Residential Segregation Indices as Tools for Identifying Socially Vulnerable Families

Flávia F. Feitosa [a,*], Gilberto Câmara [a], Antônio M. V. Monteiro [a], Thomas Koschitzki [b], Marcelino P. S. Silva [a,c]

[a]*INPE - Instituto Nacional de Pesquisas Espaciais. DPI - Divisão de Processamento de Imagen, C.P. 515, 12227-001, São José dos Campos (SP), Brazil*
[b]*Martin-Luther Universität Halle-Wittenberg. Institut für Geographie. Von-Seckendorff-Platz 4, 06120, Halle (Saale), Germany.*
[c]*UERN - Universidade do Estado do Rio Grande do Norte, BR 110, Km 48, 59610-090, Mossoró, RN, Brazil*

Abstract

Segregation measures can be regarded as useful tools for analyzing the spatial distribution of socially vulnerable families in urban areas. To achieve this, however, it is necessary the employment of measures that are able to overcome the main limitation of the traditional segregation indices: their nonspatial and global character. In other words, the commonly applied segregation measures are unable to consider the spatial arrangement of population and to show how much each areal unit contributes to the segregation degree of the whole city. Therefore, this paper presents the use of alternative spatial segregation measures, global and local, for the identification of families under a socially vulnerable condition established by the combination of poverty and segregation. The study area is São José dos Campos, a medium-sized city located in the State of São Paulo, Brazil.

Keywords**:** Residential segregation; Social vulnerability; Spatial segregation indices; Global and local segregation indices.

1. Introduction

Despite being the largest economy in South America, Brazil is characterized by a great social inequality, as well as by the persistence and reproduction of poverty (The World Bank, 2004). This reality has motivated the emergence of the concept of *social vulnerability*, which includes not only aspects related to poverty and denial of basic human needs, but also the *inability to react* against the negative effects imposed by such conditions (Busso, 2001; Abramovay and Pinheiro, 2003). The degree of social vulnerability of families is usually qualified by the combination of several variables, such as income, education, type of occupation, housing conditions, urban facilities and infrastructure (Nahas et al., 2002). Furthermore, it is important to consider that social vulnerability is also connected with the spatial proximity of families living under the same socioeconomic conditions, in other words, the *residential segregation* (Feitosa et al., 2004).

Circumstances of deep residential segregation of poor families are generally related to precarious or absent infrastructure, difficulties in accessing public services, and a higher exposure to floods, landslides and diseases. The combination between poverty and segregation also promotes the reduction of opportunities for jobs and skills upgrading, the racial and social prejudice, teenage pregnancy, higher exposure to violence, and the acquisition of values and behaviors that impede an upward social mobility (Rodríguez, 2001; Sabatini et al., 2001; Luco e Rodríguez, 2003; Maricato,

∗ Corresponding author. Tel.: +55-12-39456481; fax: +55-12-39456468
E-mail address: flavia@dpi.inpe.br

2003; Torres, 2004). Due to these negative effects of residential segregation on the social vulnerability of families, the theme has received increasing attention in public policy literature.

In spite of the emphasis given to the issue, there are few Latin-American studies focused on the measurement of residential segregation. The existing studies rely on nonspatial measures such as the dissimilarity index (Sabatini, Cáceres, & Cerdá, 2001; Telles, 1992; Torres, 2004) or variance-based indices (Rodríguez, 2001). These measures are not able to distinguish different spatial arrangements among population groups, which is an essential aspect in segregation studies.

In addition, the traditional segregation measures are global and express the segregation degree of the city as a whole. However, the residential segregation is a process that varies along the city (Wong, 2003b). Thus, the measurement of segregation requires also local indices, able to depict how much each areal unit contributes to the composition of the global index.

This work focuses on the use of alternative segregation indices as tools for identifying socially vulnerable families in São José dos Campos (SP, Brazil). These segregation indices, proposed by Feitosa et al. (2004), overcome the limitations of traditional indices by including geographical information in their formulation, and by showing the variation of the segregation degree in different areas of the city.

The remainder of this paper is organized as follows: Section 2 describes the spatial indices used in this work. In Section 3, these indices are decomposed in order to generate local indices. Section 4 exemplifies the use of the segregation indices, spatial and nonspatial, on an artificial data set. Section 5 presents the application of these indices for the identification of socially vulnerable families in São José dos Campos, during the years 1991 and 2000. Finally, the conclusion of this paper is presented in Section 6.

2. Spatial Indices of Residential Segregation

The first studies focused on the measurement of segregation date from the late 1940s and the beginning of 1950s, when several indices were proposed and discussed in the United States. However, these measures were unable to depict the spatial arrangement of population among areal units (Reardon & Firebaugh, 2002; Rodríguez, 2001).

