Application of spatial regression to investigate current patterns of crime in the north of Portugal

João Diogo Costa
NOVA IMS, Universidade Nova de Lisboa
Campus de Campolide
Lisboa, Portugal
g2016600@novaims.unl.pt

Ana Cristina Costa
NOVA IMS, Universidade Nova de Lisboa
Campus de Campolide
Lisboa, Portugal
ccosta@novaims.unl.pt

Abstract

Crime is one of the main problems of societies. However, research on this phenomenon has not yet reached a consensus on what factors influence the variability of crime rates and how it positions itself and acts in space. This work aims to contribute to a greater understanding of crime in the north of Portugal by investigating demographic and socioeconomic factors that may be associated with it. We explore the use of spatial statistics techniques through a detailed exploratory spatial data analysis (ESDA) first, and the application of regression models (Ordinary Least Squares and GWR – Geographically Weighted Regression) afterwards. The results show a crime hotspot on the coastline, and spatial homogeneity of low values between central-south municipalities. Crime patterns might be explained by population density, distance to the district capital and beneficiaries of Social Integration Income, who are individuals without socioeconomic conditions that need financial assistance from the Portuguese Government. However, predictions based on a GWR model with these factors may not be appropriate, given the limitations of this technique.
Keywords: crime pattern, spatial non-stationarity, spatial regression, ESDA, OLS, GWR.

1 Introduction

Crime is a phenomenon that accompanies societies from the earliest civilization and has a dynamic presence in time and space. Although individual incidences are unpredictable and difficult to anticipate, geographic studies have shown that crime is often concentrated in clusters (Wang et al. 2013), also named hotspots in statistical analysis, thus the phenomenon is neither random nor homogeneous in space, especially when considering urban areas (Nezami & Khoramshahi 2016). Researchers recognize the importance of considering the non-stationarity of the spatial process, and so they focus on the study of crime at the local level (Cahill & Mulligan 2007). Even though many geographers are interested in crime research, there are few attempts in the community to support and standardize this issue, and the "geography of violence" is still an emerging field of research (Springer & Le Billon 2016).

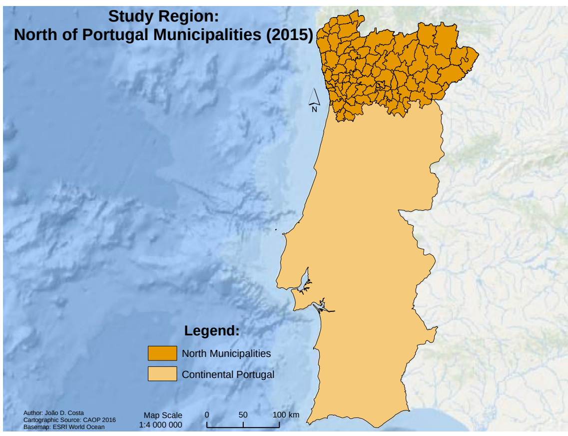
It is believed that spatial statistics techniques are little used by official entities in Portugal. Geographical Information Systems (GIS) experts, and geographers in general, can and should contribute to the development of the theoretical, territorial, spatial and planning aspects of the phenomenon. Crime is often set aside on national priorities, perhaps because of the label of "safe country" that the national territory boasts. However, low levels of crime are essential for the well-being of common citizens and society in general.

This research seeks to identify the current patterns of crime in the northern region of mainland Portugal. Additionally, we investigate potential covariates of crime.

2 Study region and data

The study variable 'crime' corresponds to the number of crimes recorded in 86 municipalities of the north of Portugal (Figure 1), in the year 2015. The types of crime accounted for include those against people, patrimony, life in society, State, pets, and others described in specific legislation. Additionally, data from 11 potential explanatory (socioeconomic) variables were collected in the PORDATA website (http://www.pordata.pt). An extra potential predictor was also computed, namely the 'Distance to the district capital' (i.e., direct distance from the municipality polygon centroid to the district capital).

Figure 1: Municipalities in the north of Portugal that define the study region.

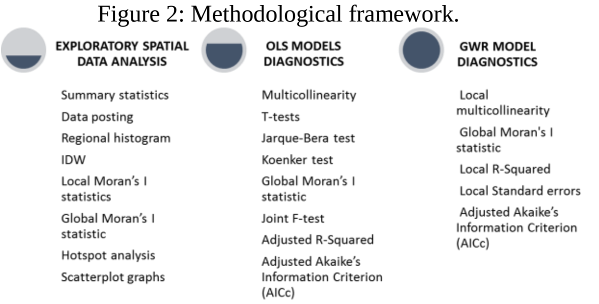


3 Methodological framework

The methodological framework included three sequential steps (Figure 2): 1) Exploratory spatial data analysis (ESDA); 2) Ordinary Least Squares (OLS) modelling and diagnosis; 3) Geographically Weighted Regression (GWR) modelling and diagnosis.

