"Mapping Social Exclusion and Inclusion in Developing Countries: Spatial Patterns of São Paulo in the 1990s"

GILBERTO CÂMARA[1], ALDAIZA SPOSATI[2], DIRCE KOGA[2], ANTONIO MIGUEL MONTEIRO[1], FREDERICO ROMAN RAMOS[1], EDUARDO CAMARGO[1], SUZANA DRUCK FUKS[3]

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

Av dos Astronautas 1758, São José dos Campos, Brazil

[2]Center for Research on Social Security and Social Assistance,

Catholic University of São Paulo (PUC/SP), Rua Monte Alegre, 984, São Paulo, Brazil

[3]Brazilian Agricultural Research Agency (EMBRAPA),

Rodovia Brasília-Fortaleza, BR 020, Km 18, Planaltina, Brazil

1 Introduction

The concept of social exclusion was born in Europe, motivated by the sharp increase in the number of poor, whose numbers in the 12 countries of the EEC went from 38 million in 1975 to 53 million in 1992. First conceived in the 1960s in France (Klanfer 1965), in the last two decades social exclusion has become a major category of social thinking (Duffy 1995; Paugam 1991). Poverty implies exclusion from goods and services; social exclusion goes beyond income inequalities, to encompass denial or non-realization of civil, political, and social rights of citizenship (Room 1995). Social exclusion is therefore linked to the approach proposed by Nobel Laureate Amartya Sen (1992), which considers that removal of inequalities in modern societies is determined by the access to basic capabilities such as the ability to be healthy, well-fed, housed, integrated into the community, participate in community and public life, and enjoy social bases of self-respect.

Social exclusion has become a major policy issue in both developing and developed nations. In Europe, the United Kingdom government has established a "Social Exclusion Unit" which was set up by the Prime Minister in 1997, and the UK government has produced the "indices of deprivation" for the UK (DETR, 2000). The Council of Europe carried out a project on "Human Dignity and Social Exclusion", 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 et al. 1997).

Since most of the socioeconomic data used in social exclusion studies are associated to geographical locations and aggregated into areas such as census tracts or boroughs, maps are a natural way of portraying local social and economical conditions. In this way, the perception of social exclusion is enhanced with the visualization of "problem-areas" which, more often that not, tend to exhibit some type of cluster patterns. Notwithstanding the impact of maps in promoting awareness about deprivation, the role of space in the patterns of social exclusion indicators has received considerably less attention. In most cities and countries, the social and economical phenomena that cause social exclusion will tend to be space-related, since the very nature of wealth concentration in most modern societies tends to produce large inequalities in land allocation, and land and house prices. Therefore, it is very relevant to inquire (following Bailey 2001): *Are the patterns of social exclusion conditioned by factors that are spatially dependent?* Or to put it directly, *why are these spatial patterns there, and how will they change if we intervene in a particular way?* Answering these questions requires using statistical methods that are spatially explicit; it also requires handling data sources at different levels of spatial aggregation and using different ways of depicting spatial information.

Given this motivation, this work examines the use of spatial analytical techniques to explore the patterns of social exclusion. The basis for the work is a set of indices of social exclusion/inclusion for the Brazilian city of São Paulo, which have had with a major impact on policy makers and the public (Sposati 1996; Sposati 2000). Taking the maps of social exclusion/inclusion as a basis, we have set out to address the question: *Can the explicit use* *of space in our analytical techniques enhance our comprehension of social exclusion in São Paulo?* This general concern was broken down in the following issues:

(a) *Are the patterns of social exclusion spatially dependent? Or are there "pockets" of local variation where social exclusion/inclusion differs significantly from the overall trends in the city?* To address these questions, we have used global and local spatial autocorrelation indices to explore the properties of the social exclusion patterns in São Paulo, and to identify clusters of social exclusion and social inclusion in São Paulo.

(b) *Given the many dimensions and components of the social exclusion/inclusion indices, what is the relative influence of the individual components to the overall social exclusion? Is there a single factor that is highly correlated to the overall pattern of social exclusion? Is this correlation spatially dependent?* This is a very relevant question for public policy, since its helps to ascertain and direct the use of public funds. To address this question, we have used spatial regression techniques, which shows that the conditioning factors of social exclusion vary considerably within the city.

(c) How can spatio-temporal trends in the components of social exclusion best be portrayed? Can we gain insight into the spatio-temporal trends of social exclusion by freeing ourselves from "the tyranny of zones" (Spiekermann and Wegener 2000)? To address this question, we have mapped spatio-temporal trends in the evolution of crime in São Paulo using geostatistics.

Each of these questions has been addressed with the use of spatial analysis techniques, as described in the next sections. These examples show how spatial analytical techniques can substantially enhance the understanding of social exclusion and inclusion patterns in large cities of the developing world.