In the 1980s, this limitation became increasingly criticized and some studies started to focus on the development of spatial measures of segregation (Morgan, 1981; Reardon & O'Sullivan, 2004; White, 1983; Wong, 2003a). Although some spatial indices have been already developed since then, most of empirical studies still rely on nonspatial indices. This happens because spatial measures always require the extraction of geographical information and are more difficult to compute than nonspatial measures.

Based on the spatial conception of segregation measurement, two conceptual dimensions of spatial residential segregation were conceived (Reardon & O'Sullivan, 2004): *evenness* (or *clustering*) and *exposure* (or *isolation*) (*Figure* 1). The dimension evenness or clustering refers to the balance of the distribution of population groups and is independent of the population composition of the city as a whole. Exposure or isolation refers to the chance of having members from different groups (or the same group, if we consider isolation) living side-by-side. Exposure depends on the overall population composition of the city (Reardon & O'Sullivan, 2004).

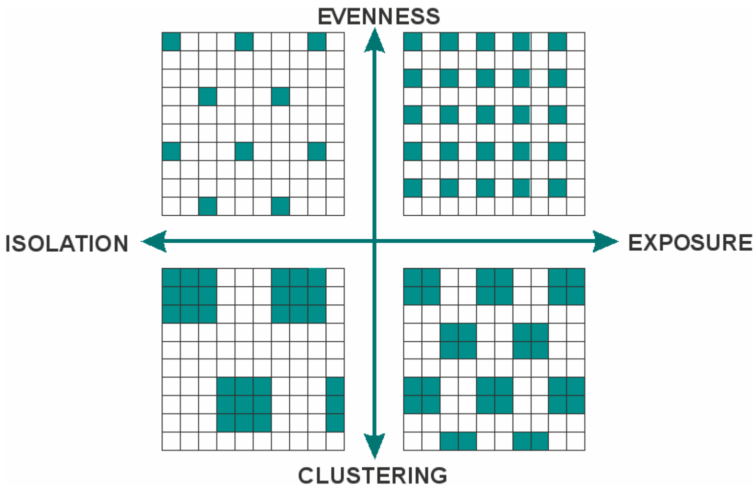

Figure 1. Dimensions of spatial segregation.
(Adapted from Reardon and O'Sullivan, 2004).

This work relies on the use of spatial indices obtained by the extension of two traditional segregation indices: the generalized dissimilarity index *D(m)* (Morgan, 1975; Sakoda, 1981) and the exposure index P^* (Bell, 1954). The index *D(m)* represent the dimension evenness/clustering, and the index P^* measures the dimension exposure/isolation.

The procedure of building spatial versions of segregation indices, presented in Feitosa et al. (2004), is based on approaches introduced by Wong (2003a) and Reardon and O'Sullivan (2004). These approaches cover analyses with different kinds of data – Wong with zones (population count) and Reardon and O´Sullivan with surfaces (population density) - and support the definition of "neighborhoods" where families interact. These neighborhoods - also called "local environment" (Reardon & O'Sullivan, 2004) - are established by proximity functions, chosen by the user according to the purposes of the study.

Like in Wong's approach, the indices used in this work deal with population count data and the neighborhood population count of areal unit *i* for group *m* is defined as (Wong, 2003a)

$$\breve{N}_{im} = \sum_{i=1}^{I} d\left(N_{im}\right), \tag{1}$$

where N_{im} is the population of the group *m* in areal unit *i*, and *d(.)* is the proximity function defining the neighborhood of *i*. The neighborhood population count in *i* is defined as the sum of the population of all areal units, where units are weighted by their proximity to *i*.

Following the same reasoning, Feitosa et al. (2004) defined the proportion of group *m* in the neighborhood of unit *i* ($\breve{\rho}_{im}$) as the ratio of the neighborhood population count of group *m* in areal unit *i* to the total neighborhood population count of *i*:

$$\breve{\rho}_{im} = \frac{\breve{N}_{im}}{\breve{N}_i}\,. \tag{2}$$

Based on the concept of neighborhood population count, the indices *D(m)* and P^* were modified to incorporate spatial information in their formulations. The spatial versions of *D(m)* and P^* used in this work are very similar to the indices proposed by Reardon and O'Sullivan (2004). In the latter case, however, the indices were developed for population density data, while all indices presented in this paper employ population count data.

