Morphological metrics and unsupervised neural networks to analyse urban sprawl and intercity commuting

Cláudia Maria de Almeida [1]
Claudia Durand Alves [1]
Madalena Niero Pereira [1]
Eliana Miglioranza [1]
Lívia Rodrigues Tomás [1]

[1]Instituto Nacional de Pesquisas Espaciais - INPE
Division for Remote Sensing – General Co-Ordination for the Earth Observation
PO Box 515 - 12201-970 - São José dos Campos - SP, Brazil
{almeida,durand,madalena,eliana}@dsr.inpe.br

Abstract. This article focuses on methods and techniques to investigate urban sprawl and intercity commuting regarding the city of São José dos Campos and cities of the Paraíba Valley region, southeast of Brazil. Upon basis of a spatial-relational database containing both its census districts and data derived from origin-destination interviews, which have been applied through stratified sampling, an assessment of its inter- and intra-regional mobility dynamics was carried out, aiming to characterize the profile of commuting regarding not only house-to-work trips, but also trips for the purposes of study, shopping, services, leisure and recreation in secondary residences, and their respective frequencies. This assessment has been conducted by means of unsupervised neural networks, meant to jointly analyse socioeconomic patterns of the population in relation to the trips purpose and frequencies, so as to recognize clusters of similar trip profiles regarding typologies (destination, transport modes, purpose) and intensity (frequencies) in São José dos Campos. These investigations have been supported by urban morphology analyses through landscape metrics, obtained from the delimitation of main urban centres and urban sprawl nuclei extracted from medium- and high-resolution satellite imagery.

Keywords: morphological metrics, landscape metrics, unsupervised neural networks, SOM, urban sprawl, intercity commuting, métricas morfológicas, métricas da paisagem, redes neurais não-supervisionadas, dispersão urbana, mobilidade regional.

1. Introduction

The phenomenon of urban sprawl is reported to take place on two distinct spatial scales: on the one hand, metropolitan regions present an ever-growing scattering of urban nuclei or poles, named by some authors as `urban nebula´, and on the other hand, on the urban level, sprawl is materialised through the gradual fragmentation of the peripheral urban tissue, consisting of dispersed urban settlements and sometimes also of neighbourhoods settled amid rural fields (Reis, 2006).

Urban sprawl is commonly described in the planning and environmental management literature as a diverse form of urban growth, with a diffusive, multinucleated and inefficient character (Ewing, 1997; Burchell and Shad, 1999). On the contrary, some studies have supported the benefits of the scattered style of urban development and argue that the `suburbanisation patterns´, particularly with respect to the American sprawl model, reflect the choices of a free market, of the general consumption behaviour, and of a democratic land control system (Carliner, 1999; Easterbrook, 1999).

In the present context, several studies have been carried out with the purpose to characterise urban sprawl by assessing its particularities as to patterns of spatial configuration, geometric parameters and other specific features. Ewing (1997) describes sprawl as a low-density reclusive and/or diffusive urban development. Burchell and Shad (1999) define sprawl as an intrusion of low-density residential and other non-residential uses in the rural zone as well as in underdeveloped areas, representing a very particular spread, unforeseeable, and sometimes also segregating form of land consumption.

In a more pragmatic approach, Torrens and Alberti (2000) developed an empirical measure of urban landscape diffusion, based on characteristics like density, spread intensity, interspersion/diffusion, and accessibility. Galster *et al*. (2001), in a similar way, employed five parameters to characterise sprawl: density (number of dwellings/area), concentration (degree of unevenness in the dwellings location), centrality (distance to the urban central zone), nuclearity (greater or smaller formation of isolated urban nuclei), and proximity (aggregation degree among similar types and densities of land use).

As stated by Hodge (1992), there is a great awareness among the scientific community that the globalisation process not only influences the economic structure of big cities, but also their spatial structure in a general manner. According to investigations in this line, emerging urban structures reveal a shift from the post-war monocentric model to what is presently called as `polycentric or spread urban region´ (Clark and Kuijpers-Linde, 1994).