2 Measuring Social Exclusion/Inclusion in São Paulo

The development of indicators for the measurement of deprivation and social exclusion has been subject to intense debate recently (for reviews of the problem, for example, Gordon and Townsend 2000 and Senior 2002). Most approaches to the calculation of social exclusion indices are based on the premise that social exclusion is made up of separate dimensions, or 'domains' of deprivation, where each domain is made up of a number of components that cover aspects of social exclusion as comprehensively as possible. For example, the UK's deprivation index (DETR, 2000) is made of six dimensions, calculated for each ward: (a) income deprivation, (b) employment deprivation, (c) health deprivation and disability, (d) education, skills and training deprivation, (e) housing deprivation, and (f) geographical access to services. Each index is computed and the different areas are ranked. Then, a normalization procedure is applied for expressing each dimension in the same scale, and all indexes are transformed, to ensure that each domain is transformed to a common distribution. Finally, the indexes are added up according to a weighting rule that reflects the relative importance of each factor (DETR, 2000).

The basis for the UK's and similar deprivation indicators is the implicit assumption that "social exclusion" tends to be associated to processes of *social disqualification* and from economical and social problems that impact urban areas, many of which 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). This assumption of social disqualification is implicit in the methodological procedures applied: for each 'dimension' of the deprivation index, wards are *ranked* and then *normalized.* The implicit assumption is that the average values for the country represent an "acceptable" level of social inclusion and that social policies should be geared towards "regeneration" of areas that have the lowest deprivation indices.

In lower income countries (LIC), social exclusion has a completely different setting. Their deprived populations have never had acceptable living conditions and there is no social protection typical of the Welfare State societies of the 20th century. In these countries, social exclusion is the result of *social apartheid* and of strong inequalities in income distribution. Therefore, measurement of social exclusion in the developing world requires a different strategy than indicators such as UK's Indices of Deprivation, which are based on *ranking* and do not provide an objective measure of whether the citizens of an area have achieved an acceptable standard of living. In LICs, social exclusion indices have to consider that the average values of dimensions such as income, health deprivation and housing quality may represent unsatisfactory living conditions. Therefore, social exclusion indices in LICs have to convey not only the *relative* position of an area (e.g., an electoral ward) in relation to a study area (e.g., a country), but also the *absolute* situation of this area in relation to the attainment of acceptable living conditions. The social exclusion/inclusion index developed by the authors starts by defining *a basic living standard*, which includes needs that are considered basic and universal according to a collective ethic of life and that incorporate attainable goals for public policy (Sposati 1996).

Our proposed *social exclusion/inclusion index* is aggregated by areal units, with four components: income, quality of life, human development, and gender equality. Each component is captured by a set of variables obtained from census and field data collection, described in Table 1. For each variable, we propose a *reference value* that marks 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 linearly to a positive [0..1] scale, and areas below such reference are assigned negative values on a [-1...0] scale. Therefore, each of the components has a range between –1 (total exclusion) and 1 (total inclusion). The social exclusion index is obtained by averaging its four components.

TABLE 1

COMPOSITION OF THE SOCIAL EXCLUSION/INCLUSION INDEX

INDEX	SUBCATEGORY	CENSUS VARIABLES	REFERENCE VALUES
Income Index	POOR FAMILY SURVIVAL CONDITIONS	Family heads below the poverty limit (without income)	0 percent
	INCOME AUTONOMY	Income per family head	3-5 minimum wage
		Job Offer	0,55
	STREET POPULATION	Adult poverty rate	0 percent
		Children at risk rate	0 percent
Quality of Life Index	ENVIRONMENTAL QUALITY	Houses with poor water service Houses with poor sewer service Houses with poor garbage collection	0.5 percent 0.5 percent 0.3 percent
	SANITATION COMFORT	Habitation density Bathroom/house offer Bathroom/person density	4 persons/house 1 bathroom/house 3 persons/bathroom
	PRIVACY COMFORT	Bedroom/house Bedroom/person density	2 bedrooms/house 2 persons/room
	POOR HOUSING	Percentage of population who lives in poor housing	0.5 percent
	TIME TO WORK	Average time spent to work	56 minutes
	SOCIAL SERVICES DEFICIT	Basic health services access potential Crèche access potential Kindergarten education access potential First level access potential	40 percent access 40 percent of children in creches 100 percent access 100 percent access

INDEX	SUBCATEGORY	CENSUS VARIABLES	REFERENCE VALUES
Human Develop. Index	POOR LITERACY	Illiterate family heads	0 percent illiteracy rate
	EDUCATIONAL DEVELOPMENT	Years of education of family head	8 years of education
	DEATH RISK	Percentage of population over 70 Children mortality Youth mortality Potential of lost life years	3 percent 25 per 1,000 births 3.76 per 100,000 43
	VIOLENCE	Larceny cases Robbery cases Vehicle robbery cases Homicide cases	0 cases 0 cases 0 cases 0 cases
Gender Equality Index		Concentration of women as family heads	2 percent
		Concentration of illiterate women as family heads	0.4 percent

This methodology was used to assess the evolution of the city of São Paulo during the 1990s. 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 that includes the largest number 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).

We have produced the "Map of Social Exclusion/Inclusion of São Paulo - 1995," which used data available from 1987 to 1995 to produce indicators for the earlier part of the 1990s (Sposati 1996), and the "Map of Social Exclusion/Inclusion of São Paulo - 2000," which concentrated on trends on population, employment and quality of life indicators during the 1990s (Sposati 2000). The 1995 map showed a significant gap between the social exclusion and the social inclusion regions of São Paulo, where 2/3 of its districts are below acceptable levels of living standards, as depicted in Figure 1.