The spatial version of the generalized dissimilarity index $\breve{D}(m)$ is defined as (Feitosa et al., 2004):

$$\breve{D}(m) = \sum_{i=1}^{I} \sum_{m=1}^{M} \frac{N_i}{2NI} \left| \breve{\rho}_{im} - \rho_m \right|, \tag{3}$$

where

$$I = \sum_{m=1}^{M} (\rho_m)(1 - \rho_m). \tag{4}$$

In Eq. (3) and (4), N is the total population of the city, N_i is the total population in area i, ρ_m is the proportion of group m in the city, and $\breve{\rho}_{im}$ is the proportion of group m in the neighborhood of i. The symbol I represents the interaction index, a diversity measure of population (White, 1986). The index $\breve{D}(m)$ can be interpreted as a measure of how different the population composition of all neighborhoods is, on average, from the population composition of the city as a whole (Reardon & O'Sullivan, 2004).

The spatial version of the exposure index of group m to group n (${}_m\breve{P}_n^*$) is defined as the average proportion of group n in the neighborhood of each member of group m (Feitosa et al., 2004):

$${}_m\breve{P}_n^* = \sum_{i=1}^{I} \frac{N_{im}}{N_m} \breve{\rho}_{in}. \tag{5}$$

Equally, the spatial isolation index of group m can be defined as the spatial exposure of group m to itself:

$${}_m\breve{P}_m^* = \sum_{i=1}^{I} \frac{N_{im}}{N_m} \breve{\rho}_{im}. \tag{6}$$

3. Local Indices of Segregation

All the indices presented until now are regarded as global measures, which summarize the degree of residential segregation of the entire city. However, segregation is a process that varies along the city and the exclusive use of global measures is useless for the identification of socially vulnerable families. Therefore, it is important to rely on local indices that can be geographically displayed as maps.

In this paper, we use two local indices of segregation, obtained by decomposing the global indices $\breve{D}(m)$ and $\breve{P}^*$ (Feitosa et al., 2004). These local indices show how much each unit and its neighborhood contribute to the global segregation measure of a city. The local version of $\breve{D}(m)$ - $\breve{d}(m)$ - is defined as

$$\breve{d}(m) = \sum_{m=1}^{M} \frac{N_i}{2NI} \left| \breve{\rho}_{im} - \rho_m \right| . \tag{7}$$

Likewise, the local version of the isolation index of group m (${}_m\breve{p}_m$) is defined as

$${}_m\breve{p}_m^* = \frac{N_{im}}{N_m} \breve{\rho}_{im} . \tag{8}$$

4. Example Using Artificial Data Sets

For the purpose of exemplifying the use of the presented segregation measures and its advantages, three artificial datasets were employed (Figure 2). The datasets are composed by 144 areal units with equal dimension (10m X 10m) and four population groups with the same proportion (0,25 of each group). In each data set, the population groups were distributed in a different way: (a) the dataset A represents a case of extreme segregation, where each areal unit has just individuals of one group and the units characterized by the same group are clustered; (b) in dataset B, each areal unit has also just individuals of one group, but these units are distributed in a balanced way; and (c) the dataset C represents a case of extreme integration, where each areal unit has the same population composition of the entire set.

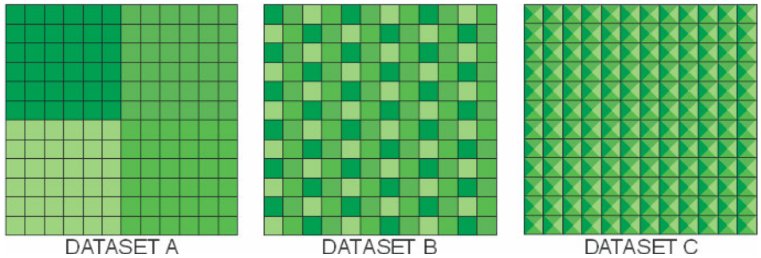

Figure 2. Artificial datasets

For every dataset, the nonspatial index $D(m)$ and the spatial index $\breve{D}(m)$ were computed (Table 1). Gaussian functions with bandwidth 10m and 30m were used to define the neighborhoods required in the spatial indices.

After the computation of global indices, the local index $\breve{d}(m)$ was also calculated for the three datasets, using the same proximity functions (Figure 3).

Table 1. Comparison between $D(m)$ and $\breve{D}(m)$.

