THE USE OF SPATIAL STATISTICS TO EXPLORE SPATIAL PATTERN AND RELATIONSHIP BETWEEN DRUG ADDICT AND SOCIO-ECONOMY

Nur Aisah Abu Bakar, Mohamad Nor Said and Anuar Ahmad

Department of Geoinformatics, Faculty of Geoinformation and Real Estate,
Universiti Teknologi Malaysia, 81310 UTM Skudai,
Johor, Malaysia
Tel:+067-5530808; Fax: +607-5566163
E-mail: sha_aisah@yahoo.com, m.nor@utm.my, anuarahmad@utm.my

KEY WORDS: Geographical Information System, Drug Addict, Socio-economic, Geographically Weighted Regression

ABSTRACT: Drug addiction is one of the major problems in many countries in the world. People become addicted to drug due to a number of reasons including emotional problems, poverty; influence by friends, stress and many more. However, these factors alone do not clearly show the phenomenon of the problem. It is subjective without describing the relationship between the contributing factors and the existence of the phenomenon. This research aims to determine the spatial pattern of the drug addicts and also to investigate the spatial correlation between their status and socio-economic influences by using spatial statistical analysis model together with Geographical Information System (GIS) technology. The area of study is Pontian district, in the state of Johor, Malaysia. The dependent variable of this study is the status of drug addicts (new addicts) while the independent variables are on the socio-economic side, the focus of which is on the education and employment levels. In order to gauge the relationship between these variables, a number of statistical analyses is carried out including the Mean Center, Standard Deviational Ellipse, Ordinary Least Square (OLS) and Geographical Weighted Regression (GWR). The results show that the distribution (pattern) of the new drug addict concentrates more in the urban and sub-urban areas. Furthermore, the correlation between the occurrences (arrests) of new drug addicts and their level of education and employment show a varying level of significance and also with different spatial distribution. The use of spatial statistics approach aided by GIS tools prove that the social problem such as drug abuse can be better explained and visualized in terms of the pattern of occurrences as well as the correlation between the contributing parameters. This certainly provides a useful assistance to the relevant authorities in their planning to more effectively combat such a very chronic problem

INTRODUCTION

In many developing countries, rapid changes in political alignment, community cohesiveness, unemployment, education level, economic and social marginalization have increased the number of drug addicts and crime cases (UNDP, 1993; 2001). In Malaysia, drug abuse has been one of the major social problems. For many years the government has relentlessly combated to eradicate this menace (UNDP, 2004). The statistics produced by the National Anti Drug Agencies (AADK) for the years of 2000 until 2010 indicate a significant increase in the number of drug users (addicts). Despite strict laws and penalties, illicit drug use is still high and continues to rise. Indeed, drug addiction is one of the major problems that reflects and contributes to the national issues.

Numerous studies at the provincial and national levels have been conducted to investigate the problem of drug addiction in Malaysia. However, the epidemiology and geographic distribution of the new status of the drug addicts and possible links to the socio-economic factors is not very much documented (AADK, 2011). Indeed, according to the United Nations Development Programme (2001), the factors of people becoming addicted to drug are due to a number of reasons including emotional problems; influence by adolescence; poverty; socializing; lack of mutual attachment and nurturing, availability and affordability of drugs; a chaotic home environment in which parents use substances and; stress and sex.

The factors influencing the people that get addicted to drug are only partially understood and subjective without describing the relationship among them and the existence of the phenomenon. This study focuses on this issue with an approach of exploiting the spatial statistics tools as provided by a Geographical Information System (GIS) in providing

more effective way of visualizing the phenomena. This includes the mapping of the drug addict cases (and study their distribution pattern) as well as the investigation of relationship between this pattern and the socio-economic influences, in this case the education and employment levels of the arrested addicts. Understanding the socio-economic influence that correlates between drug addict hotspots is important in urban public safety and planning for combating the spread of drug addiction in a certain area.