The GWR model used the same predictors of the final OLS model, and an adaptive kernel with a near-Gaussian weighting function given by $w_{ij} = exp[-(d_{ij}/b)^2]$, where d_{ij} is the distance between i and j, and b is the bandwidth, which is determined using the Adjusted Akaike's Information Criterion (AICc).

Figure 2: Methodological framework.



4 Results and discussion

4.1 Exploratory spatial data analysis

In the reference year, a total of 109 281 crimes have been recorded in the municipalities of the north of Portugal. The lowest value was 59 and the highest 16 056. Despite the large range of values, 75% of the municipalities had less than 1 321 crime records. We identified a global trend with values increasing from east to west, region where it is located the highest crime density. The regional histogram confirmed Porto and Vila Nova de Gaia, which share boundaries, as extreme values of the crime data distribution.

Setting a cut-off distance of 26 931m, the Local Moran's I statistic identified positive spatial autocorrelation (clusters of low and high values) at the 1% significance level (Figure 3), and the Getis-Ord Gi* identified a large hotspot of 13 municipalities (Figure 4).

Figure 3: ESDA – local spatial autocorrelation

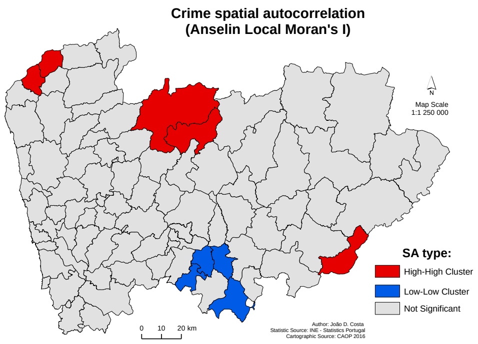
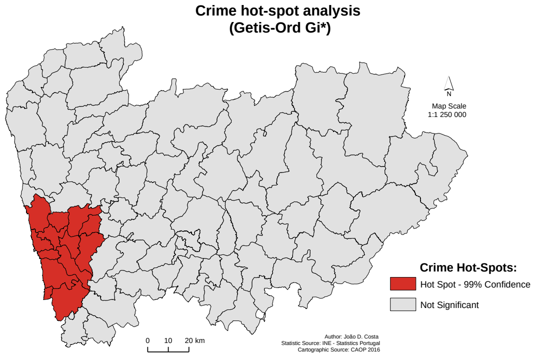


Figure 4: ESDA – hot-spot analysis



4.2 OLS models and diagnostics

In order to improve the OLS results, we transformed the dependent and independent variables into log2. Considering the transformed variables, the best model found was the one assuming that crime increases in the study region when population density and beneficiaries of the Social Integration Income (SII) increase, and when distance to the district capital decreases, *ceteris paribus*. SII beneficiaries are individuals with poor socioeconomic conditions that need financial support from the Portuguese Government. Those predictors do not exhibit multicollinearity (VIF values smaller than 3.5).

Overall, the model verifies the regression assumptions (Table 1), except in what concerns the residuals spatial autocorrelation. The residuals map (not shown) exhibits the highest positive errors in the north, and the lowest negative ones in the south.

Table 1: Final OLS model results and diagnostics.

Diagnostics	**Statistic**	**p-value**
Adjusted R-Squared	0.7887	
AICc	206.9527	
Joint-F statistic	106.7370	0.0000
Jarque-Bera statistic	2.3559	0.3079
Koenker statistic	5.4979	0.1388
Global Moran's I statistic	0.5096	0.0000
Model variables	**Coefficient**	**t-value**
Intercept	4.3857	6.6877
Population Density	0.2575	4.0011*
SII Beneficiaries	0.4540	5.8528*
Distance	–0.0737	–3.0488*
* Significant at the 1% level		

4.3 GWR model

Comparing with the OLS model, the GWR model improved the goodness of fit results: the AICc decreased to 155, and the GWR model now explains 90% of the crime variability (Adjusted R-Squared). It used 50 neighbours for each local estimation, which is an important parameter for the model. The effective number of parameters in the model was equal to 17. The sum of the squared residuals was 20, and the estimated standard deviation for the residuals was 0.54.

The regional variation of the coefficients has different patterns (Figure 5): population density has more impact in criminality on the west; distance to the district's capital is more relevant for municipalities located in the centre and south; and the coefficients of the SII beneficiaries have increasing values from west to east.