	Dissimilarity Indices (range from 0 to 1)		
	$D(m)$ - nonspatial	$\breve{D}(m)$ - gaussian, bw 10	$\breve{D}(m)$ - gaussian, bw 30
Dataset A	1	0.86	0.54
Dataset B	1	0.05	0.04
Dataset C	0	0	0

bw: bandwidth

The results of these tests, presented in Table 1 and Figure 3, show the following:

a) The nonspatial indices indicate the distribution of dataset A and B as cases of maximum segregation ($D(m)$ = 1), despite the fact that dataset A has a much more segregated distribution than B. It means that, if each unit is occupied just by individuals of the same group, the result of nonspatial indices will be always extreme, regardless the spatial arrangement of the units. This limitation has been called as the *checkerboard problem* (Reardon & O'Sullivan, 2004; Sabatini et al., 2001; Wong, 2003a;).
b) The spatial indices indicated a high level of segregation in dataset A, but not the maximum value. This can be justified by the presence of some integrated units, like the ones located at the central area, where different groups are close to each other. By means of local measures (Figure 3), it is possible to observe this.
c) Unlike the nonspatial measures, the spatial indices indicated no segregation in dataset B because of the diversified neighborhood of all areal units. Due to the ability of considering the units neighborhood, the spatial indices overcome the checkerboard problem of nonspatial indices.
d) All indices indicated the dataset C as a case of extreme integration.
e) The local indices are useful in showing which areas are more or less segregated (Figure 3), justifying the global indices results. This ability is essential in the process of identifying families which live under a socially vulnerable condition.
f) Neighborhoods defined by proximity functions with larger bandwidths usually result in lower indices of segregation. This is a scale effect of MAUP (modifiable areal unit problem) and it is usually called as the *grid problem* (Reardon & O'Sullivan, 2004; Sabatini et al., 2001). Due to this problem, it is not appropriate the comparison of results calculated with different bandwidths or different scale of units (e.g., census tracts and districts).

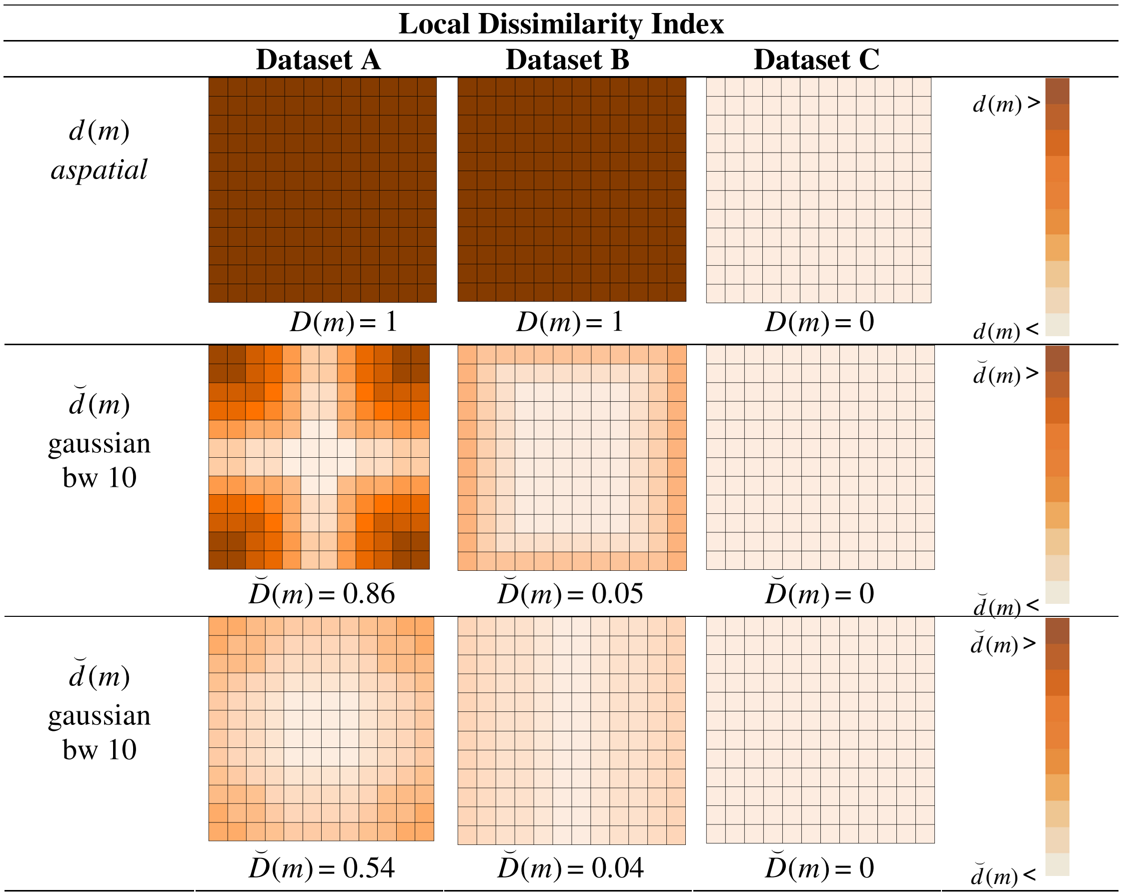

Figure 3. Local dissimilarity index applied to an artificial dataset

6. Socially Vulnerable Families in São José dos Campos, Brazil

In this section, we evaluate the use of spatial measures of segregation for identifying spatial patterns of distribution of socially vulnerable families. The study area is São José dos Campos, a city with 532,711 inhabitants, located in the State of São Paulo, Brazil.