STUDY AREA

The area chosen for this study is Pontian District – one of the districts that form the Malaysian southern state, i.e. Johor. The whole district comprises of 31 planning blocks and covers an area of approximately 19,700 hectares with a total population of 171,291 (as of 2010). This area is chosen because of the increasing number of drug addicts and are considered (by National Anti Drug Agency- AADK) as a 'grey area' – an area representing the percentage of drug cases between 0.1% to 0.49% of the population of a certain bounded geographic bondary. According to AADK (2010), white area is categorized as less than 0.1% while black area is 0.5% and higher. In this case, the district of Pontian has a record of 0.15% of drug addict cases.

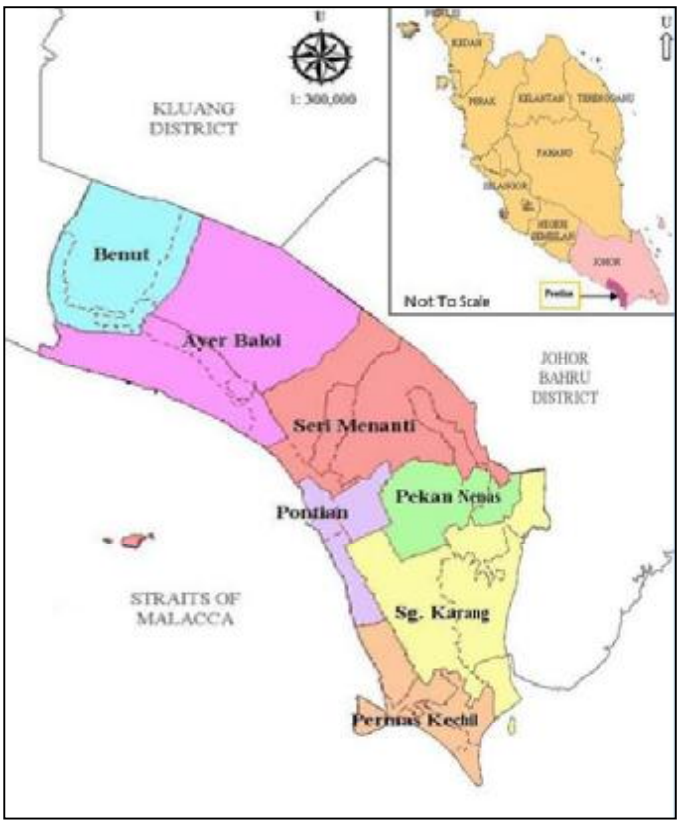

Figure 1: The study area – Pontian District, Johor

METHODOLOGY

In achieving the objectives of the research, two main approaches are adopted:

a) Modeling and analyzing the socio-economic influence
b) Geostatistical analysis of spatial pattern

Modeling and Analyzing the Socio-economic Influence

This study is conducted to achieve the following two main objectives:

i) to determine the spatial pattern of drug addict from 2007 to 2010 in Pontian District, Johor and
ii) to explore the spatial correlation between the new drug addicts with their level of education and employment (socio-economic variables)

In supporting the analyses, a number of statistical methods are decided and used i.e. Mean Center, Standard Deviational Ellipse, Ordinary Least Square (OLS) Regression and Geographically Weighted Regression (GWR). The Mean Center and Standard Deviational Ellipse methods are performed to visualize the geographic distribution of the drug addict cases (arrests) for the study years of 2007 to 2010. While the OLS and GWR are performed to investigate the spatial variation among the new drug addicts with the socio-economic variables, i.e. the level of education and employment.

The OLS method is a traditional approach where the underlying independent and dependent assumptions associated with the use of spatial data need not hold. Thus, alternative estimation strategies are required including those based on the local and global specifications, i.e. the use of GWR. This approach extends the OLS by taking spatial structure into account and estimates the local rather than the global model parameters. The variables used for the study are shown in Table 1. The dependent variable is drug addict and new status of drug addict (by *mukim* – province). On the other hand, the independent variables are the level of education and the level of employment. The status of drug addicts is provided by the Johor's National Anti Drug Agency (AADK), while the socio-economic data are obtained from the state's Department of Statistics.