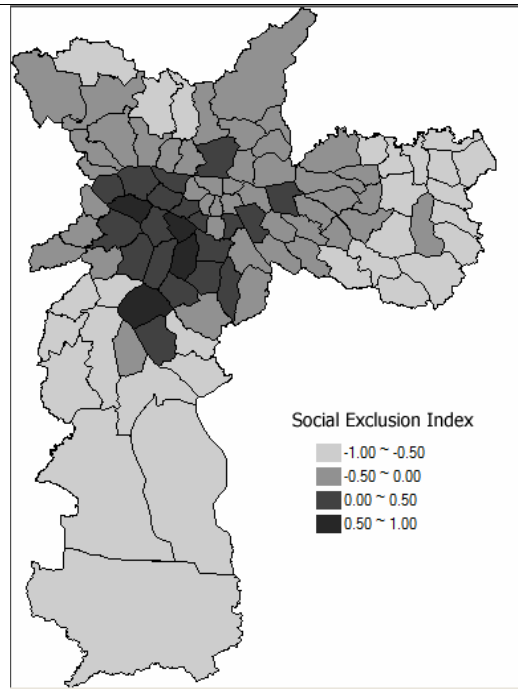

Figure 1 – Social Exclusion Index in São Paulo obtained by the 1995 Social Exclusion/Inclusion Map (96 districts). Values vary from -1 (extreme exclusion) to +1 (extreme inclusion).

The main findings of the 2000 Map were a significant change in population dynamics and a strong relation between education and unemployment trends. Although has been only a small increase of the overall population from 1991 to 1996 (from 9,646,185 inhabitants to 9,839,066 inhabitants or a 2 percent-growth), the poorest regions of the city have registered population increases up to 130 percent. 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 consequence has been a large increase in violence and homicide rates, 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 1990s.

3 Exploring the Patterns of Social Exclusion/Inclusion for Special Dependence

Taking the "Map of Social Exclusion/Inclusion of São Paulo" as our basic data set, we have set out to explore a number of questions regarding the rôle of space in the social exclusion patterns. Our first concern was to address the nature of spatial dependence in these patterns. As a starting point, it was necessary to determine the possible existence of regional trends in the data; that estimation is necessary, since if the data would exhibit a trend, it would cause the indices to be naturally spatially autocorrelated. The regression fit for the trend surface was very poor ($R^2 = 0.12$), to the extent that it can be inferred that there are no strong spatial trends in the data. The next step was to estimate the global spatial autocorrelation in the social exclusion index, by using Moran's I index:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1}$$

In this equation, n is the number of areas, x_i is the value of the attribute in area i, $\bar{x}$ is the mean value of attribute for the whole region, and the weights w_{ij} are such that they are $1/neigh_i$ if area i and area j are contiguous and 0 (zero) otherwise, where $neigh_i$ is the number of neighbors of area i. The Moran index is a global correlation coefficient, where a value of 0 (zero) would indicate no spatial correlation and a value of 1 (one) a complete spatial dependency. For the composite social exclusion/inclusion index, we have obtained a Moran's I index of 0.642. In order to test the significance of this index, we have used a permutation test, where the value of the attributes associated to the regions are shuffled in random, generating 999 new spatial arrangements (Anselin 1992). The Moran index associated to each new spatial arrangement is computed, producing an empirical distribution of 1000 values, from which we can derive that the value obtained is significant within a 99% confidence interval.

Given the existence of a strong global pattern of spatial association for the social exclusion/inclusion index in São Paulo, the next question to be asked concerns the regional distribution of this index: *Are there "pockets" of local variation where social exclusion/inclusion differs significantly from the overall trends in the city?* The idea is to find clusters of local variation where the social exclusion/inclusion index has a stronger association than the overall trends in the city. To address these questions, we have used two exploratory data analysis tools: the Moran scatterplot and the Local Moran spatial autocorrelation index.

The *Moran scatterplot* (Anselin 1996) is a tool for visualization of the patterns of spatial autocorrelation. The idea is to compare the spatial distributions of an *attribute* and of its *local mean*. As a first step, both variables are *normalized*, subtracting its values from the global mean and dividing by the standard deviation. The resulting normalized variables will have a mean of 0 (zero) and a standard deviation of 1 (one). Following Anselin (1996), we refer to the normalized variable as **Z** and to its local mean by **WZ**, where **W** is the normalized weights matrix, as described in equation (1). By constructing a graph of **Z** versus **WZ** (Figure 2), we can express four different types of spatial association:

- Quadrant Q1 ("High-High"), that shows areas whose both its normalized values and its local mean values are positive;
- Quadrant Q2 ("Low-Low"), that shows areas whose both its normalized values and its local mean values are negative,
- Quadrant Q3 ("High-Low"), with positive values and negative local means,