The experiments were based on Census Tracts data from 1991 and 2000. The variable income of householders was chosen to represent the socioeconomic status of households and used in the computation of the segregation indices. The same experiment was repeated using the variable level of education of householders and very similar results were obtained.

Before the computation of the indices, a compatibility procedure between 1991 and 2000 Census tracts was necessary to avoid an inappropriate comparison of results obtained over units with different geometries (grid problem). For the neighborhood population count, Gaussian functions with different bandwidths (400 m and 2000 m) were chosen to define the neighborhoods of each areal unit and an application was developed to compute them. The selection of different bandwidths allows the examination of segregation on different scales, an issue which has been discussed in several studies (Rodríguez, 2001; Sabatini et al., 2001; Torres, 2004). According to Sabatini et al. (2001), both dimensions of segregation (evenness/clustering and exposure/isolation) can show different trends if we analyze them on different scales.

After the neighborhood population count of all units, the spatial dissimilarity index $\breve{D}(m)$ was computed in order to verify the degree of difference between the population composition of the neighborhoods and the population composition of the city as a whole (Table 2). According to the results presented in Table 2, the index $\breve{D}(m)$ revealed the intensification of segregation during the period 1991-2000 on both scales (bandwidth 400 and 2000).

Table 2. Segregation indices (income of householders).

Index	bw 400		bw 2000	
	1991	2000	1991	2000
$\breve{D}(m)$	0.22	0.24	0.10	0.14
${}_{20}\breve{P}^{*}_{20}$	0.20	0.28	0.10	0.16
${}_{<2}\breve{P}^{*}_{<2}$	0.37	0.33	0.32	0.29

For the identification of socially vulnerable families, however, it is essential to complement this information with the use of indices of the dimension exposure/isolation. By means of isolation indices, it is possible to find out whether the increase in segregation, pointed by $\breve{D}(m)$, is connected with the concentration of groups with better or worse socioeconomic conditions. Therefore, the spatial isolation index of householders with income greater than 20 minimum wages (${}_{20}\breve{P}^{*}_{20}$) and the spatial isolation index of householders with income up to 2 minimum wages (${}_{<2}\breve{P}^{*}_{<2}$) were computed. The indices presented in Table 2 show that the isolation of the low-income group (${}_{<2}\breve{P}^{*}_{<2}$) decreased during the period 1991-2000 in both scale, while the isolation of the richest group increased (${}_{20}\breve{P}^{*}_{20}$). Although this result can be explained by the relative improvement of the income indicator in the city, it also shows that the increase in the segregation of São José dos Campos was mainly promoted by the voluntary segregation of high-income families.