For Wardwell (1980), the industrial decentralisation promoted by governmental initiatives in the second half of last century, as a means to decrease industrial locational costs and drive development to smaller cities and underdeveloped regions, does not isolately account for the ongoing changes in the distribution and location of population. According to this author, these changes are also induced by the current technological innovations in transport and telecommunications systems and also by a decrease in transportation costs resulting from the globalisation itself. These topics have become recurrent in recent debates about efficient urban spatial organisation, commuting, traffic jams, and the progressive independence between the locations of home and work (Clark and Kuijpers-Linde, 1994).

In this sense, a parallel phenomenon to urban sprawl is the increasing intercity commuting, what has been also called as `the regionalisation of the daily life´. Commuting is no more restricted in terms of scale and finality. Longer distances are travelled daily, and commuting is not just related to home-to-work trips, but also refers to trips for studies, shopping, medical treatment or leisure and recreation.

This paper is concerned with the quantitative characterisation of the urban sprawl and intercity commuting by means of morphological metrics and unsupervised neural networks. The following section approaches the study area and its recent historical aspects. The third section introduces the construction of the regional commuting database. In the forth section, an analysis of commuting through self-organising maps (SOM) is provided. Section five introduces some morphological metrics and applies them to municipalities in the Paraíba Valley region throughout the four latest decades. In section six, the results of the SOM and metrics analyses are presented and explained in the context of the present changes in regional mobility. Finally, section seven is reserved for final comments and conclusions.

2. The Study Area

The analysis of urban sprawl was conducted for fifteen municipalities of the Paraíba River Valley. In this region, urban sprawl started in the early 1950s with the construction of the Dutra road and the resulting mushrooming of industries along it, mostly of which former settled in the Metropolitan Region of São Paulo. Particularly in the city of São José dos Campos, this process of urban sprawl was intensified by the implementation of the Aerospace Centre (*Centro Tecnológico Aeroespacial – CTA*) in the late 1950s, the National Institute for Space Research (*Instituto Nacional de Pesquisas Espaciais – INPE*) in 1971, and the foundation of the National Aeronautic Industry (*Empresa Brasileira de Aeronáutica – EMBRAER*) in 1970. These huge installations impelled urban development towards its surroundings, what led to the delineation of a spatially diffusive urban pattern in the city.

In the latest years, new drivers of urban sprawl came into scene. Gated communities (or walled neighbourhoods) have been built on pleasant sites outside the cities, around 5 to 10 km away from the urban boundaries. Malls, big services and recreational facilities also displace themselves to settle along the roads, striving to reach a greater number of consumers coming from neighbouring towns. Along with that, universities campii have also been implemented in farther areas. According to Reis (2006), urban sprawl extends in several directions towards the Mantiqueira Ridge, with gated communities for living or leisure settled along roads leading to the mountain villages of São Francisco Xavier and Campos do Jordão, reaching the south of Minas Gerais State. In the opposite side of the Paraíba Valley region, sprawl extends in the direction of cities like Jambeiro (**Figure 1**), and Paraibuna, and southwards in the direction of Guararema, a municipality belonging to the Metropolitan Region of São Paulo. This phenomenon has been intensified with the construction of the Carvalho Pinto Road and may be further strengthened with the enlargement of the Tamoios Road.

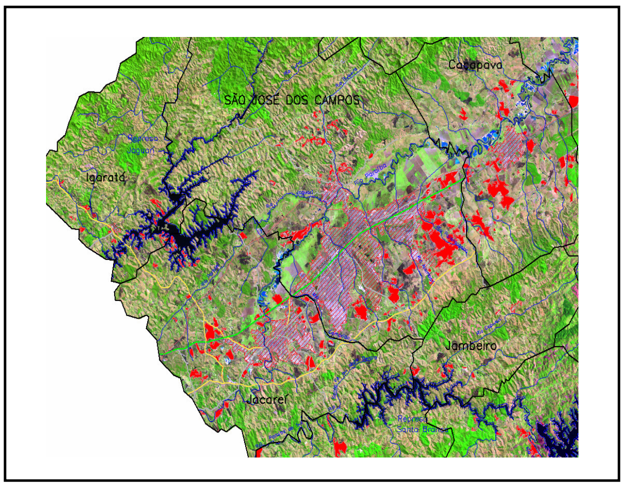

Figure 1. Progress of urban sprawl in São José dos Campos - 2000. Main urban agglomerations in stripes and sprawl nuclei in red, superimposed onto a LS 7 – ETM + image, 219/76, of 09/03/99. (Pereira *et al*., 2006).