Table 1: Dependent and independent variables

Variables	Definition and year
Dependent variables	
Drug addict by mukim	Percentage number of drug addict, 2007 to 2010
Drug addict status (new) by mukim	Percentage number of drug addict, 2007 to 2010
Independent variables	
Education by mukim (Primary school, secondary school University, non-educated, others)	Percentage number of each type of education, 2010
Employment Level by mukim (management, sales, services, student, labor, unemployed, construction worker, factory workers, agro transportation services)	Percentage number of each type of employment, 2010

In order to support the geospatial components of this study, administrative boundary map is prepared. The boundaries represent the Small Planning Blocks (BPK) of the Pontian District. Meanwhile, the associated attributes include the percentage of drug addicts per population, as well as the percentage of each type/ level of education (primary school, secondary school, university, non-educated and others) and employment (management, sales, services, student, labor, unemployed, construction worker, factory workers, agro**,** transportation services). In the stage of designing and

developing the database, all appropriate database development requirements have to be fulfilled. This includes formatting, error cleaning and topological construction, coordinate registration and attribute tagging. This is established using ArcGIS software package.

Geostatistical Analysis of Spatial Pattern

In this study, three methods of GIS spatial statistics are applied. It includes Mean Center, Standard Deviation Ellipse and clustering which consist of Ordinary Least Square (OLS) and Geographically Weighted Regression (GWR). The clustering method is well known because of its contribution to find the correlation between dependent and independent variables.

The analysis is divided into four parts. The first involves the use of mean center and standard deviational ellipse to portray the spatial pattern of the drug addict arrest locations (from 2007 to 2010 in the study area). The results of the mean center output are analyzed to determine the geographic center of a set of points as adjusted for the influence of a value associated with each point. The second part is to produce the drug addict hotspots. The application of standard deviational ellipse enables the determination of the range of these hotspots.

In the third part, the OLS regression approach is adopted as an attempt to explain the global relationship between the dependent and independent variables. The OLS tools calculate the coefficient for each explanatory variable in the model and perform a statistical test to determine whether the variable explains the phenomenon or not. The coefficient of r-squared (R^2) values should range from -1.0 to 1.0. Values closer to 1.0 represent a better predictive performance. OLS regression estimation is calculated using the formula below (Charton and Fotheringham, 2009):

$$y_i = \beta_0 + \sum_{k=1}^{n} \beta_k x_{ik} + \varepsilon_i$$

Where y_i is estimated value of the dependent variable for observation $_i$, β_0 is the intercept, β_k is the parameter estimate for variable x_{ik}, is the value of the k^{th} variables for $_i$, k is equal to 1,2 until n and ε_i is the error term (charton and Fortheringham, 2009).

In this study, y_i is the target value referring to the status of drug addicts and new drug addict. β_0 is the regression intercept which represents the location relationship over space of the expected value of the dependent variable if all of the independent variables are zero. β refers to the coefficient that represents the strength and type of relationship between the new drug addicts with socio-economic influence (education level and employment level). x_{ik} are the explanatory variables which are the percentage of education level and employment level.

Finally, a GWR model is performed in the fourth part of the analysis. This model is applied to analyze the spatial non-stationarity, defined as when the measurements of relationship among variables differ from location to location (Fortheringham *et al*.,2002). GWR is a localized multivariate regression that allows the parameter of a regression estimation to change locally. GWR detects spatial variation relationship in a model and produces maps for exploring and interpreting the spatial non-stationarity.

Using GWR estimation, the parameter for a regression model is calculated using the following formula (Mitchell, 2005):

$$y_i = \beta_0(u_{ik}, v_i) + \sum_{k=1}^{n} \beta_k(u_{ik}, v_i)\ x_{ik} + \varepsilon_i$$

Where $\beta_0(u_{ik}, v_i)$ is the estimated value of the correlation between dependent and independent variable for observation $_i$, and ε_i is the error term (Mennis, 2006).

The GWR output contains coefficient and t-value for every feature and explanatory variable. The choropleth mapping based on the coefficient value is performed in colour scheme and the estimated correlation should range between -1.96 to +1.96. On the contrary, the range of local R^2 values is between 0.0 and 1.0. These coefficients and R^2 values provide an indicator of how well the local regression model fits the observed values.

