"Mapping Social Exclusion/Inclusion in Developing Countries: Social Dynamics of São Paulo in the 1990s"

GILBERTO CÂMARA¹, ANTONIO MIGUEL MONTEIRO¹, FREDERICO ROMAN RAMOS¹

ALDAIZA SPOSATI², DIRCE KOGA²

¹Image Processing Division, National Institute for Space Research (INPE),

Av dos Astronautas 1758, São José dos Campos, SP, Brazil

²Social Studies Research Center, Catholic University of São Paulo (PUC/SP), Brazil

DRAFT VERSION

1 Introduction

This work presents a methodology for mapping social exclusion and inclusion in large urban areas in developing countries and for using spatial analytical techniques for identifying significant spatial patterns of this phenomenon.

Social exclusion and inclusion have become a major policy issue, both in developing and developed nations. The United Kingdom government has established a "Social Exclusion Unit" which was set up by the Prime Minister in 1997 and in the United States, initiatives include the "National Neighborhood Indicators Partnership", which aims to use geographical information as a means of improving the awareness of citizens and urban planners about the different dimension of deprived urban areas (Kingsley, Coulton et al., 1997). In this context, the use of maps has been recognized as a major tool for enabling new ways of thinking about urban issues, including social exclusion/inclusion. In many cases, local authorities have made extensive effort to use GIS as a basis for portraying local social and economical conditions.

Notwithstanding the impact of maps in promoting awareness about living conditions, considerable less attention has been devoted to the use of spatial analytical techniques in conjunction with social exclusion/inclusion indicators. Although the phenomena of spatial dependence has long been recognized as an intrinsic feature of spatial data (Goodchild, 1988; Bailey and Gattrel, 1995), many urban planners and policy makers do not take this issue into account when producing spatial information.

The quantitative expression of this *spatial dependence* principle is the effect of *spatial autocorrelation*: the observed values will be spatially clustered, and the samples will not be independent. Additionally, most spatial data sets, such as those obtained from geo-demographic and health surveys, not only possess *global* spatial autocorrelation, but also exhibit significant patterns of spatial instability, which is related to regional differentiations within the observational space. In this light, the results presented in this chapter illustrate how spatial analytical techniques can enhance the understanding of social exclusion/inclusion patterns in large cities of the developing world.

The work is structured as follows. Section 2 discusses the problem of mapping social exclusion in the Third World. Section 3 presents the methodology used for producing the social exclusion/inclusion indicators for São Paulo and Section 4 describes the "Map of Social Exclusion/Inclusion of São Paulo", which has had a major public and policy impact. Section 5 indicates how local spatial autocorrelation indexes were used to identify significant clusters of social exclusion/inclusion in São Paulo. Section 6 describes the use of spatial econometrics techniques for identification of spatial regimes in São Paulo and calculation of spatial regressions, which portray the influence of individual components in the overall exclusion/inclusion indices. Section 7 illustrates the use of geostatistical techniques for mapping spatio-temporal trends in the evolution of important socio-economical variables. Section 8 concludes the work, with a discussion of the impact of spatial analytical techniques in the study of social exclusion/inclusion and an indication of how interested researchers can obtain the data sets and software used in this study.

2 Mapping Social Exclusion in the Developing World

The concept of social exclusion was born in Europe, motivated by the sharp increase in the number of poor, whose numbers in the twelve countries of the EEC went from 38 million in 1975 to 53 million in 1992. But, it is fair to ask: What precisely does exclusion mean? Who is excluded? And why? How does this concept differ from that of poverty? Does exclusion refer to a problem in the distribution of wealth or a loosening in the ties that bind society together? (Bessis, 1995).

This discussion is particularly relevant in the case of developing countries, whose societies have never had the kind of social protection typical of the Welfare State societies of the 20th century, especially in Europe. In higher-income countries, "social exclusion" tends to be associated to processes of *social disqualification* and from economical and social problems that affect urban areas, many of whom have had previously better living conditions. In the definition of UK's Social Exclusion Unit, "Social exclusion is a shorthand term for what can happen when people or areas suffer from a combination of linked problems such as unemployment, poor skills, low incomes, poor housing, high crime environments, bad health and family breakdown" (Blair, 1998).