3. The Intercity Commuting Database

In order to assess the intercity (regional) commuting dynamics of São José dos Campos, origin-destination (O-D) interviews have been applied throughout the city, upon basis of a stratified sampling approach, taking into account its census districts established in 2000. A total of 4,260 interviews have been realised in the period from August 2005 to April 2006, aiming to diversify the profile of interviewed people in terms of age, social status, and gender. The interviews focussed on aspects like trips purpose, frequency, transport mode, and destination, which included not only cities of the Paraíba Valley but also cities belonging to the Metropolitan Regions of São Paulo, Campinas, and Santos, as well as other cities located in the inland of São Paulo State, in Rio de Janeiro State and in the south of Minas Gerais State.

A database was built in Access to store the information obtained by the O-D interviews. The database comprises nine tables (relations), which refer to numerical codes for the person interviewed and his/her level of occupation as well as for trips destination, purpose, usual and eventual frequency and transport modes (**Figure 2**). Each trip represents a row (tuple) in the trips table, and hence, a composed primary key was created consisting of "*personal identification code – trip destination code – trip purpose code*".

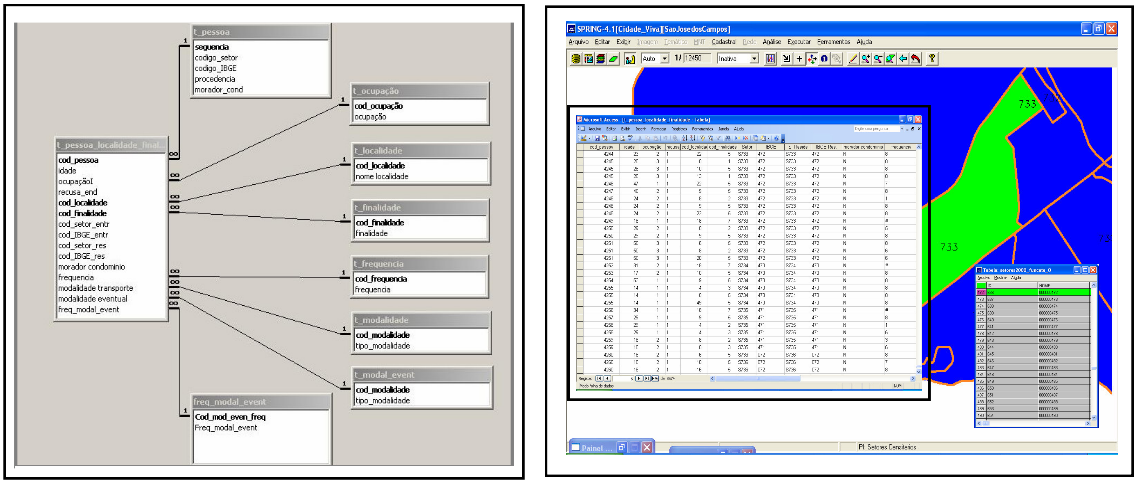

Figure 2. (a) Tables and relationships of the intercity commuting database. (b) Spatial-relational database showing the linkages of the census districts polygons to the intercity commuting database in Access.

An individual `Index of Regional Mobility´ (*Índice de Mobilidade Regional - IMR*) was formulated in order to express in numerical terms the intensity of mobility capacity of each interviewed person. The individual IMR´s were then averaged to generate an aggregated IMR for each census district. The IMR calculation is as follows:

$$IMR_{CDi} = ln\left[\sum_{i=1}^{n} \frac{Fin^{fr}}{NR_{CDi}}\right] \qquad (1)$$

where *Fin* corresponds to a numerical value for each type of trip finality or purpose; *fr*, a numerical value for each type of trip frequency (**Table 1**), and NR_{SCi} represents the number of rows (trips) in a given census district (*CD*) *i*.