Unlike the OLS model, the spatial distribution of the GWR residuals exhibits a random pattern, which was confirmed by a non-significant value of the Global Moran's I statistic (p-value = 0.2052). The Condition Number exhibits an increasing pattern from west to east (Figure 6), but there is no evidence of multicollinearity among the predictors because all values are less than 30. The local R^2 are higher over the whole western region and smaller in the centre (Figure 7).

Even though the GWR model provided good diagnostic results, the interpretation of map patterns for individual coefficients should be done with caution, given the methodological limitations of GWR (e.g., Wheeler 2014).

Figure 5: GWR – coefficients

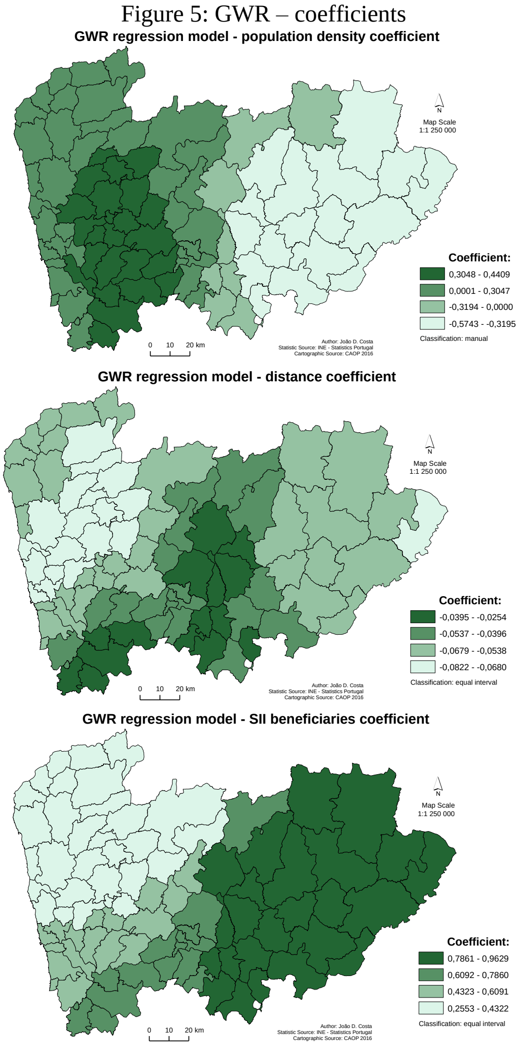


Figure 6: GWR – condition number

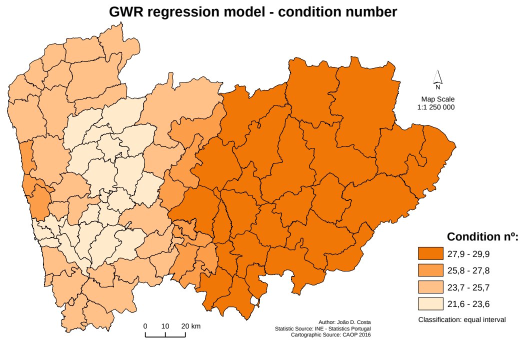
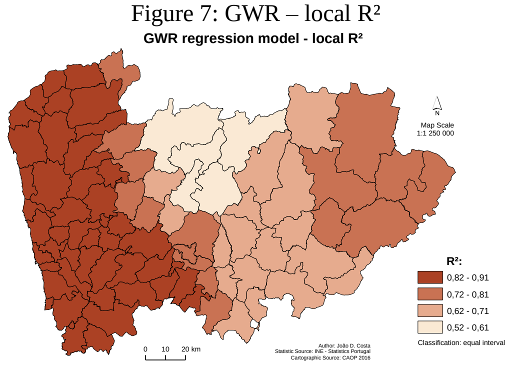


Figure 7: GWR – local R^2



References

Cahill, M., & Mulligan, G. (2007). Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, 25(2), 174–193.

Nezami, S., & Khoramshahi, E. (2016). Spatial Modeling of Crime by Using of GWR Method. In: *2016 Baltic Geodetic Congress (BGC Geomatics)*, Gdansk, June 2016, pp. 222–227. doi: 10.1109/BGC.Geomatics.2016.47.

Springer, S., & Le Billon, P. (2016). Violence and space: An introduction to the geographies of violence. *Political Geography*, 52, 1–3.

Wang, D., Ding, W., Lo, H., Morabito, M., Chen, P., Salazar, J., & Stepinski, T. (2013). Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns. *Computers, Environment and Urban Systems*, 39, 93–106.

Wheeler, D. C. (2014). Geographically weighted regression. In: *Handbook of Regional Science*, M. M. Fischer & P. Nijkamp (eds.), Springer Berlin Heidelberg, pp. 1435–1459. doi: 10.1007/978-3-642-23430-9_77