Quadrant Q4 ("Low-High"), with negative values and positive local means,

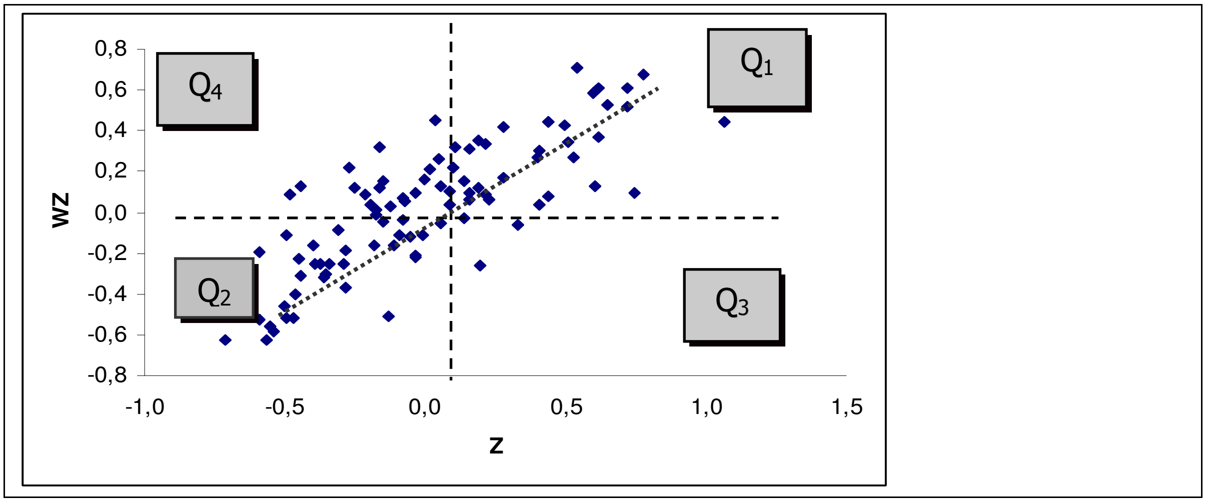

Figure 2 – Moran Scatterplot for the social exclusion/inclusion index of São Paulo, for the 1995 data set.

The interpretation of the Moran scatterplot in this context is that most of the districts of the city are located in quadrants Q1 and Q2, which are areas of positive spatial association. However, a significant number of districts are located in quadrants Q3 and Q4, and can be considered as areas that do not follow the same pattern of spatial association. These districts can therefore be considered as transition regions between regions of social inclusion (the "high-high" districts) and regions of social exclusion (the "low-low" areas). To further investigate this hypothesis, we have created a map, shown in Figure 3, in which each district is labeled according to the quadrant occupied by its social exclusion/inclusion index in its Moran scatterplot. Such a map is called a "Moran map" (Anselin 1996), and shows that Q1 ("high-high") districts are all located in the center of São Paulo, and Q2 ("low-low") districts are mostly located in the Eastern and Southern parts of the city, which are the regions of greater social exclusion. Districts associated to quadrants Q3 ("high-low") and Q4 ("low-high") are mostly associated to intermediary regions between the city's center and its two great regions of social exclusion.

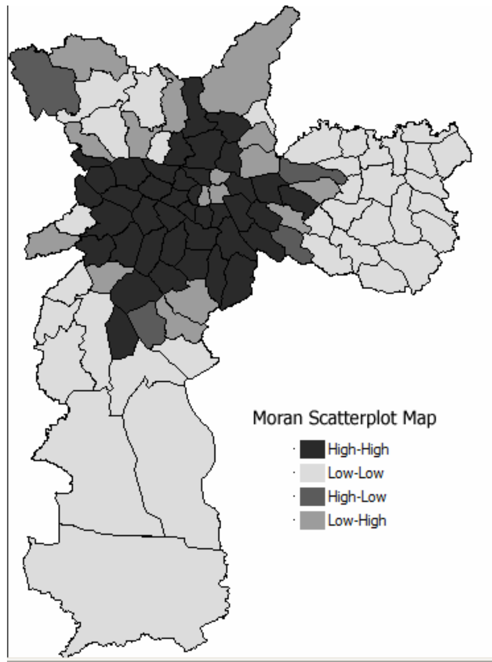

Figure 3 – Moran scatterplot map for Social Exclusion/Inclusion Index of the "Map of Social Exclusion/Inclusion of São Paulo – 1995."

The Moran map indicates strong patterns of spatial association, and therefore suggests the presence of clusters on the distribution of the social exclusion/inclusion index. To find such clusters, we used the *local Moran index* (Anselin 1995):

$$I_i = \frac{(x_i - \bar{x})\sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{2}$$

where the terms are defined as in equation (1). To establish the significance of the local Moran index, we simulated a pseudo-distribution by permutation of the attribute values among the areas; statistical tests were then used to establish confidence intervals. Local index values with significance of 95 percent, 99 percent and 99.9 percent were then mapped and posited as 'hot-spots' of local non-stationarity (Anselin 1995). We found two 'hot spots' of social exclusion, located in the South and East of the city, and one 'hot spot" of social inclusion located in the Center of the city, as shown in Figure 4.