By contrast, social exclusion south of the Equator has a significantly different setting. In most areas suffering from social exclusion, its population has never had acceptable living conditions. Therefore, in developing nations, social exclusion has to be measured from the viewpoint of considering what would be a *basic standard of living* (Sposati, 2000). Establishing the basic standard of life implies defining what needs are considered basic and universal according to a collective ethic of life.

In order to express social exclusion as a spatial pattern, our approach was to derive a composite *social exclusion/inclusion index* (**IEx**), which is aggregated by areal units. The social/exclusion index (**IEX**) is divided into four dimensions: IEX-autonomy (**IEX-a**), IEX-life quality (**IEX-qv**), IEX-human development (**IEX-dh**) and IEX-equality (**IEX-eq**). Each dimension is captured by a set of variables, obtained from census and field data collection, as outlined in Table 1. For each variable, (Sposati, 2000) proposes a *reference value* which indicates what should be considered as the attainment of *a basic standard of*

inclusion. Areas which achieve such levels are assigned a value of 0 (zero), whereas areas with values above such reference are mapped to a positive [0..1] scale, and areas below such reference are assigned negative values on a [-1...0] scale. Therefore, each of the indexes has a range between -1 (total exclusion) and 1 (total inclusion).

This methodology is fundamentally different from widely used indicators such as the human development index (HDI), developed by Nobel Prize economist Amartya Sen and produced by the United Nations Development Program (Jolly, 1999). HDI is a *ranking* indicator, in that it provides an ordering of human development in the world, based on three components: life expectancy at birth, education standards and per capita income. The HDI does not provide an objective measure of whether a country's citizens have achieved an acceptable standard of living. By contrast, the IEx aims at informing citizens, decisionmakers and the media with an appraisal of the gap that separates the richest part of the society to the less privileged ones.

3 The Map of Social Exclusion/Inclusion in São Paulo

The methodology described in Section 2 was used to assess the evolution of the city of São Paulo during the 1990's. São Paulo presents an important challenge to social and urban planners, a city that it is simultaneously one of the world's largest (9 million people), Brazil's richest (as measured by the industrial and service goods output) and the one which includes the largest amount of socially excluded citizens in Brazil. To allow for a common geographical basis for the different data sets, the study used the official division of São Paulo in 96 districts. The data sets used included: (a) the 1991 census and the 1996 population assessment, from IBGE (Brazil's Bureau of Census); (b) the 1987 and 1997 living conditions surveys by the Companhia do Metropolitano de São Paulo (Subway Authority); (c) the 1996 and 1999 homicide rates produced by Fundação SEADE (São Paulo State Statistics Bureau); (d) Information on infant mortality rates by PROAIM (Public Safety Secretariat of the São Paulo city).

In this period, two main results have been produced:

- The "Map of Social Exclusion/Inclusion of São Paulo 1995", which used data available from 1987 to 1.95 to produce indicators for the earlier part of the 90's (Sposati, 1996);
- The "Map of Social Exclusion/Inclusion of São Paulo 2000", which concentrated on trends on population, employment and quality of life indicators during the 1990's (Sposati, 2000).

The 1995 Map had a large impact on media, political and academic awareness of the issue of social/exclusion inclusion in São Paulo and its results were used by many researchers and public policy administrators. One of its important results of the 1995 map was to indicate a significant gap between the social exclusion and the social inclusion regions of São Paulo, where 2/3 of its districts of were below acceptable levels of living standards, as shown in Figure 1.



Figure 1 – Social Exclusion Index in São Paulo obtained from IBGE 1991 census data (96 districts grouped in sextiles). Values vary from -1 (extreme exclusion) to +1 (extreme inclusion).

The 2000 Map has had a similar impact, with substantial media coverage and public and academic interest. Its results are being used by the mayor of São Paulo elected in 2000, Marta Suplicy, to subsidize public policy and government investment in the city. The main results of the 2000 Map were a significant change in population dynamics and a strong relation between education and unemployement trends.

In terms of population, whilst there has been a small increase of the overall population from 1991 to 1996 (from 9,646,185 inhabitants to 9,839,066 inhabitants or a 2%-growth), there has been an important inter-urban trend in that the poorest regions of the city have registered population increases up to 129.96%. This trend is also markedly skewed in the range of 15 to 24 year olds, which has grown by 75,000 people, mostly in deprived neighborhoods. One of the consequences is a large increase in violence and homicide rate, since youngsters have access to information, but do not have the means to obtain consumer goods. Therefore, teenage violence has grown markedly in São Paulo during the 90's.