Another important value for assessing model performance is Akaike's Information Criterion (AIC). The AIC value is a useful measure for comparing models that have same dependent variables. The model with a smaller AIC value performs better because it provides a better fit for the observed data (Lauren *et al*., 2011).

RESULTS AND DISCUSSION

The map as shown in Figure 2 shows the results of computing the Mean Centre and Standard Deviational Ellipse of the drug addicts for the four study years (2007 through 2010). The mean centers estimate the focal point location of the drug addicts while the ellipses represent the coverage of highest drug addict activities (the hotspots). These analyses conclude that, the hotspots of drug addiction in this district are in the area of Pontian Tengah, Pontian Besar and Pontian Kechil.

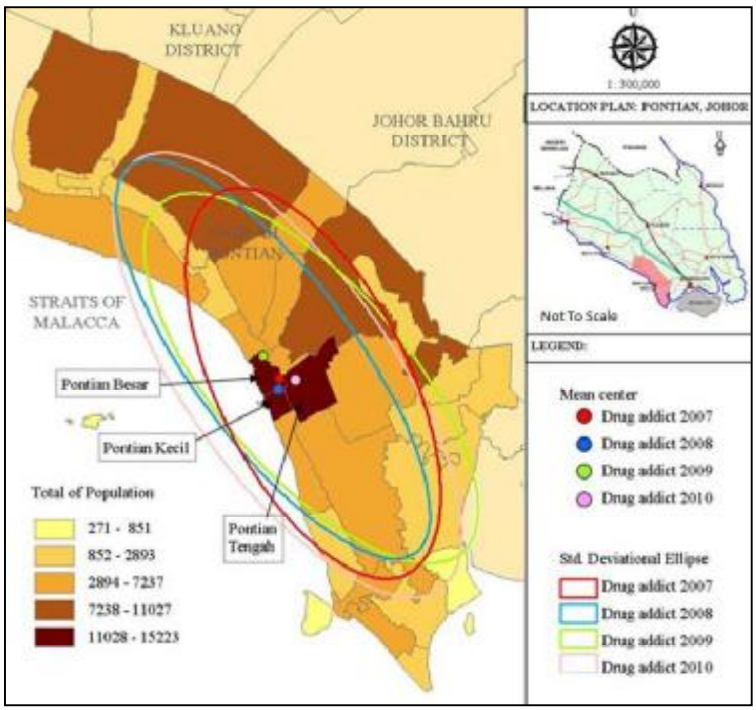

Figure 2: Drug addict hotspots and ellipse clusters

The results also provide information that the main concentration of the drug addict activities seems to take place in the high density population of the urban and sub-urban areas. This is represented by the dark color of the map. The change of the size of the ellipses (major and minor axes) for each year provides an indicator of the drug activities in the area, either increasing or decreasing. For example, comparing the distances of the axes of 2007 (major 14.74km; minor 17.8km) and 2010 (major 36.14km; minor 16.64km) will result in a conclusion that the activities decrease. The orientation of the ellipses demonstrates the pattern of movement of these activities.

A further analysis is to find the correlation between drug addict and socio-economic factors using OLS regression and GWR methods. Table 2 shows the statistical analysis performed using OLS model for the coefficient and t-value estimates. The null hypothesis is accepted and the correlation exists if the t-value or t-statistic is higher than 2, whereas the coefficient values must be close to 1 to reflect a strong correlation as compared to the smaller one. On the other

hand, the null hypothesis is not accepted and is considered no correlation between the variables if there is any negative value in t-value and the coefficient.