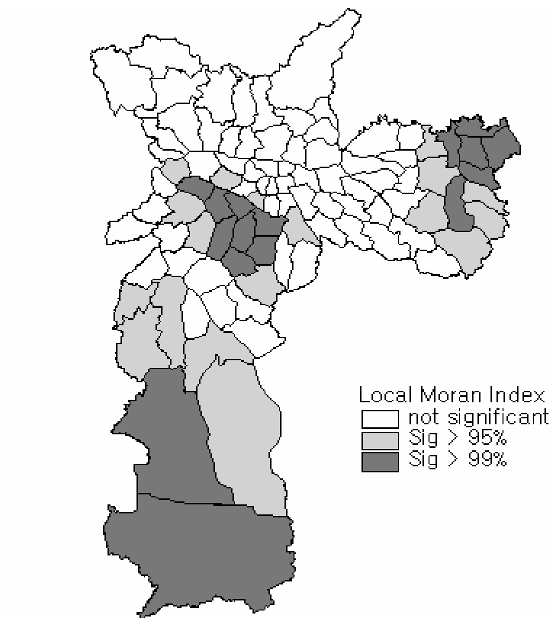

Figure 4 – Significant values of local Moran index for social exclusion/inclusion index for São Paulo, for the 1995 data set.

The clusters in Figure 4 correspond to areas that concentrate a significant amount of the city's disparities: (a) the so-called "Deep South" and the "Far East" regions of São Paulo, areas of high social exclusion; and (b) the center of the city, an area of high social inclusion. The two most excluded regions of the city differ in the causes for such exclusion. The city's "Deep South" region has had an explosive growth in recent times. Migrant workers have come to São Paulo from other parts of the country and have occupied the region, resulting in population growth rates of over 100 percent in most areas. This process has not been matched by public investment, and as a result its inhabitants have the worst conditions of the city in terms of public services (health, education and social care). In the "Far East" of São Paulo, concentration of low-income population is a direct consequence of public policies of the 1970s and 1980s, which removed poor people from slums located in the central (and wealthiest) part of the city towards publicly-built housing estates in the eastern periphery. These housing estates were inadequately built and rapidly degraded into crime-infested areas. By contrast, in the center of São Paulo, the map shows a significant cluster of high-income areas, where the wealthiest part of the population lives. Consequently, the exploratory data analysis tools have proven effective in distinguishing the extreme concentrations of wealth and poverty in São Paulo.

4 Analysis of Social Exclusion Factors by Spatial Econometrics

The investigation of social exclusion process in São Paulo also requires an assessment of 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 in developing nations. This raises an important question: *what is the minimum set of variables, which can still produce a credible result for the social exclusion/inclusion index? Is there a variable that is a determinant factor for social exclusion?* To establish a relation between the relevant factors and the composite indices, we investigated three different types of spatial regression models: the *spatial autoregressive lag* model, *the spatial autoregressive error* model, and the *spatial regimes* regression. Each of these models is briefly described, following Anselin (1988). The linear regression model formulation can be described as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \varepsilon,\ \varepsilon \sim N(0,\sigma^2), \text{ or} \tag{3}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ .. \\ .. \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & .. & x_{1k-1} \\ 1 & x_{21} & .. & x_{2k-1} \\ .. & .. & .. & .. \\ .. & .. & .. & .. \\ 1 & x_{n1} & .. & x_{nk-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ .. \\ .. \\ \beta_{k-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ .. \\ .. \\ \varepsilon_n \end{bmatrix} \tag{4}$$

where $\boldsymbol{Y}$ is an *(n x 1)* vector of observations on a dependent variable taken at each of n locations, $\boldsymbol{X}$ is an *(n x k)* matrix of exogenous variables, $\boldsymbol{\beta}$ is an *(k x 1)* vector of parameters, and **ε** is *(n x 1)* an vector of disturbances. The *spatial lag model* includes a spatial dependence term, through a new term that incorporates the spatial autocorrelation as part of the explanatory component of the model:

$$\boldsymbol{Y} = \rho \boldsymbol{WY} + \boldsymbol{X\beta} + \varepsilon \tag{5}$$

where $\boldsymbol{W}$ is the spatial weights matrix, as described in equation (1), and the product $\boldsymbol{WY}$ expresses the spatial dependence on $\boldsymbol{Y}$, where ρ is the *spatial autoregressive coefficient*. The *spatial autoregressive error model* considers that the spatial effects are a perturbation that should be removed. In this case, the spatial effects are associated to the error term and the model is expressed as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \varepsilon, \qquad \varepsilon = \lambda \boldsymbol{W} + \xi \tag{6}$$

where λ is a scalar spatial error parameter, and ε 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. These global spatial regression models are based on the hypothesis that the spatial process whose observations are being analysed is stationary. This implies that the spatial autocorrelation patterns can be captured in a single regression term. In practice, medium to large sized data sets, such as the São Paulo ones, exhibit different patterns of spatial dependence, as seen in the previous section, and suggest the use of regression techniques suitable for non-stationary spatial processes, whose regression coefficients must reflect the spatial heterogeneity. One possibility is modelling the spatial trends in a discrete fashion, subdividing the study region into in sub-regions, called *spatial regimes* (Anselin 1988). Each spatial regime is posited to have its own spatial pattern, and its own coefficients. For example, consider the case where the study region is divided in two subregions:

$$\boldsymbol{Y} = \boldsymbol{X_1\beta_1} + \boldsymbol{X_2\beta_2} + \varepsilon \tag{7}$$

