Evaluating Transportation Accessibility with Spatial Statistics Tools in a GIS Environment

A.P. Krempi [1], N.C.M. Brondino [2] and A.N.R. Silva [1]
[1] University of São Paulo – USP
São Carlos School of Engineering - Department of Transportation
Av. Trabalhador São-Carlense, 400, 13566-590 São Carlos, SP, Brazil
[2] São Paulo State University - UNESP
Faculty of Sciences - Department of Mathematics
Av. Eng. Luiz Edmundo Carrijo Coube, s/n, 17033-360 Bauru, SP, Brazil

ABSTRACT

In several developing countries it is often assumed that low-income segments of the population living at the periphery of the cities are those affected the most by poor conditions of transportation accessibility. In order to gain a better understanding of the way transportation accessibility is distributed across different regions of an urban area, the main aim of this work is to analyze, making use of Spatial Statistics tools in a GIS (Geographical Information System) environment, the relationship between accessibility and geographical locations in a medium-sized Brazilian city. Data of an origin-destination (O-D) survey carried out in the city of Bauru, which brings information about four different transportation modes, were used in this study. Such data, grouped following the census tracts, were carefully examined in a Geographic Information System in order to look for spatial patterns of accessibility that are not visible in the traditional approaches. One of the interesting outcomes of the application was the identification of regions with particular dynamics, which go against the pattern found in the overall urban area. This and other results of the case study clearly indicate that Spatial Statistics analyses in a GIS environment create a powerful tool to extend conventional transportation accessibility analysis.

1 INTRODUCTION

Several Brazilian cities have experienced a very fast growth process of their urbanized areas in recent years, what has caused significant impacts on their transportation systems. One such a problem is an overall reduction of transportation accessibility. Transportation accessibility is directly related to the level of transportation supply and land uses and the way they affect individuals in their trip desires for accomplishing regular-basis activities. Due to the economic and social characteristics of Brazil, the accessibility distribution in urban areas seems to follow a general pattern found in several other developing country cities. The large distances resulting from urban expansion make difficult the displacements of those people who do not have access to motorized transportation modes. Consequently, the general belief is that low-income segments of the population living at the periphery of the cities are those affected the most by poor transportation accessibilities. With long distances to traverse and usually relying only on deficient public transportation systems, those citizens spend a large share of their daily time traveling, mostly to and from work.

Although some authors, such as Silva *et al.* (1998); Silva *et al.* (2000); Silva *et al.* (2001); and Goto *et al.* (2001), had tried to gain a better understanding of the way transportation accessibility is distributed across different regions and for various population segments in Brazilian urban areas, several points remain unclear. Are really those who live at the periphery of medium-sized cities the people most affected by low accessibility levels? Or, considering that in some cities the population distribution pattern is changing, giving place to developments for high-income groups, should we not look, when evaluating accessibility distribution, for differences in the income levels of those living at the periphery also in developing countries? If that is the case, those developments certainly have available a good transportation infrastructure, what might indicate a high accessibility level. Actually, it may be in some cases higher than the accessibility levels of other urban regions, which are in theory better located in geographically terms, i.e., closer to attraction nodes.

Seeking for answers to these questions, the main aim of this work is to analyze the relationship between accessibility and geographical locations in a medium-sized Brazilian city, based on the assumption that the accessibility distribution pattern is affected both by local and global attributes of an urban area. Data of an origin-destination (O-D) survey carried out in the city of Bauru, bringing different transportation modes information, were used in this study. Such data, grouped following the census tracts, were carefully examined with Spatial Statistics tools in a Geographic Information System environment in order to look for spatial patterns of accessibility that are not visible in the traditional analysis approaches.

The present work is organized as follows. In section 2, some characteristics of the Spatial Statistics tools that were selected for the present study and the advantages of their use in a GIS environment are briefly highlighted. Next, the method steps are described in section 3, followed by a description of the results obtained in part 4, and the overall conclusions of the paper in section 5.

2 SPATIAL STATISTICS AND GIS

According to Levine (1996), some statistics can be used to describe the spatial distribution of data. The variables that can be spatially described can refer to points (such as in the case of accidents locations, buildings, people, etc.) or they can be aggregated as areas (e.g., traffic zones, urban boundaries, etc.). The statistics used to describe both points and areas fall into three general categories:

- *Measures of spatial distribution:* they describe the center, dispersion, direction, and shape of a variable's distribution;
- *Measures of spatial autocorrelation:* they describe the relationship among the different locations for a single variable, therefore indicating the degree of concentration or dispersion;
- *Measures of spatial association between two or more variables:* they describe

the correlation or association between variables distributed over space, for example, the correlation between liquor stores locations and points with high traffic accidents rates.