Table 2: OLS model parameter estimates for drug addict and socio-economic

Dependent variables/ correlation	***Drug addict***		***Correlation***		
Independent	***Coefficient (OLS)***	***t-value (OLS)***	***Yes***	***No***	***Significance***
Level of education					
Non-education	0.04	5.15	✓		
Others	0.01	1.64		✓	
Primary School	0.17	8.99	✓		
Secondary School	0.76	35.78	✓		✓
University	0.01	1.64		✓	
Level of employment					
Construction worker	0.01	4.33	✓		
Management	0.01	4.87	✓		
Sales	0.05	13.52	✓		
Technical	0.05	12.38	✓		
Transportation	0.04	14.32	✓		
Factory worker	0.09	23.86	✓		
Services	0.10	31.18	✓		
Agro	0.34	82.07	✓		✓
Labour	0.15	20.38	✓		
Unemployed	0.06	14.57	✓		
Student	0.01	9.16	✓		

***significance<0.05, N =31 (each of independent variables)

From Table 2, it can be identified that all levels of education and employment are generally related to the drug addict problem, as far as this case study is concerned. The only difference is the level of significance of the correlation. Based on the OLS summary table, the most significant factor is secondary school level of education and agro (farmers and fisherman) in the level of employment. The results reveal that, the correlation is almost positive in agro as compared to the other types of employment. This may be due to the low level of income of the addicts, the working environment, peer pressure and stress. On the contrary, the most positive group taking drug is children in secondary school. Most of them are at the age of 13 to 18 years. It may be due to the societal environment, influence of friends and social life.

Table 3, on the other hand, are the results used to find if there is any correlation between 'new drug addicts' and socio-economic factors (referring to secondary school and agro). The calculation is made on the coefficient for each variable in the model and a statistical test is performed to determine whether the variables explain the phenomenon. Looking at the coefficient and t-value in this table, the secondary school parameter is significantly correlated with the new status of the drug addicts. The coefficient shows that the secondary school is very close to the value of 1 and t-value is higher than 2. The agro (employment), on the other hand, is also positive but the value is very small.

Since secondary school and agro are not significantly correlated, GWR analysis is performed to further explore the finding by looking at more local level data. These results are used to identify the location with a positive relationship between the new drug addicts and socio-economic factors (level of education and employment).

Results revealed that GWR model exhibits a significant improvement in explaining the variance as compared to the OLS regression model. This is proven in the AIC score for the GWR model which decreases in all correlation i.e. from 54.5202 to 50.1103. Based on these AIC results, it suggests that GWR model for the Pontian district is better than the

OLS model. The R^2 results of all independent variables are explained by the value closer to 1 which indicates that the model has a better predictive performance.

Table 3: OLS and GWR Model Parameter Estimates for New Drug Addict

Dependent variables / Independent variables (X)	New Drug Addict (Y)					
	Coefficient **(OLS)**	**t-value (OLS)**	***Adjusted R^2(OLS)***	***Adjusted R^2(GWR)***	AIC (OLS)	AIC (GWR)
Significance			**0.980**	**0.986**	**54.5202**	**50.1103**
Secondary school	0.83	18.83				
Employment (Agro)	0.18	4.50				

***significance<0.05, N =31 (each of independence variables)

Figure 3 shows the spatial mapping that addresses the relationship between the new drug addicts with the socio-economic influences (secondary school and agro) by GWR method. In each map the classification is based on the significance of correlation (ranging from -1.96 to +1.96) shown in a variation of different colours. The correlation might be under predicted (less than -1.96) or over predicted (greater than +1.96).