4 Finding Social Exclusion/Inclusion Clusters with Spatial Autocorrelation Indexes

In order to provide additional insight into the inner-city dynamics of São Paulo, we have used spatial analytical techniques. Initially, we investigated if local spatial autocorrelation indicators (Getis and Ord, 1996) could indicate clusters of social exclusion/inclusion in the "Map of Social Exclusion/Inclusion-1995". To that end, we chose the composite **IEx** (social exclusion/inclusion index), who exhibits a significant global spatial autocorrelation (Moran's I = 0,65, with significance of 99%). We used the local Moran index I_i , which is computed by multiplying the local normalized value z_i , by the local mean (Anselin, 1995):

$$I = \frac{z_i \sum_{i}^{n} w_{ij} z_i}{\sum_{i}^{n} z_i^2}$$

To establish a significance test for the local Moran index, we used the method proposed by Anselin (1995): a pseudo-distribution simulation by permutation of the attribute values among the areas. Based on this pseudo-distribution, statistical tests are used to indicate local index values with significance of 95%, 99% and 99,9%. The 'significant' indexes are then mapped and posited as 'hot-spots' of local non-stationarity.

The local Moran index significance map indicated three 'hot spots', two of them related to low values of inclusion (located to the South and East of the city) and one related to high values of inclusion (located in the Center of the city), shown in Figure 2. These clusters correspond to areas that concentrate a significant amount of the city's disparities.



Figure 2 - Significant values of Local Moran Index for Social Exclusion Indicator for São Paulo, for the 1995 data set (blue=95% significance; green=99% significance)

The results of the LISA indicators correlate very well with the inequalities of São Paulo. The city's South region corresponds to an expansion area that has had an explosive growth in recent times, occupied by migrant workers which come to São Paulo from other parts of the country, with very limited public investment. In the East of São Paulo, the concentration of low-income population is a direct consequence of public policies of the 70's and 80's, which removed poor people from slums located in the central (and wealthiest) part of the city towards publicly-built housing estates in the region. These housing estates were inadequately built and rapidly degraded into crime-infested areas. In the center of São Paulo, the LISA map shows a significant cluster of high-income area, where the wealthiest part of the population lives. As a result, the LISA indicators have proven effective in distinguishing the extreme concentrations of wealth and poverty in São Paulo.

5 Spatial Autocorrelation Patterns as a Basis for Redistricting

We have also investigated the use of spatial autocorrelation indicators as a basis for the design of administrative zoning systems for the city of São Paulo. As is well known, zone design is a major challenge for urban and regional planners, since it involves major decisions on how to distribute public resources. Given the enormous socio-economic inequalities of São Paulo, the rational planning of the city requires a careful division of the urban space into administrative regions that are homogenous by some objective criteria. Unfortunately, the current grouping of the 96 districts of São Paulo into 11 Administrative Regions has been driven by historical and political forces, and does not reflect a rational attempt to challenge the city's disparities.

Taking the social exclusion index as a basis, we have grouped the 96 districts into a set of administrative zones, each containing a significant number of districts, and homogeneous with respect to social exclusion status. We used two exploratory spatial analysis tools: the local Moran index significance map (figure 2) and the Moran Scatterplot Map.

The *Moran scatterplot map* (Anselin, 1996) is a tool for visualization of the relationship between the observed values Z and the local mean values WZ, where Z indicates the array of attribute values (expressed as deviations from the mean), and $WZ_{:}$ is the array of local mean values, computed using matrix W. The association between Z and WZ can be explored to indicate the different spatial regimes associated to the data and display in a graphical form. The Moran Scatterplot Map divides spatial variability into four quadrants:

• Q1 (positive values, positive local mean) and Q2 (negative values, negative local means): indicate areas of positive spatial association.

• Q3 (positive values, negative local means) and Q4 (negative values, positive local means): indicate areas of negative spatial association.

Since the *Iex* variable exhibits global positive spatial autocorrelation (Moran I = 0.65, significance= 99%), districts in quadrants Q3 and Q4 are interpreted as regions that do not follow the same global process of spatial dependence and these points indicate transitional regions between two different spatial regimes. Figure 3 shows the Moran scatterplot map for the social exclusion index of the 1995 map.