Table 1
Numerical values associated with the trips finality and frequency

Trip Finality (Purpose)	Numerical Value	Trip Frequency	Numerical Value
		Daily	*100*
Study	*4*	*Until three times a week*	*80*
Work	*4*	*Weekly (week days)*	*70*
Shopping	*5*	*Weekly (weekends)*	*50*
Services	*4.5*	*Fortnight*	*35*
Leisure/Recreation	*3*	*Monthly*	*25*
Secondary Residence	*2*	*Bimestrial*	*15*
No trips	*0*	*Others*	*10*
		No trips	*1*

The numerical values were assigned so as to give greater importance to the very specialized finalities and higher frequencies, which indicate a stronger mobility capacity.

4. Self-Organising Maps (SOM) for the Intercity Commuting Analyses

Self-organising maps (SOM) are a very special type of unsupervised artificial neural networks based on competitive learning. According to Haykin (1999), on a self-organising map, the neurons (processing units) are located in the nodes of a grid, which is commonly uni- or bi-dimensional. Maps of higher dimensions are also possible, but unusual. The neurons are selectively tuned to the several input data (stimuli) or classes of input data throughout the learning process. The locations of the tuned neurons (i.e. the winners) are reiteratively reorganised, in a way that a coordinate system for different input patterns is created over the grid (Kohonen model). A self-organising map is thus characterised by the delineation of a topographical map of the input patterns, in which the neurons spatial location or coordinates over the grid indicate the intrinsic statistical characteristics contained in such patterns.

Briefly stated, the SOM algorithm comprises three basic steps – sampling, similarity matching and actualisation – which are continuously repeated until the map delineation is completed. The similarity matching concerns the best match of the winner neuron $\boldsymbol{i}(\mathrm{x})$ in time step n through the criterion of minimum Euclidean distance:

$$i(\mathbf{x}) = \arg\min_{j} \|\mathbf{x}(n) - \mathbf{w}_j\|,\ \mathrm{j}=1,2,\ldots,l \qquad (2)$$

The actualisation stage regards the adjustment of the synaptic weights vectors referring to all neurons according to the equation below:

$$\mathbf{w}_j\,(n+1) = \mathbf{w}_j\,(n) + \eta\,(n)\,\mathrm{h}_{j.i\,(\mathbf{x})}\,(n)\,(\mathbf{x}\,(n) - \mathbf{w}_j\,(n)) \qquad (3)$$

where $\eta\,(n)$ is the learning factor and $\mathrm{h}_{j.i\,(\mathbf{x})}\,(n)$ is the neighbourhood function centred around the winner neuron $i(\mathbf{x})$; both $\eta\,(n)$ and $\mathrm{h}_{j.i\,(\mathbf{x})}\,(n)$ dynamically vary during learning in order to obtain the best results.

The IMR together with ten variables related to socio-economic aspects of the city census districts - 729 according to the Census of 2000 - were used to drive the SOM neural network in the Finnish programme SOMPak. These eleven variables have been previously normalised in function of their respective mean and standard deviation. Initially, twenty-five clusters (hexagons with circumference inside) have been employed. A regrouping in the clustering results was further made, aiming to come out with six clusters only. On a first stage, this reclassification was realised visually, where the closest neurons, i.e. those separated by hexagons (without circumference) with lighter tons of grey, were grouped (**Figure 3**). On a