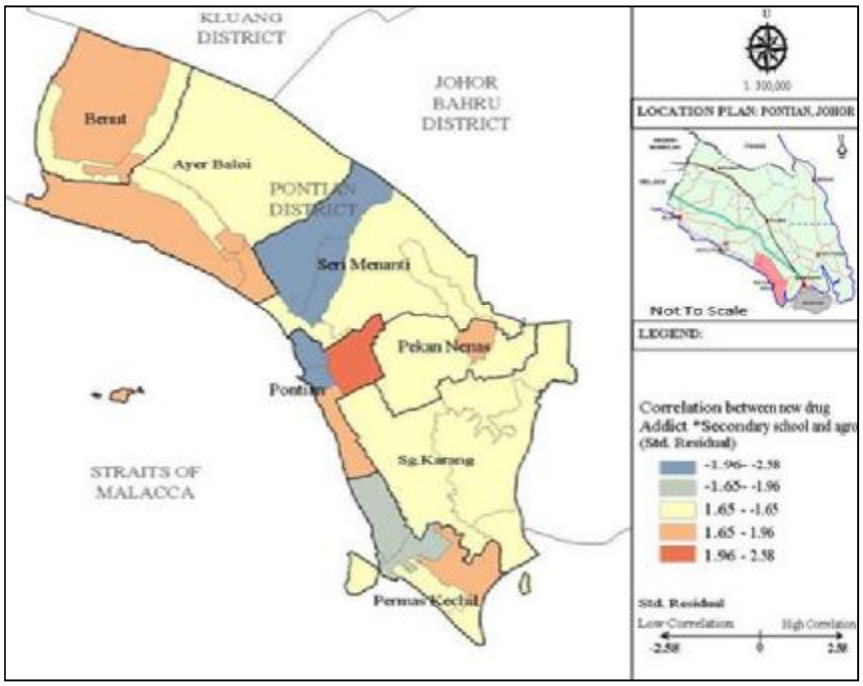

Figure 3: GWR model - correlation between new drug addicts with level of education (secondary school) and level of employment (agro)

The figure shows there is a correlation between the new addicts and the two variables being investigated and it occurs in a random pattern. The positive values indicate the area with the positive correlation between the variables while the negative values represent no correlation exist. Areas with high positive residuals include Pontian Tengah, Pontian Selatan, Chokoh, Bandar Pekan Nenas, Kuala Benut, Bandar Benut, Bandar Ayer Baloi and Parit Betak. It can be concluded that most of the new drug addict activities in this district concentrate in the urban areas. It may be due to a better accessibility to a road network and close proximity to the towns that encourage the supply of the drugs. Besides, the high population in the urban areas is advantageous to the addicts to move around with very less alert of the police observation.

CONCLUSION

This study has successfully explored the spatial pattern of drug addicts and its correlation between socio-economic influences (in this case the levels of education and level of employment) by analyzing a range of variables recorded by the National Anti Drug Agencies (AADK) related to where and why incidents occurred. GIS spatial statistical tools have many advantages in proving social problems such as drug abuse. It is best studied in terms of correlation and spatial distribution pattern. Knowing the location of arrests with a specific level of socio-economic of the addicts may provide a better picture of how drug addiction circulates. The advantage of this method is that the precise information of hotspot and positive correlation can be better achieved. This could assist the relevant authorities to do a better planning in combating drug-abuse related problem towards a safety environment.

ACKNOWLEDGEMENT

The authors would like to thank the Sustainability Research Alliance, Universiti Teknologi Malaysia for providing the research fund and the National Anti Drug Agency (AADK) for providing drug addict data that enable this study successfully conducted.

REFERENCES

Data Jumlah Penagihan Dadah Negeri Johor, Malaysia, 2011. Agensi Dadah Johor, Johor , Malaysia.

Fortheringham, A.S., Charlton, M., 2002. Geographically weighted regression, the analysis of spatial varying relationship, Chichester, UK

Kementerian Keselamatan Dalam Negeri, 2010. Maklumat Dadah Tahun 2007, 2008, 2009 dan 2010, Agensi Anti Dadah Kebangsaan (AADK) Malaysia.

Lauren, R. Lauren, S. and Monica, P., 2011. Findings a meaningful model, ArcUser Winter, ESRI

Mennis, J., 2006. Mapping the results of geographically weighted regression, Temple University, Philadelphia, USA. *The Cartographic Journal*, 43, 171-179

Mitchell, A., (2005) *The Esri Guide to GIS Analysis*, Carlifornia: Environmental System Research Institute Inc

United Nations Development Programme (UNDP), 2001. Governance, Poverty and Social Capital. *www.beta.undp.org.*

United Nations Development Programme (UNDP), 1993. Governance, Poverty and Social Capital. *New York, Oxford University Press.*

Zili, S. and William J. B., 2003. Handbook of drug abuse prevention; theory, science and practice, *Kluwer Academic Publishers*. New York