Isolation index of householders with income up to 2 minimum wages - bw = 400 m

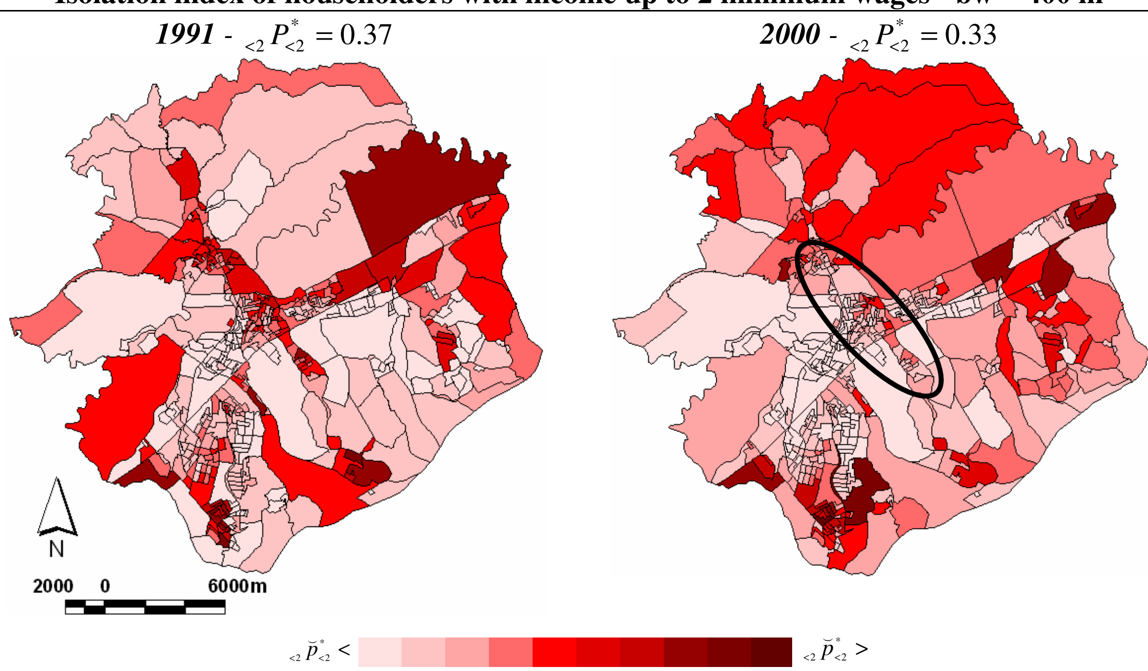

Figure 4. Local isolation index maps - householders with income up to 2 minimum wages (1991 and 2000), bandwidth 400.

The maps of the local isolation index of the householders with income up to 2 minimum wages, presented in Figure 4, show that the pattern of isolation of this group is sparse and located in the outskirts of the city. Concerning the period 1991-2000, it is possible to observe that the isolation of low-income families was maintained in the Northern, Eastern and Southern Region of the city. However, the area encircled in black in Figure 4 presented a clearly decrease in the isolation of the group. This area corresponds to Downtown and surroundings, including some neighborhoods in the Northern and Southern direction.