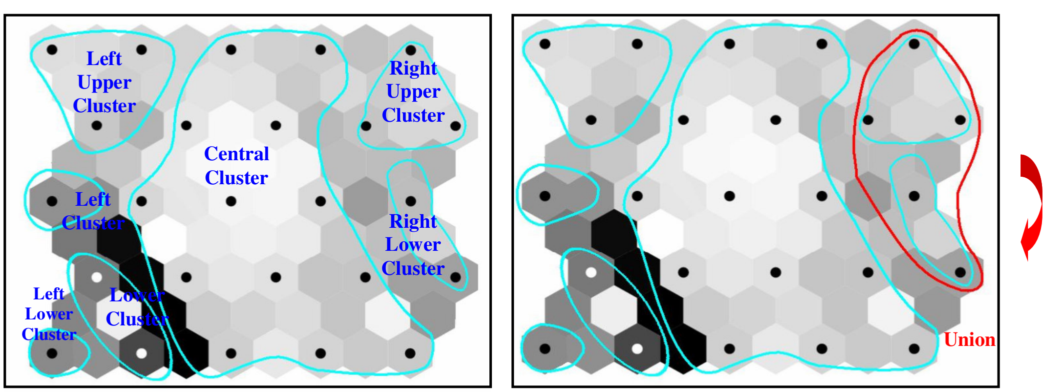

Figure 3. (a) Output matrix of SOMPak with 25 neurons subject to a preliminary visual regrouping into seven clusters. (b) Redefinition of clusters with the aid of graphics analyses, and the union of the clusters to the right.

second stage, graphical analyses were performed, in which the mean regarding the eleven variables of each neuron in the considered cluster was plotted against the mean of each bordering neuron belonging to the possible clusters to be aggregated.

5. Morphological Metrics to Assess Urban Sprawl

A diverse set of morphological or landscape metrics was used to assess urban sprawl, mostly implemented in the programme Fragstats, of which three will be described in detail below.

The Aggregation Index (*AI*) indicates how often pair of patches of the same class show up side by side on the landscape, and is given by:

$$AI = [g_{ii} / max\text{-}g_{ii}] * 100 \quad (4)$$

where g_{ii} is the number of adjacencies between pixels of class i, and $max\text{-}g_{ii}$ is the maximum number of adjacencies between pixels of class i, both estimated through simple counting.

The Fractal Dimension Index (*FRAC*) reflects shape complexity throughout a wide range of spatial scales (patch sizes), overcoming constraints imposed by the area/perimeter ratio.

$$FRAC = 2 \ln \ (0.25\, p_{ij}) \, / \, \ln a_{ij} \quad (5)$$

where p_{ij} is the perimeter (m) of patch ij, and a_{ij} is the area (m^2) of patch ij.

The Contiguity Index (*CONTIG*) assesses the spatial connectivity or contiguity of cells in a patch, so as to provide a measure of the patch boundary spatial configuration, and as such, of the patch shape. It is given by:

$$CONTIG = [\sum_{r=1}^{z} c_{ijr} / a_{ij}] - 1 / (v - 1) \quad (6)$$

where c_{ijr} is the contiguity value for pixel r of patch ij (computed according to a function that assigns value 2 to orthogonally contiguous pixels, and value 1 to diagonally contiguous ones, upon basis of a 3x3 moving window), v is the sum of values found in the moving window; and a_{ij} is the area of patch ij in terms of cell numbers.

The delimitation of urban sprawl was made directly upon topographic charts for the year 1970 and upon basis of satellite imagery for the other years referring to each of the following decades. The first LS 5-TM free of clouds in the study area was obtained in 1984. For the 1990s, an image of 1991 was chosen, since this was the year when the national census was accomplished. LS7-ETM+ images were used to analyse sprawl in 2000. In all cases, ISOSEG was used in the classification. In the particular case of São José dos Campos, an IKONOS image was also employed to identify urban areas. The above-mentioned indices and further metrics have been applied to municipalities of the Paraíba Valley, as indicated in **Figure 4.**