In this model, the non-zero values of vector $\boldsymbol{X_1}$ are only those values of $\boldsymbol{X}$ for areas that are within the first spatial regime, as expressed in:

$$\begin{bmatrix} y_1 \\ y_2 \\ .. \\ .. \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & .. & x_{2k-1} \\ 1 & x_{21} & .. & x_{2k-1} \\ .. & .. & .. & .. \\ .. & .. & .. & .. \\ 0 & 0 & .. & 0 \end{bmatrix} \begin{bmatrix} \beta_0^1 \\ \beta_1^1 \\ .. \\ .. \\ \beta_{k-1}^1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ .. & .. & .. & .. \\ .. & .. & .. & x_{n-1,k-1} \\ 1 & x_{n1} & .. & x_{n,k-1} \end{bmatrix} \begin{bmatrix} \beta_0^2 \\ \beta_1^2 \\ .. \\ .. \\ \beta_{k-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ .. \\ .. \\ \varepsilon_n \end{bmatrix} \qquad \textbf{(8)}$$

As an initial approximation to the spatial regimes for the São Paulo data set, we used the exploratory techniques described in the previous section, and divided the city into three distinct regions, as shown in Figure 5: the city's centre, which is a locus of social inclusion, the Southern and Eastern parts of the city (extremes of social inclusion), and a transition region, which contains most areas that do not follow the main spatial dependence trend, and which fall into quadrants Q3 and Q4 of the Moran map (see Figure 3).

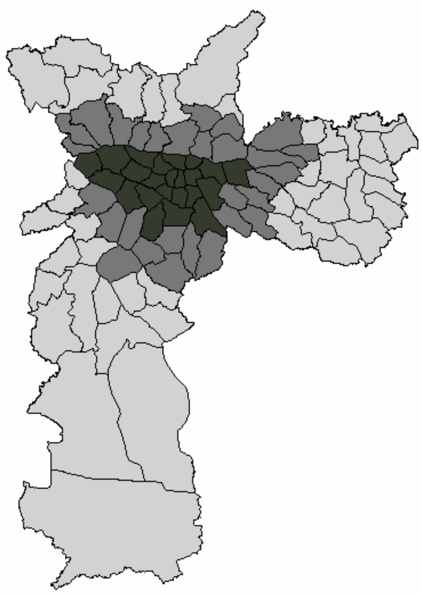

Figure 5 – Spatial regimes for the city of São Paulo

We performed a regression analysis on the correlation between the percentage of family heads[1] with more than 15 years of schooling (as the independent variable) with the social exclusion index (as the dependent variable). Four regression techniques were used: standard OLS regression, spatial lag model, spatial error model and the spatial regimes, as shown in Figure 5. The results are summarized in Table 1, based on four comparison criteria: (a) R^2 - the standard "goodness-of-fit" measure, which is inadequate for spatially dependent data, as discussed in Anselin (1988); (b) *Likelihood* – maximized log-likelihood assessment of model fit, a preferred measure according to Anselin (1992); (c) *MI-error* – global spatial autocorrelation indicator of residuals; (d) *LM-error* – Lagrange Multiplier indicator (assesses the extent to which there remains spatial autocorrelation in the residuals), a test proposed by Anselin (1992).

TABLE 2

EDUCATION X SOCIAL EXCLUSION IN SÃO PAULO

(RESULTS FROM REGRESSION MODELS)

	OLS	**Spatial Lag**	**Spatial Regimes**
R^2	0.75	0.77	0.86
Likelihood	14.9	20.53	47.86
MI-error	0.384	-	-0.007
LM-error	29.43	12.19	0.006

The spatial regimes regression was significantly superior to the other regression models; it had a better fit to the observed data, and the model residuals exhibited

[1] In the Brazilian census, a "family head" is the person of the family who is responsible for bringing the most income into the family (usually, but not always the father).

significantly less spatial autocorrelation. This performance is explained by the existence of different regimes of spatial association for the social exclusion/inclusion index in São Paulo. Since the spatial error and the spatial lag regression techniques only model the global autocorrelation patterns, they fail to account for local instabilities.

The practical implication of this regression study is that a significant proportion of the social exclusion/inclusion index can be related to the education of the family head, when the spatial dependence is taken into account. In fact, similar regressions indicated that, out of all 45 basic variables, education has the strongest relation to the social exclusion/inclusion index. In São Paulo's poor regions, our results indicate that a small increase in years of schooling will most often be translated into a substantial improvement in social inclusion.

5 Trend Surfaces of Homicide Rates in São Paulo

Spatial data models for socioeconomic phenomena usually involve aggregation of census-type data over area units. The boundaries of such areas are defined by operational or political criteria and therefore essentially unrelated to the phenomena being modeled (Martin 1996). This fact leads to idea of dissolving zonal data into continuous surfaces; these surfaces provide a useful framework for the spatial analysis of socioeconomic data. For the social scientist, the removal of the boundaries is not as important as the information gained by having a distribution that depicts the major trends of a variable (or a set of variables) over the entire study area (Goodchild, Anselin, and Deichmann 1993; Spiekermann and Wegener 2000).

