Social segregation in urban areas – an exploratory data analysis using geographically weighted regression analysis

Levers, Christian; Brückner, Maria; Lakes, Tobia
Geomatics Lab, Geography Department, Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin

Introduction

Social segregation in urban areas is a phenomenon with distinct spatial patterns. It is influenced by differences in socio-economic (e.g. income) and environmental (e.g. share of green space) factors. Since these influential factors vary in their intensity throughout space it is important to identify and to analyse the spatial patterns of these influencing factors in order to develop adequate coping strategies. Analysis of spatial autocorrelation, cluster analysis and geographically weighted regression may give new insights into spatial patterns and processes of the social segregation context in urban areas (Cahill & Mulligan 2007). In contrast to traditional global statistical regression analyses such as logistic regression or ordinary least square analyses spatial processes are addressed in this study in their very distinct behaviour using geographically weighted regressions (Brunsdon et al. 1998). The aim of this paper is to explore spatial patterns and underlying processes of social segregation in Berlin.

Material and Methods

We use an integrated dataset on environmental (e.g. noise pollution), socio-economic (e.g. property values) and demographic (e.g. population density) variables to explore the spatial patterns of the social status which is taken as an indicator for social segregation. All data was made available for the whole area of Berlin, Germany, on a detailed and spatially explicit level of spatial units characterised by the distinct living environment of their inhabitants. In a first step we check the social status data for spatial autocorrelation using the global Moran's I (Moran 1950). By detecting spatial autocorrelation we conclude that locality is an essential point in describing and explaining the social status. Thus, we check the data for spatial autocorrelation on the local level utilising Anselin's local Moran's I and univariate Local Indicators of Spatial Association (LISA) to investigate "hot- and cold-spots" of social status (Anselin 1995). In order to explain the given distribution of the social status we perform a geographically weighted regression to account for spatial variations in the social status data. To do so we first reduce the number of independent variables in a stepwise multiple regression model to avoid multi-collinearity between the variables. Finally, we apply the geographically weighted regression for the dependent variable "social status" and the selection of independent explanatory variables.

Results and Discussion

On a global (Berlin-wide) scale we could observe a significant clustering of the social status with a Moran's I of 0.284 (Z score = 24.745, p = 0.000, variance = 0.000134). Thus, we can assume that there is less than 1 % likelihood that this observed pattern could be the result of random chance. Like it was expected locality does play a major role in the spatial distribution of the social status in Berlin. On the same Berlin-wide level we could identify clusters of high and low social status and map these for the area of Berlin (figure 1). Local autocorrelation analysis showed Moran's I values of -0.043 to 0.063 ($-13.43 \leq Z$ score ≤ 14.294, $0.000 \leq p$ value ≤ 0.999). A total of 153 out of 447 spatial units were identified to be significant on the 0.05 level (see fig. 1). HH clusters indicate city areas with low social statuses (indicated by high values due to data source characteristics) which are located next to low social status areas forming a "low social status cluster". LL clusters indicate city areas with high

