intervals. Each location is 400 meters by 400 meters square. The entire space time cube spans an area 53,600 meters west to east and 36,400 meters north to south. Each of the time step intervals is 1 week in duration so the entire time period covered by the space time cube is 318 weeks. Of the 12,194 total locations, 4,288 (35.16%) contain at least one point for at least one-time step interval. These 4,288 locations comprise 1698048 space time bins of which 76,082 (4.48%) have point counts greater than zero. There is a statistically significant crime increase in point counts over time. The result as show in Figure 8 for EHSA.
The results generated by the tool space time cube show trend crime data is statistically significant increase in point counts time study in 2011 year. However, trend crime data shows statistically significant decrease in point counts in year 2014, 2016 and 2017. Only trends crime data in 2012, 2013 and 2015 years are not statistically significant increase or decrease in point counts over time study based on the calculation of Mann-Kendall algorithm in EHSA as show in Table 2.

<table>
<thead>
<tr>
<th>Num</th>
<th>Year</th>
<th>Total</th>
<th>Trend bin</th>
<th>Trend statistic</th>
<th>Trend p-value</th>
<th>Crime Trend direction</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2017</td>
<td>116928</td>
<td>-1.7143</td>
<td>0.0865</td>
<td>Decreasing</td>
<td>Significant</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2016</td>
<td>122400</td>
<td>-2.6743</td>
<td>0.0075</td>
<td>Decreasing</td>
<td>Significant</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2015</td>
<td>130248</td>
<td>0.8248</td>
<td>0.4095</td>
<td>Not Significant</td>
<td>Not Significant</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2014</td>
<td>95472</td>
<td>-2.5372</td>
<td>0.0112</td>
<td>Decreasing</td>
<td>Significant</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2013</td>
<td>134064</td>
<td>-1.3029</td>
<td>0.1926</td>
<td>Not Significant</td>
<td>Not Significant</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2012</td>
<td>118524</td>
<td>-0.3429</td>
<td>0.7317</td>
<td>Not Significant</td>
<td>Not Significant</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2011</td>
<td>122740</td>
<td>3.7784</td>
<td>0.0002</td>
<td>Increasing</td>
<td>Significant</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4828824</td>
<td>1.9950</td>
<td>0.0460</td>
<td>Increasing</td>
<td>Significant</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Overall trend result by Space Time Cube (2011-2017)

### 3.2 Urban Crime Factors

The maximum confidence level set for OLS Regression (maximum coefficient p-value cut off) is \( p < 0.05 \). Explanatory Regression is carried out first to get the validation model through standard passing model. From Exploratory Regression, only model population size and rate of urbanization indicators meet the requirements of passing model’s predetermined test (Table 3 and Table 4) with Adjusted R-Squared \( R^2 \) is larger than \( > 0.3 \) (30%) where is \( 0.56 \) (56%), coefficient (Koenker BP Statistic) p-value cut off is less than \( < 0.05 \) (95%) where is \( p > 0.055 \) (stationary), VIF is less from \( < 7.5 \) (normally distributed), and spatial autocorrelation p-value is also larger than \( > 0.1 \) where is \( 0.2 \) (random). The survey data also did not show multicollinearity problems. Summary of variable significance shows that the exploratory variable for population density indicator has no significance \( (p > 0.10) \) compared to rate of urbanization indicators has positive significance level \( (p < 0.05) \) with 95% confidence level with negative linear relationship and size population have positive significance level \( (p < 0.01) \) with 99% confidence level with positive linear relationship. Significance, this model is substantial factors and contributes 56 percent \( R^2 = 0.56 \) variance to the index crime rate throughout the study year (2011-2017).