Figure 3 – Moran scatterplot map for Social Exclusion Index for 1995 data set in São Paulo.

We used the clusters found in the LISA map as "seeds" in the zoning procedure. The remaining regions were defined interactively, taking into account the Moran scatterplot map, which indicates a number of transition regions between the regions of Q1 and Q2 locations (to so-called "high-high" and "low-low" areas). These regions were grouped into separate zones. The method proceeded interactively, until a final set of spatial regimes was produced. These set of spatial regimes can be also considered as a first approximation to a redistricting proposal for the city of São Paulo, which can be confronted with the current administrative regions. Figure 4 shows the current administrative regions in São Paulo and the proposed redistricting, according to a spatial analysis procedure outlined above.



Figure 4 – Left: Current Administrative Zones in São Paulo. Right: Spatial regimes identified by the social exclusion index.

A comparison of the two maps shown in Figure 4 indicates that the proposed set of spatial regimes provide support for a more socially oriented administrative division of São Paulo. This results are a positive indication of the possible use of local spatial statistics as a basis for zoning procedures and show how indicators such as the social exclusion index of Sposati (1996) can be used as a support for urban planning.

6 Spatial Econometrics Analysis of Social Exclusion/Inclusion Factors

In our investigation of social exclusion/inclusion process in São Paulo, we also wanted to assess the relative influence of the factors that produce the overall index. The 1995 Map used 45 variables, but such a large data set may not be always available for researchers. This raises an important question: *what is the minimum set of variables, which can still produce a credible result for the exclusion/inclusion index? Is there a variable that is a determinant factor for social exclusion/inclusion ?* This issue is very relevant in connection to spatial statistical studies in developing countries, where only limited data sets exist.

In order to establish a correlation between the relevant factors and the composite indexes, we investigated three different types of spatial regression models, as discussed in (Anselin, 1988): the *spatial autoregressive lag* model, *the spatial autoregressive error* model and the *spatial regimes* regression.

Each of these models is briefly described, following (Bivand, 1998). Starting from a linear regression model formulation,

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0,\sigma^2)$$

where is y is an $(n \ x \ l)$ vector of observations on a dependent variable taken at each of N locations, X is an $(n \ x \ k)$ matrix of exogenous variables, β is an $(k \ x \ l)$ vector of parameters, and ε is $(n \ x \ l)$ an vector of disturbances, the spatial autoregressive lag model can be expressed as

$$y = \rho Wy + X\beta + \varepsilon$$

where ρ is a parameter that expresses the autoregressive relationship between the observations. The spatial autoregressive error model can be expressed as

$$y = X\beta + \lambda Wu + \varepsilon$$

where λ is a scalar spatial error parameter, and u is a spatially autocorrelated disturbance vector.

The spatial autoregressive lag model and the spatial autoregressive error model both aim at exploring the global patterns of spatial autocorrelation in the data set. By contrast, the idea of the spatial regimes regression model is to divide the observation set into subsets, each with a different spatial pattern, and to perform separate regressions, one for each subset.

The existence of significant "clusters" of social exclusion/inclusion in São Paulo, as discussed above, indicates that several regimes of spatial association are present in the city and that the assumption of global stationarity is not supported for this data set. In this case,

regression models which only take in account the global autocorrelation (such as the spatial error or the spatial lag model) may fail to account for local correlation structures.

In order to assess the explanatory potential of the spatial regression techniques, we performed a regression analysis on the correlation between the percentage of family heads with more than 15 years of schooling (as the independent variable) with the social exclusion index (as the dependent variable). Four regression analyses were performed using the SpaceStat software (Anselin, 1992): standard OLS regression; spatial lag model; spatial error model; spatial regimes using the proposed new zoning (discussed in the earlier section). The results as summarized in Table 1, where the comparison criteria used were:

- R² the standard "goodness-of-fit" measure, which is inadequate for spatially dependent data, as discussed in (Anselin, 1988).
- Likelihood maximized log-likelihood assessment of model fit (a preferred measure, according to (Anselin, 1992).
- MI-error global spatial autocorrelation indicator of residuals.
- LM-error Lagrange Multiplier indicator (assesses the extent to which there remains spatial autocorrelation in the residuals)