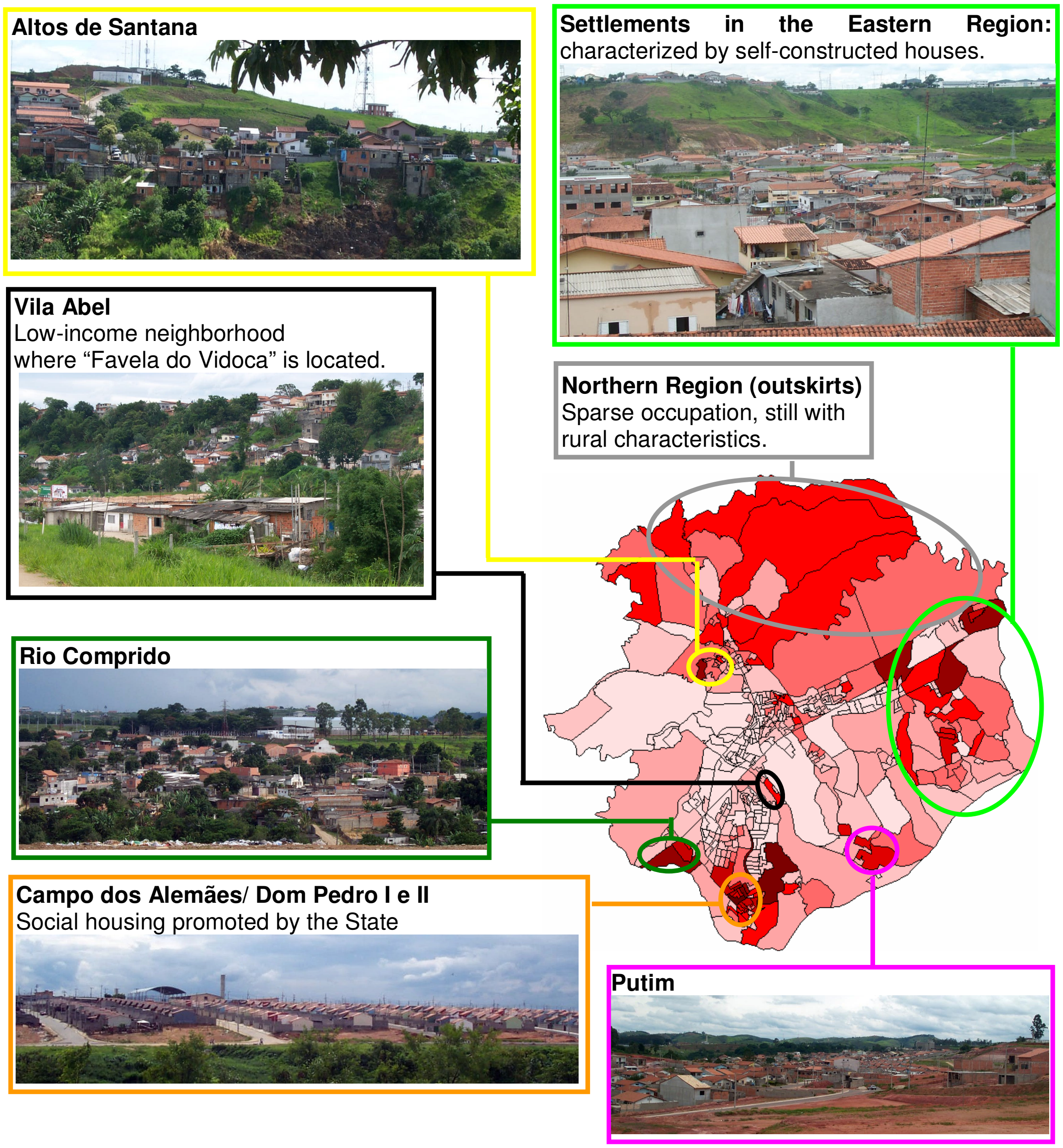

Figure 5. Some areas identified as clusters of isolation of low-income families (year 2000, bandwidth = 400 m).

Figure 5 presents some pictures of areas of isolation of low-income families in 2000. These areas, which combine poverty and segregation, are mainly: (a) settlements, usually produced by irregular or illegal parceling, where self-constructed houses prevails; (b) slums (or “favelas”) located in poor neighborhoods, frequently in high-risk areas of floods or landslides; and (c) settlements characterized by low-income houses constructed by the State, located in the outskirts of the city.

In the maps of local isolation of householders with income up to 2 minimum wages in Figure. 6, which consider a broader scale (bandwidth 2000 m), it is possible to identify *macrosegregation* patterns in the city (Villaça, 1998), in other words, larger regions or group of neighborhoods where this group is clustered. By means of these maps, we observe peripheral clusters of low-income householders in the northern, eastern and southern region (encircled in black in Figure 6), which appear much more consolidate in 2000 than in 1991.

Isolation index of householders with income up to 2 minimum wages - bw = 2000 m

1991 - $\breve{D}(m) = 0.11$ ***2000*** - $\breve{D}(m) = 0.11$

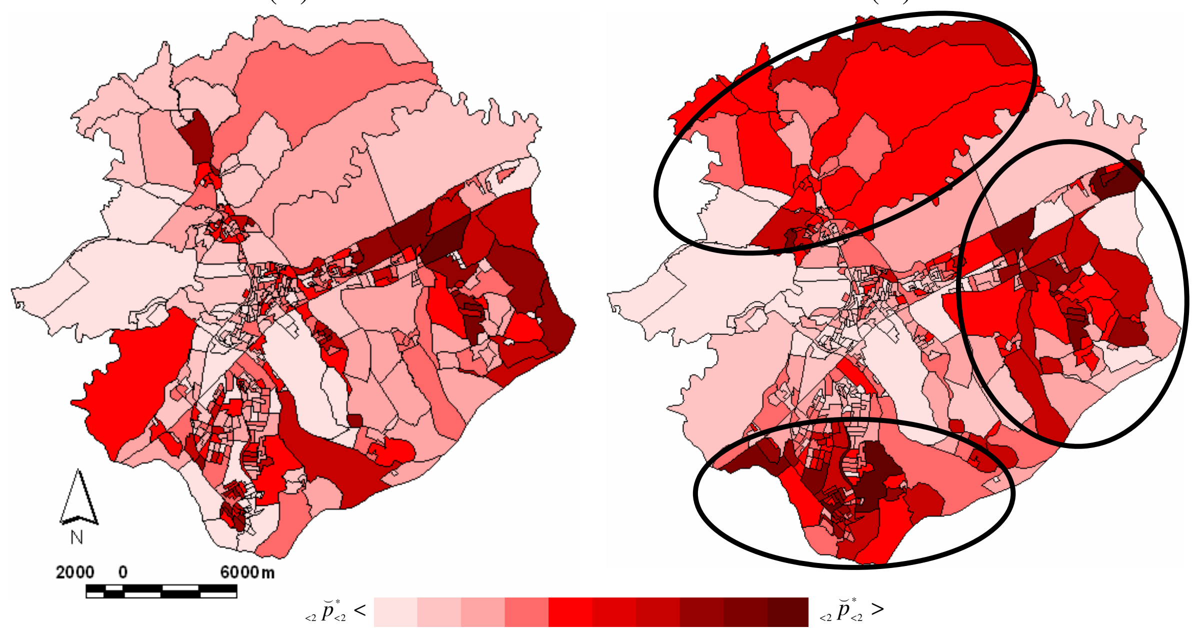

Figure 6. Isolation index of householders with income up to 2 minimum wages, bandwidth = 2000 m .

7. Conclusions

Measures of residential segregation can be applied as an instrument for analyzing patterns of distribution of socially vulnerable families and, therefore, play an important role in public policies. However, most of empirical studies have so far employed nonspatial and global measures, which are unable to consider the spatial arrangement of population and to indicate how much each area contribute to the segregation of the city.