We used geostatistical techniques to produce trend surfaces for the homicide rates in São Paulo in 1996 and 1999, as shown in Figure 6 The original data consisted of estimates of homicide rate per 100,000 inhabitants, aggregated by the 96 districts. To produce these maps, we obtained a sample set by assigning a sample at the center of each district. These samples were then used as a basis for computing a variogram that models the spatial correlation structure, and a surface was interpolated by ordinary kriging (Bailey and Gatrell 1995). The trend surfaces depict a significant decrease in the areas with lowest homicide rate (below 30 deaths per 100,000 inhabitants) in 1999 in relation to 1996. Since the lower homicide rate correlates very strongly with the wealthiest regions of the city (compare with Figures 1 and 2), these results indicate a spatial spreading of crime. Violence is thus not confined to the poorest areas of the city and the inhabitants of richer areas are increasingly prone to be victims of violent assaults. These results have had a major impact on public awareness of the spatial trends of crime in the city; *Folha de São Paulo*, Brazil's largest newspaper, ran a major story with Figure 6in its front page.

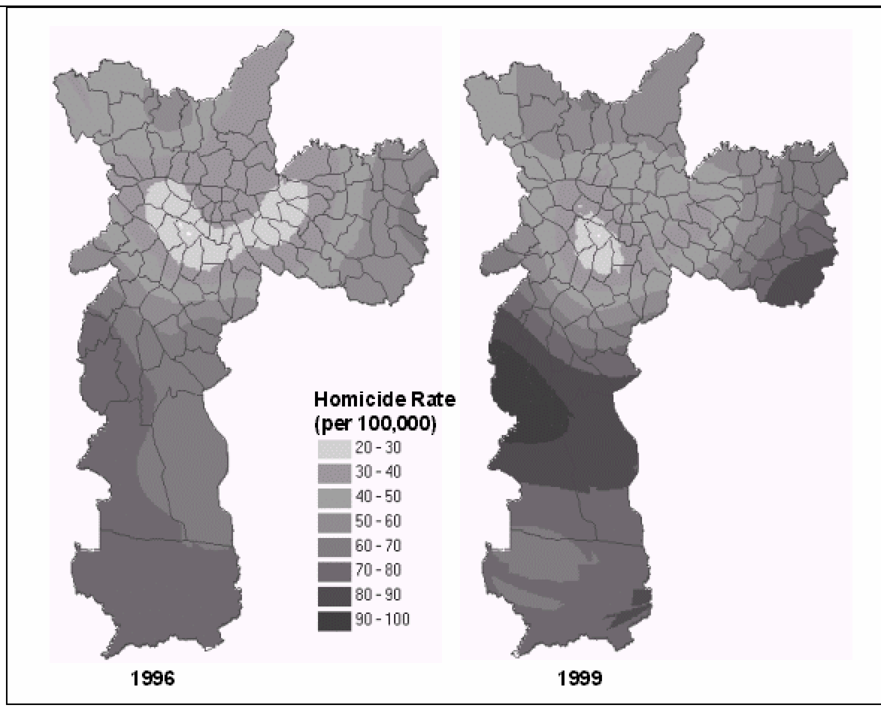

Figure 6– Trend surfaces for homicide rates in São Paulo, with values shown as homicides per 100,000 population. Left: 1996 data Right: 1999 data.

Despite its usefulness for estimating trend surfaces, there is a potential problem of using ordinary kriging techniques in connection with this type of socioeconomic data. Ordinary kriging relies on the normal distribution hypothesis, which is not the most appropriate assumption for data sets such as homicide rates. Given the characteristics of homicides as rare events, a more appropriate assumption would be a Poisson or a binomial probability distribution (Bailey and Gatrell 1995). One alternative to apply kriging procedures to this kind of phenomenon would be to use the corrections proposed by McNeil (1991) and Oliver et al (1992). These corrections estimate the "risk" of homicide, expressed as the probability that an individual in an area be killed by homicide. Another alternative would be to use "model-based geostatistics", a set of kriging estimators for distributions other than the normal (Diggle, Moyeed, and Tawn 1998). Unfortunately, such techniques are not widely available in connection with GIS packages, a situation we hope will be solved in the near future.

6 Conclusions and Future Work

What can be gained in social exclusion studies by using spatial analytical techniques? A lot, as it turns out. Exploratory techniques such as global and local Moran indexes and Moran scatterplot maps are very useful to indicate the existence of global trends of spatial autocorrelation for the social exclusion patterns, and also to point out regions where this trend was significantly weaker or stronger. Spatial regression analysis (especially the spatial regimes technique) enables measuring the relation between the various phenomena that comprise social exclusion, and can help establish how the relationships between the components of social exclusion and the combined indices vary in space. Finally, by freeing ourselves from the "tyranny of zones", surfaces provide a powerful means of apprehension of spatial variation.

Further work being carried out by the authors concerns one specific limitation of our data set: the use of 96 districts, which represent an aggregated perspective of most areas. In countries with great social contrasts such as Brazil, it is frequent that different social groups are aggregated in the same administrative areas, resulting in indices that can misrepresent the diversity of these populations. The reason for the use of the district-level data was its availability at this spatial level of aggregation, for the 1991 and 1995 data sets. Detailed data from the 2000 Census is currently being made available to the authors, who will refine and review the indices for São Paulo, using data at the census tract level.