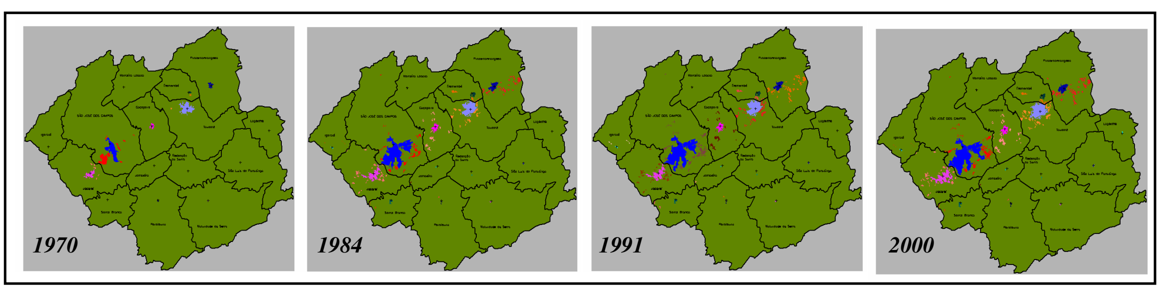

Figure 4. Main urban agglomerations and sprawl nuclei in the Paraíba Valley region in the latest four decades.

6. Results and Discussion

The final six classes identified through SOM were spatialised over the census districts map of São José dos Campos. The low/very low standard, in yellow, corresponds to peripheral neighbourhoods, semi-rural areas, and even squatter settlements located in more central areas. The medium-low standard, in orange, refers to medium-low neighbourhoods, also located in peripheral areas. The medium standard, in olive, is observed in more central neighbourhoods, including most of CTA. The salmon colour accounts for the high socioeconomic and commuting standard, and is found in wealthier neighbourhoods, including the high officials residence sector of CTA. The red and brown correspond to the very high and extremely high standards respectively, and are mostly found in gated communities of the city (**Figure 5**).

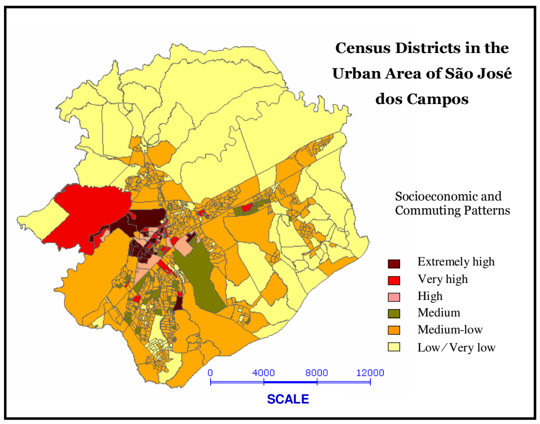

Figure 5. Current socioeconomic and commuting patterns in São José city resulting from the SOM analysis.

Regarding the morphological metrics, it was observed a growing disaggregation (distancing) of sprawl nuclei in the Paraíba Valley in the latest decades, once sprawl increases its reach. Cities with no sprawl in 1970 presented an AI of 100. Bigger cities have lower values of *AI*, since the greater incidence of nuclei tend to drive them closer. Complexity (*FRAC*)

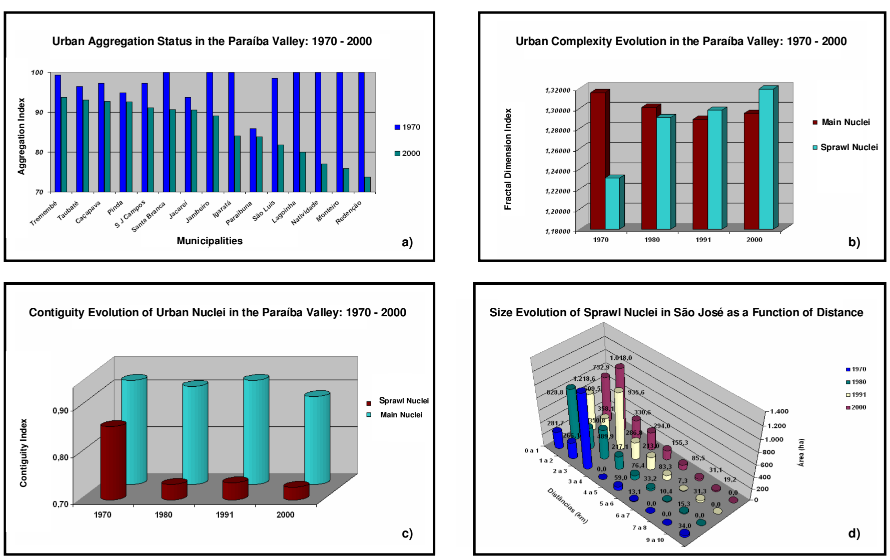

Figure 6. Morphological metrics to assess sprawl in the Paraíba Valley region: (a) Aggregation Index; (b) Fractal Dimension Index; (c) Contiguity Index, and (d) Size of sprawl nuclei (ha) according to distances ranges (ha).