The importance of spatial attributes in some analyses and the characteristics of the statistical tools used to carry out those analyses may explain the growing interest in combining spatial statistics tools with Geographical Information Systems. Even though, applications are in general still incipient. Examples of authors who started to explore that combination are Bullen (1997) and Câmara *et al.* (1999). The former used regression models to assess property values in the United Kingdom, while the latter analyzed if the population longevity could be somehow explained by dwelling locations in the city of São Paulo, Brazil. These and several other studies, such as Haining (1995), Fischer and Getis (1996), Patkar (1997), and Forkenbrock *et al.* (2001) clearly show the advantages of using spatial statistics in a GIS environment, since the resources they offer complement each other in terms of understanding the information that may be comprised in spatial data and sometimes is not easily classified or visualized.

3 STUDY METHOD

The analysis presented in this paper has been carried out with the tools available in TransCAD, which is a GIS-T (Geographic Information System for Transportation) software package. The main source of data for this study was an Origin-Destination survey carried out in the Brazilian city of Bauru, which had over 300,000 inhabitants. Interviews were conducted in a sample of 4,000 households, which is about 4.5 % of the total number of households in the city (Reyes 1999). The survey recorded data of 23,314 trips using four transportation modes: car/motorcycle (as driver), car/motorcycle (as passenger), bus, and walk/bicycle. The data was grouped following 306 census tracts boundaries and the mode share originated in each area was recorded as a percentage of the total trips per zone.

Thematic maps showing the mode split per zone were initially built in order to have a first grasp of the accessibility distribution across the urban area. Subsequently, the first step of the analysis was the estimation of spatial autocorrelation values, which made use of three basic elements:

- *Spatial proximity matrix (P):* matrix of dimension n x n, in which every p_{ij} element receives a value of one if zones *i* and *j* are neighbors and zero if not. The matrix is normalized by the division of each element equal to one of a line *i* by the total sum of the same line, what originates matrix *W*,
- *Vector of deviations (Z):* each element of the vector is obtained by the subtraction of the total mean (μ) from the attribute value of each zone ($z_i = y_i - \mu$),

- *Vector of weighed averages (W_Z):* product of W by Z. Each element of the vector is the average of the deviations of zone *i* neighbors.

The spatial autocorrelation index applied here was Moran`s I. That coefficient, which varies from -1 to +1 and has an expected value approaching zero for a large sample size in the absence of autocorrelation, is calculated through equation (1).

$$I = \frac{Z^t W_Z}{Z^t Z} \qquad (1)$$

where the superscript t denotes transposed vector.

A careful examination of expression (1) suggests that Moran`s coefficient can be interpreted as the linear regression coefficient, looking at W_Z as the dependent variable and Z as the independent variable. In such a way, the index I could be thought as the slope of the regression line adjusted to the set of pairs. From that analogy, one can conclude that:

- If the slope is zero, the W_Z values do not vary according to Z values, so there is no relationship between the value of the attribute measured for any particular zone and the values of the same attribute measured for its neighbors;
- If the slope is positive, when the Z_i´s values grow, the W_Z´s values also increase. It means that when the value of the attribute measured for a particular zone increases, the average value of the same attribute in the neighbor zones also increases;
- If the slope is negative, when the Z_i´s values grow, the W_Z´s values decrease. It means that when the value of the attribute measured for a particular zone increases, the average value of the same attribute in the neighbor zones decreases.

In order to graphically observe those relationships, the values of W_Z x Z were plotted here for each transportation mode. The graph allows, by visual comparison, an immediate comprehension of the existing relationship between the attribute values of any single zone and its neighbor zones. By splitting the graph with lines that cross each other at point zero, four possible combinations of W_Z and Z_i values can be identified (Figure 1). The two lines divide the space in four quadrants, described here as Q1, Q2, Q3, and Q4.

Points located in quadrants Q1 and Q2 indicate that the attribute value of a particular zone is similar to the average value of the same attribute in neighbor zones (positive value for the zone and positive average value for neighbors in Q1 and negative value for the zone and negative average value for neighbors in Q2). This is an indication of positive spatial autocorrelation.