The work related to the “Map of Social Exclusion/Inclusion of São Paulo” has had a large impact on increasing political and academic awareness of the issue of social exclusion in São Paulo. Many researchers and public policy administrators have been using its results; the “Social Exclusion/Inclusion” maps have appeared in more than 50 news articles since 1995. The current mayor of São Paulo is using its results to subsidize public policy and government investment in the city, and one of the authors (Sposati) has been appointed as Social Services Secretary of São Paulo, during the 2000-2004 period. The practical impact of the work only enhances the benefits than can be gained by spatially aware social science research.

Acknowledgments

The results presented in this paper have mostly been financed by FAPESP (Fundação de Amparo à Pesquisa em São Paulo), by grant 00/01965-0. The research of Gilberto Câmara and Antônio Miguel Monteiro is also supported by CNPq (National Research Council of Brazil - award number CNPq 480322/99) and by Brazil’s MCT/PCTGE (Ministry of Science and Technology—Program for Science and Technology on Ecosystems Management). 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 et al. 1996) and *SpaceStat*, a spatial analytical software developed by Luc Anselin (Anselin 1992).

References

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

———. 1992. SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data. Santa Barbara: NCGIA (National Center for Geographic Information and Analysis).

———. 1995. Local indicators of spatial association - LISA. *Geographical Analysis* 27:91-115.

———. 1996. The Moran scatterplot as ESDA tool to assess local instability in spatial association. In *Spatial Analytical Perspectives on GIS*, edited by M. Fisher, H. J. Scholten and D. Unwin. London: Taylor & Francis, pages 25-32.

Bailey, T., and A Gatrell. 1995. *Spatial Data Analysis by Example*. London: Longman.

Bailey, T. 2001. Debate on paper by Câmara,G. and Monteiro, A.M.V., Geocomputation Techniques for Spatial Analysis: Are They Relevant to Health Data?. *Cadernos de Saúde Pública*, 17(5):1059-1081, set/out. 2001.

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. *Computers and Graphics* 15 (6):13-22.

Cressie, N. 1991. *Statistics for Spatial Data*. New York: Wiley.

DETR. 2000. Indices of Deprivation 2000. *UK Department of Environment, Transport and Regions Research Summary 31*. DETR, London.

Diggle, P., R. Moyeed, and J. Tawn. 1998. Model-based geostatistics. *Applied Statistics* 47:299-350.

Duffy, K. 1995. *Social Exclusion and Human Dignity in Europe*. Strasbourg: Council of Europe.

Fotheringham, A.S., C. Brunsdon and M.E. Charlton. 2000. *Quantitative Geography*. London: Sage.

Goodchild, M. 1988. A spatial analytic perspective on geographical information systems. *International Journal of Geographical Information Systems* 1:327-334.

Goodchild, M., L. Anselin, and U. Deichmann. 1993. A Framework for the Areal Interpolation of Socioeconomic Data. *Environment and Planning A* 25:383-97.

Gordon, D. and Townsend, P. (eds). 2000. *Breadline Europe: The measurement of poverty*. Bristol: The Policy Press.

Jolly, R. 1999. Human Development Report. New York: United Nations Development Program (UNDP).

Kingsley, 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.

Klanfer, J. 1965. *L'Exclusion sociale. Étude de la marginalité dans les sociétés occidentales (The social exclusion. Study on the marginality on the western societies)*. Paris: Bureau de Recherches Sociales.

Martin, D. 1996. An assessment of surface and zonal models of population. *International Journal of Geographical Information Systems* 10:973-989.

McNeill, L. 1991. Interpolation and smoothing of binomial data for the Southern African Bird Atlas project. *South African Statistical Journal* 25:129-146.

Oliver, M., R. Webster, C. Lajaunie, and K. Muir. 1998. Binomial cokriging for estimating and mapping the risk of childhood cancer. *IMA Journal of Mathematics Applied in Medicine and Biology* 15:279-297.

Paugam, S. 1991. *La Disqualification sociale. Essai sur la nouvelle pauvreté (Social disqualification. Essay on the new poverty)*. Paris: Presses Universitaries de France.

Room, G. 1995. *Beyond the Threshold. The Measurement and Analysis of Social Exclusion*. Bristol: Polity Press.

Sen, A. 1992. *Inequality Reexamined*. Cambridge: Harvard University Press.

Senior, M. 2002. Deprivation Indicators. In *The Census Data System,* edited by Philip Rees, David Martin, and Paul Williamson. Chichester: John Wiley, pages 123-138.

Spiekermann, K., and M. Wegener. 2000. Freedom from the Tyranny of Zones: Towards New GIS-based Models. In *Spatial Models and GIS. New Potential and New Models*, edited by A. S. Fotheringham and M. Wegener. London: Taylor & Francis, pages 45-60.

Sposati, A. 1996. *Mapa de Exclusão/Inclusão Social de São Paulo*. São Paulo: EDUC.

———. 2000. Map of Social Exclusion/Inclusion for São Paulo - 2000. Social Dynamics of the 1990s. São Paulo: Catholic University.