In this work, we used alternative spatial segregation indices to identify areas where the combination poverty and segregation provides a condition of social vulnerability for its residents. The applied indices are based on the definition of “neighborhoods”, defined by proximity functions chosen by the user, which allow analysis on different scales. Another positive aspect of these measures, essential for the identification of families under a socially vulnerable condition, is that they can be decomposed to generate local segregation indices. By displaying these indices as maps, it is possible identify the most segregated areas of the city. The advantages of the segregation indices used in this work were demonstrated by experiments on an artificial dataset.

The paper showed that the segregation dimension "exposure/isolation" is the most important for the identification of socially vulnerable families. By means of this dimension, it is possible to recognize areas which are poor and segregated. In these areas, people are surrounded by others in the same conditions and have less chances of an upward social mobility. Concerning the identification of socially vulnerable families in São José dos Campos, the indices pointed out that the isolation of low-income families decreased during the period, although it has become more defined and concentrated in the outskirts of the city.

References

Abramovay, M., & Pinheiro, L. C. (2003). *Violência e Vulnerabilidade Social*. In Fraerman, A (eds.). Inclusión Social y Desarollo: Presente y Futuro de La Comunidad IberoAmericana. Madrid, Comunica.

Bell, W. (1954). A probability model for the measurement of ecological segregation. *Social Forces*, 32, 337-364.

Busso, G. (2001). La vulnerabilidad social y las políticas sociales a inícios do siglo XXI. Una aproximación a sua potencialidades y limitaciones para los países. In *Seminário Internacional "Las diferentes expressiones de la Vulnerabilidad social en América Latina y el Caribe"*. Santiago de Chile, CEPAL/CELADE.

Feitosa, F.F., Câmara, G., Monteiro, A. M. V., Koschitzki, T., & Silva, M. P. S. (2004). Spatial Measurement of Segregation. In *Proceedings of the VI Brazilian Symposium on Geoinformatics*. São José dos Campos, INPE, p.59-73.

Luco, A., & Rodríguez, J. (2003). *Segregación Residencial en Áreas Metropolitanas de América Latina: Magnitud, Características, Evolución e Implicaciones de Política*. Santiago de Chile: Naciones Unidas.

Maricato, E. (2003). Metrópole, Legislação e Desigualdades. *Estudos Avançados*, 17, 151-166.

Morgan, B. S. (1975). The segregation of socioeconomic groups in urban areas: A comparative analysis. *Urban Studies,* 12, 47-60.

Morgan, B. S. (1981). A distance-decay interaction index to measure residential segregation. *Demography,* 18, 251-255.

Nahas, M. I. P; Oliveira, A. M.; Neto, A. C. (2002). Acesso à Ocupação e à Renda Versus Escolarização no Espaço Intra-urbano de Grandes Cidades: o Caso de Belo Horizonte. In *X Seminário sobre a Economia Mineira*. Diamantina, CEDEPLAR.

Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, 32, 33-67.

Reardon, S. F., & O´Sullivan, D. (2004). *Measures of Spatial Segregation*. Population Research Institute, Working Paper 04-01, Pennsylvania State University.

Rodríguez, J. (2001). *Segregación Residencial Socioeconómica: Que És?, Cómo se Mide?, Que Está Pasando?, Importa?*. CELADE/UNFPA, Santiago de Chile.

Sabatini, F., Cáceres, G., & Cerdá, J. (2001). Segregación residencial en las principales ciudades chilenas: Tendencias de las tres últimas décadas y posibles cursos de acción. *EURE (Santiago)*, 27, 21-42.

Sakoda, J. (1981). A generalized index of dissimilarity. *Demography*, 18, 245-250.

Telles, E. (1992). Residential segregation by skin color in Brazil. *American Sociological Review,* 57, 186-198.

The World Bank (2004). Annual Report 2004. <www.worldbank.org/annualreport/2004>. Nov. 2004.

Torres, H. (2004). Segregação residencial e políticas públicas: São Paulo na década de 1990. *Revista Brasileira de Ciências Sociais*, 54, 41-56.
Villaça, F. (1998). *Espaço Intra-Urbano no Brasil*. Studio Nobel: São Paulo.
White, M. J. (1983). The measurement of spatial segregation. *American Journal of Sociology*, 88, 1008-1018.
White, M. J. (1986). Segregation and diversity measures in population distribution. *Population Index*, 52, 198-221.
Wong, D. W. S. (2003a). Implementing spatial segregation measures in GIS. *Computers, Environment and Urban Systems*, 27, 53-70.
Wong, D. W. S. (2003b). Spatial decomposition of segregation indices: A framework toward measuring segregation at multiple levels. *Geographical Analysis*, 35, 179-94.