increases for the sprawl nuclei, for they are formed by the gradual incorporation of new settlements, with irregular borders. Main nuclei, however, tend to slightly smooth their boundaries as they grow in size. Provided that sprawl nuclei have very irregular perimeters, their contiguity (*CONTIG*) or compactness decreases, whilst main urban nuclei present a more stable behaviour, with mild oscillations. Finally, the analyses revealed that in the city of São José, the predominance of bigger sprawl nuclei occur within 4km-distance from the main urban agglomeration boundary, since longer distances imply more recent (and thus smaller) nuclei as well as a decreasing locational attractiveness.

7. Final Remarks and Conclusions

This work was meant to quantitatively describe and analyse the correlated phenomena of intercity commuting and urban sprawl. The spatial analysis of regional commuting together with income and educational indicators through the SOM algorithm presented a result that reflected consistent current socioeconomic and mobility patterns in the city of São José dos Campos. The urban sprawl analysis, on its turn, focused not only on the present time but rather on the latest four decades, once sprawl is essentially a temporal phenomenon. It is worth mentioning that the methods employed in this research proved to be robust and efficient for the quantitative characterisation of both phenomena under consideration.

References

Burchell, R. W.; Shad, N. A. The evolution of the sprawl debate in the United States. **West-Northwest Journal of Environmental Law and Policy**, v. 5, n. 2, p. 137-160, 1999.

Carliner, M. S. Comment on Karen A. Danielsen, Robert E. Lang, and William Fulton´s “Retracting suburbia: smart growth and the future of housing”. **Housing Policy Debate,** v. 10, n. 3, p. 549-553, 1999.

Clark, W. A. V.; Kuijpers-Linde, M. Commuting in restructuring urban regions. **Urban Studies** (Special Issue: Globalisation, World Cities and the Randstad), v. 31, n. 3, p. 465-484, Apr. 1994.

Easterbrook, G. Comment on Karen A. Danielsen, Robert E. Lang, and William Fulton´s “Retracting suburbia: smart growth and the future of housing”. **Housing Policy Debate**, v. 10, n. 3, p. 541-547, 1999.

Ewing, R. Is Los Angeles-style sprawl desirable? **Journal of the American Planning Association**, v. 63, n. 1, p. 107-126, 1997.

Galster, G.; Hanson, R.; Wolman, H.; Coleman, S.; Freihage, J. Wrestling sprawl to the ground: defining and measuring an elusive concept. **Housing Policy Debate**, v. 12, n. 4, p. 681-717, 2001.

Haykin, S. S. **Neural networks:** a comprehensive foundation. Upper Saddle River, NJ: Prentice Hall, 1999. 842p.

Hodge, D. C. Urban congestion: reshaping urban life. **Urban Geography**, v. 13, p. 577-588, 1992.

Pereira, M. N. ; Gonçalves, C. D. A. B. ; Souza, I. M. E. ; Garcia, S. ; Portela, A. G. ; Almeida, C. M. ; Florenzano, T. G. Uso de Imagens de Satélite como Subsídio ao Estudo do Processo de Urbanização. **Revista de Estudos sobre Urbanização, Arquitetura e Preservação**, v. 46, p. 6-33, 2006.

Torrens, P.; Alberti, M. **Measuring sprawl**. Disponível em <http://www.casa.ucl.ac.uk/measuring_sprawl.pdf>. Acesso em 27/01/2006.

Reis, N. G. **Notas sobre urbanização dispersa e novas formas de tecido urbano**. São Paulo: Via das Artes, 2006. 201p.

Wardwell, J. Toward a theory of urban-regional migration in the developed world. In: Brown, D. L.; Wardwell, J. ed. **New directions in urban-rural migration:** the population turnaround in rural América. New York: Academic Press, 1980.