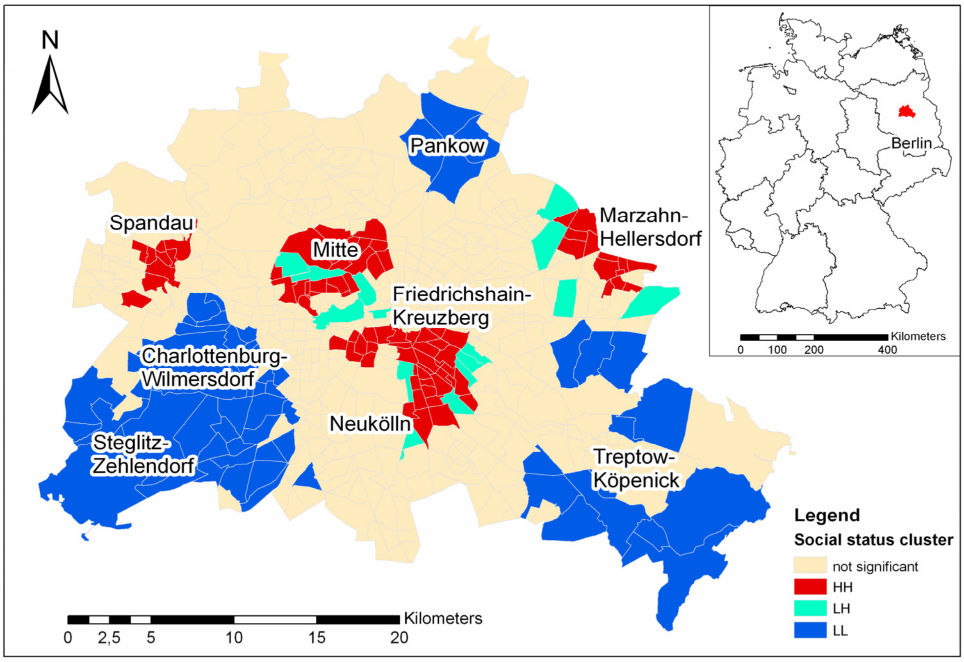

Figure 1: Clustering of social status in Berlin, Germany.

social statuses (indicated by low values) which are located next to high social status areas forming a "high social status cluster". LH clusters indicate outliers, in this case areas of high social statuses which are located next to areas with low social statuses. City districts with a high social status (indicated by low values due to data source characteristics) can be found in the south eastern and south western part of Berlin namely in Steglitz-Zehlendorf, Charlottenburg-Wilmersdorf and Treptow-Köpenick. In a part of Pankow in the northern part of Berlin a high social status is observable. Areas of low social status are located in the centre of Berlin (Mitte, Friedrichshain-Kreuzberg and Neukölln) and in the western (Spandau) and eastern (Marzahn-Hellersdorf) part. Generally it can be stated that city areas with a high social status are located in the border area of Berlin whereas areas with a low social status tend to be in centre of Berlin. Furthermore it is noticeable that there are several clusters of high social status next to clusters with low social status (Mitte, Neukölln, Marzahn-Hellersdorf). These "outliers" show that even in areas of low social status good living quarters can be found. To explain the given distribution of social status we employ a geographically weighted regression analysis (GWR) (Brunsdon et al. 1998 and Fotheringham et al. 2002). First we identify the most influencing variables on social status by using a stepwise regression model to avoid multi-collinearity within the set of variables. With this set of selected variables we perform the GWR. Thus, we can identify the likely influencing socio-economic and environmental factors for the derived spatial clusters of social status in Berlin. So the different processes in space which underlie these patterns can be approximated by the geographically weighted regression in their very distinctive spatial characteristics.

Conclusions and outlook

The variability of processes and factors leading to social segregation in urban areas is pointed out in our study using geostatistical and geographically weighted regression analysis. Social status data is analyzed regarding spatial autocorrelation and explained by socio-economic and environmental variables. Knowledge about the location of socially disadvantaged areas and the underlying driving factors for this disadvantage can support decision-making, improve social equality policies and offer a screening for further detailed analysis.

BIBLIOGRAPHY

Anselin, L. (1995). Local indicators of spatial association – LISA. Geographical Analysis, 27, 93-115.

Brunsdon, C., Fotheringham, A.S., Charlton, M. (1998). Geographically weighted regression - modelling spatial non-stationarity. The Statistician, 47, 431-443.

Cahill & Mulligan (2007). Using Geographically Weighted Regression to Explore Local Crime Patterns. Social Science Computer Review 2007; 25; 174.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). Geographically weighted regression: The analysis of spatially varying relationships. Chichester: Wiley.

Häußermann H, (2009) Monitoring Soziale Stadtentwicklung Berlin 2009 (Senatsverwaltung für Stadtentwicklung Berlin, Berlin).

Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena. Biometrika, 37, 17-33.