TABLE 1

EDUCATION X SOCIAL EXCLUSION IN SÃO PAULO

	OLS	Spatial Lag	Spatial Error	Spatial Regimes
\mathbf{R}^2	0.75	0.77	0.79	0.88
K	0,75	0,77	0,75	0,00
Likelihood	14,9	20,53	26,68	49,93
MI-error	0,384	-	-	-0,006
LM-error	29,43	12,19	16,58	0,006

(RESULTS FROM REGRESSION MODELS)

The spatial regimes regression model was shown to be significantly superior to the other regression models by all of the comparison criteria used. It had a better fit to the observed data, and the model residuals exhibited significantly less spatial autocorrelation. As a conclusion, it can be shown that, in data sets which combine global and local spatial autocorrelation patterns such as the case of São Paulo, the estimation of a set of spatial regimes is a useful intermediate step before using spatial regression models.

7 Mapping Trends: São Paulo in the 90's using Geostatistics

One of the aims of the social exclusion study is the identification of trends in the city. To that end, and in order to "filter out" local instability, we conducted a study using geostatistical techniques to produce "social topography" surfaces which indicate citywide trends. To produce these maps, we obtained a sample set using the city's districts by assigning a sample at the center of each district. These samples were then used as a basis for establishing a global autocorrelation pattern by means of a variogram analysis, and a surface was interpolated by ordinary kriging.

As an example of the trend surfaces we obtained, consider the homicide rate for São Paulo in the years 1996 and 1999, shown in Figure 5. These figures show a clear spatial trend, in which the areas with lowest homicide rate (below 30 deaths per 100,000 inhabitants) have significantly decreased in 1999 in relation to 1996. Since these are correlate very strongly with the wealthiest regions of the city (compare with Figure 1 and 2), these results indicate a spatial spreading of crime, in that violence is not confined to the poorest areas of the city and that the inhabitants of richer areas are increasingly prone to be victims of violent assaults. These results were published in the national media (Folha de São Paulo, Brazil's largest newspaper) and have had a major impact on public awareness of the spatial trends of crime in the city.



The authors are aware of the potential problems of using geostatistical techniques in connection with socio-economical data aggregated by areas, especially with the fact that ordinary kriging uses an implicit hypothesis of normal distribution which may not be supported by data sets such as homicide rates. We consider that emerging techniques such as "model-based geostatistics" (Diggle, Moyeed et al., 1998) should be used in preference to ordinary kriging. Unfortunately, such techniques are not widely available in connection with GIS packages, a situation we hope will be solved in the near future. Additionally, given the presence of local instabilities, as shown in earlier Sections of this work, the production of trend surfaces by geostatistical techniques should be taken in context. Such data sets should be analyzed in connection with the original data and are useful as overall trend indicators.

8 Conclusions

This paper has described the use of spatial analytical techniques in connection with social exclusion studies for the city of São Paulo in the 90's. These results are being used as part of a major effort in understanding intraurban dynamics in Brazil's largest city. We have shown different examples of spatial analysis methods, including: local autocorrelation indexes, spatial autocorrelation maps as a basis for spatial clustering, spatial regression analysis and trend surfaces derived from geostatistical interpolation. Our results show that the use of spatial analysis can provide social scientists with tools that enable significant insights for understanding urban dynamics.

The data sets used in this work are available on the Internet, on the webpage http://www.dpi.inpe.br/geopro/exclusao. The software used in this work was *SPRING*, a free GIS available on the Internet (www.dpi.inpe.br/spring) developed by INPE (Câmara, Souza et al., 1996) and *SpaceStat*, a spatial analytical software developed by Luc Anselin (Anselin, 1992).