**Table 3: Exploratory Regression result for urban crime factors**

**Table 4: OLS Regression result for urban crime factors**

The result of data analysis using OLS regression (Table 4) explains that the combination of exploratory variable model; rate of urbanization, population size and population density contributed 56 percent \( R^2 = 0.559 \) variance in crime index rate incident \([F(3,39) = 18.779, p < 0.01])\). Rate of urbanization \((\beta = -88.067, t = -2.647, p < 0.01)\) and population size \((\beta = 0.556, t = 5.245, p < 0.01)\) are significance factor to crime index area. While the population density \((\beta = 0.045, t = 0.700, p < 0.10)\) is not a significance factor to the change in incidence rate of crime index. The coefficients value \((p < 0.01)\) for
population size would mean that for every unit people increase in population size, crime index rate incident (dependent variable), will increase by 0.556 per crime rate. This variable relationship is linearly positive. While coefficients value for variable urbanization rate would mean for every unit increase in the rate of urbanization, the dependent variable (crime index rate) decreases by 88.067 units per crime rate. In other words, the increasing population size variable will increase the number of crime index count. The lower the urbanization rate, the higher the index crime rate in the urban neighbourhood.

Result OLS spatial regression in the form of standard residual map (Figure 9) allows the data model of the study in visualization spatial showing the model population size and rate of urbanization performed to explain crime index rate incident in the study area.

![OLS Regression Model](image)

Figure 9: OLS spatial regression result residual map

Figure 9 shows the map of the red areas are under predictions (where the actual number of crime index rate incident high than the model predicted); the blue areas are over predictions (actual crime index rate incident is lower than predicted). As can be seen in the map, the population rate area for the police station neighbourhood, Sg. Way, S 17 (Section 17), Kota Damansara in Petaling Jaya and Bukit Puchong in Subang Jaya were major contributors to the crime index rate that was higher than the model predicted followed by Klang, Bandar Baru Klang, Sek. 6 Shah Alam, Subang Jaya, Kelana Jaya and Serdang. This neighbourhood is a major contributor to population size and urbanization rate to the crime index rate. The high under prediction area should be given a high priority in the policing policy following this model contributing 56 percent (R² = 0.56) variance changes to the rate of crime index year of the study.

4. CONCLUSION

The urban crime pattern shows half the neighbourhood-reaches 99.9 percent significance with a very large z-score of 3.2381 and a very small p-value (p <0.001) at SS 2 Sea Park, Petaling Jaya. Interestingly, there are 26 locations of new hot spot categories within the study area with each year of study. New hot spot locations are changing location position but cluster pattern within 12 months of time step interval analysis for each year. This condition can be shown in the base year of 2011, Taman Sri Medan, Petaling Jaya is categorized as a new hot spot category until 2014 and turned into a category of diminishing hot spot in the year 2015 and return to the persistent hot spot category in 2017. Despite several changes in pattern categories, Taman Sri Medan is a hot spot crime area that needs to be given priority over crime reduction in policing policy in Petaling Jaya.

Overall in urban crime pattern and factors, neighbourhood areas in Sg. Way (Taman Sri Medan), S 17 (Section 17), Kota Damansara in Petaling Jaya and Bukit Puchong in Subang Jaya were major contributors to the crime index rate that have positive significance level (p < 0.01) with 99% confidence level. This study is able to provide effective tools of interpreting hot spot crime-based risk population performance results statistically by hot and cold crime location and substantial factors contributes to crime for provide the Ministry of Home and Royal Malaysia Police to plan strategic implementation for preventing crime such as crime prevention through environmental designs (CPTED), Safe City Programs and Omnipresence Initiative.

5. RECOMMENDATIONS FOR FUTURE WORK

This study uses all 13 types of crime category indexes that include crime of violence and property crime for the general result. Therefore, studies by each type of violence and property should be a future study to get a more in-depth decision on the crime pattern and the relationship with factors influencing the crime pattern in study area. Other factors such as economics and lifestyle need to be given priority in linking causality and crime consequences in future studies.

ACKNOWLEDGEMENTS

The first author is the PhD candidate at Centre of Studies for Surveying Science & Geomatics, Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA. He is also the member of International Association of Crime Analysts (IACA) and Institution of Geospatial and Remote Sensing Malaysia (IGRSM).

REFERENCES


Science 277: 918-924.

Population density and criminal victimization. 
Criminology 17(2), 184-193.