APPENDIX A

TABLE 1

COMPOSITION OF THE SOCIAL EXCLUSION/INCLUSION INDEX

CATEGORIES	COMPOSED INDICATORS	VARIABLES – 1995 MAP	REFERENCE VALUES
Autonomy Index	IEX – POOR FAMILY SURVIVAL CONDITIONS	Family heads below the poverty limit (without income)	0%
	IEX – INCOME AUTONOMY	Income per family head	3-5 minimum wage
		Job Offer	0,55
	IEX – STREET POPULATION	Adult poverty rate	0%
		Children at risk rate	0%
Quality of	IEX – ENVIRONMENT AL QUALITY	Houses with poor water service	0,5%
		Houses with poor sewer service	0,5%
Life Index		Houses with poor garbage collection	0,3%
	IEX – SANITATION COMFORT	Habitation density	4 persons/house
		Bathroom/house offer	1 bathroom/house
		Bathroom/person density	3 persons/bathroom
	IEX – PRIVACY COMFORT	Bedroom/house	2 bedrooms/house
		Bedroom/person density	2 persons/room
	IEX – POOR HOUSING	Percentage of population who lives in poor housing	0,5%
	IEX – TIME TO WORK	Average time spent to work	56 minutes
	IEX – SOCIAL SERVICES DEFICIT	Basic health services access potential	40% access
		Crèche access potential	40% of children in
		Kindergarten education access potential	100% access
		First level access potential	100% access
Human Development Index	IEX – POOR LITERACY	Illiterate family heads	0% illiteracy rate

	IEX - EDUCATIONAL DEVELOPMENT	Years of education of family head	8 years of education
	IEX – DEATH RISK	Percentage of population over 70	3%
		Children mortality	25 per 1,000 births
		Youth mortality	3,76 per 100,000
		Potential of lost life years	43
	IEX – VIOLENCE	Larceny cases	0 cases
		Robbery cases	0 cases
		Vehicle robbery cases	0 cases
		Homicide cases	0 cases
Equality Index		Concentration of women as family heads	2%
		Concentration of illiterate women as family heads	0,4%

References

Anselin, L. (1988). Spatial econometrics: methods and models. Dordrecht, Kluwer.

- Anselin, L. (1992). SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data. Santa Barbara, NCGIA (National Center for Geographic Information and Analysis).
- Anselin, L. (1995). "Local indicators of spatial association LISA." <u>Geographical Analysis</u>
 27: 91-115.
- Anselin, L. (1996). The Moran scatterplot as ESDA tool to assess local instability in spatial association. In: <u>Spatial Analytical Perspectives on GIS</u>. M. Fisher, H. J. Scholten and D. Unwin (ed). London, Taylor & Francis: 111-126.
- Bailey, T. and A. Gattrel (1995). Spatial Data Analysis by Example. London, Longman.
- Bessis, S. (1995). From social exclusion to social cohesion :towards a policy agenda. Paris, UNESCO Management of Social Transformations Policy Paper -2.

- Bivand, R. (1998). A review of spatial statistical techniques for location studies, Norwegian School of Economics and Business Administration. http://www.nhh.no/geo/gib/gib1998/gib98-3/lund.html.
- Blair, A. (1998). Bringing Britain together: a national strategy for neighbourhood renewal. London, UK Cabinet Office Social Exclusion Unit, CM report 4045.
- Câmara, G., R. Souza, U. Freitas and J. Garrido (1996). "SPRING: Integrating Remote Sensing and GIS with Object-Oriented Data Modelling." <u>Computers and Graphics</u> 15(6): 13-22.
- Diggle, P. J., R. A. Moyeed and J. A. Tawn (1998). "Model-based geostatistics." <u>Applied</u> <u>Statistics</u> **47**: 299-350.
- Getis, A. and J. K. Ord (1996). Local spatial statistics: an overview. In: <u>Spatial Analysis:</u> <u>Modelling in a GIS Environment</u>. P. Longley and M. Batty (ed). New York, John Wiley: 261-277.
- Goodchild, M. (1988). "A spatial analytic perspective on geographical information systems." <u>International Journal of Geographical Information Systems</u> 1: 327-334.
- Jolly, R. (1999). Human Development Report. New York, United Nations Development Program (UNDP).
- Kingsley, G. T., C. J. Coulton, M. Barndt, D. S. Sawicki and P. Tatian (1997). Mapping Your Community: Using Geographic Information to Strengthen Community Initiatives. Washington, D.C, U.S. Department of Housing and Urban Development.

Sposati, A. (1996). Mapa de Exclusão/Inclusão Social de São Paulo. São Paulo, EDUC.

Sposati, A. (2000). Map of Social Exclusion/Inclusion for São Paulo - 2000. Social Dynamics of the 1990's. São Paulo, Catholic University.