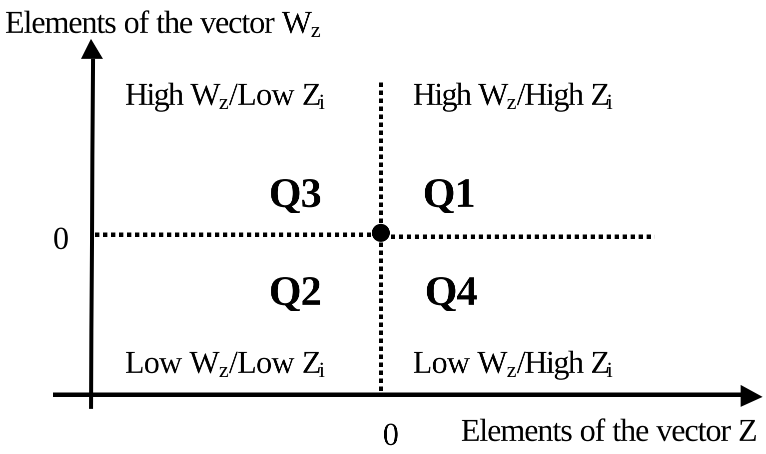

Figure 1: **Interpretation of Moran's scatterplot**

Points are located in quadrants Q3 and Q4 if the attribute value of a particular zone is dissimilar to the average value of the same attribute in neighbor zones (positive value for the zone and negative average value for neighbors in Q3, and negative value for the zone and positive average value for neighbors in Q4). This is an indication of negative spatial autocorrelation. Zones located in quadrants Q3 and Q4 can be seen as extreme cases regarding the variable under consideration, since the attribute values do not follow the pattern of the close neighbors.

Thematic maps based on the point's location in Figure 1 were built next, for each transportation mode. This kind of representation allows the identification of each element according to its classification (Z_i and W_Z values), directly by the identification of the quadrant they belong to. In that way, one can visually identify the correlation between the attribute value measured for a particular zone and the average value obtained for its neighbors.

4 RESULTS

For the analysis carried out in the present study, the variables were identified as follows:

- *% car_driver:* percentage of car trips as driver
- *% car_pass:* percentage of car trips as passenger
- *% non_mot:* percentage of trips by non-motorized modes
- *% bus:* percentage of bus trips

The general Moran´s I index obtained for the variable *% car_driver* with equation (1) was 0.375, which indicates a relatively low spatial autocorrelation for that variable. Figure 2 shows the distribution of the Z_i and W_z values for this case, in which one can detect the presence of points in all quadrants. Points located in Q1 and Q2 indicate that the attribute value of a particular zone is similar to the average value of the same attribute in neighbor zones, while the attribute value of points located in Q3 and Q4 do not follow the average pattern of the close neighbors. The location of all regions to which those points are referring to can be seen in the thematic map of Figure 3.

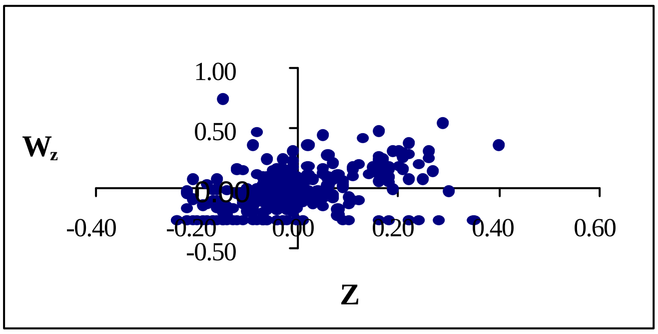

Figure 2: **Moran's scatterplot for the variable *% car_driver***

In the map shown in Figure 3 one can observe that several regions associated to points in quadrant 1 are geographically located in the center and south sectors of the city. In those zones, there is a positive correlation between the attribute value of the zone and the average value of the same attribute in neighbor zones, i.e., the attribute value is similar to the neighborhood average value. For zones in quadrant 1 the deviation value per zone and the average of deviations are both positive, what indicates that the percentage of car trips (as driver) is higher than the average percentage measured for the entire city. Points in Q2 are predominantly located in the west, north, and east regions of the city. Again, there is similarity between the attribute value of the zone and the average value of the same attribute in neighbor zones. In that case, however, the deviation value per zone and the average of deviations are both negative. Thus, the percentage of car trips (as driver) is lower than the average percentage measured for the entire city.

It is interesting to observe, also in Figure 3, the fact that most points in quadrants 3 and 4 are associated to peripheral zones of the city. For regions in Q3, which are visible mainly in the north of the city, the attribute value of the zone is less than the average value of the surrounding area. On the other hand, in Q4 zones, which are mostly located in the south of the city, the percentage of car trips (as driver) is higher than the average percentage value measured for the entire city, although the surrounding zones are below that average.

The same kind of analysis was conducted for all other transportation modes under investigation. For *% car_pass*, the index obtained with equation (1) was 0.189, which reveals a low spatial autocorrelation for the variable. Figure 4 shows the distribution of the Z_i and W_z values for that variable and Figure 5 is a thematic map in which the zones to which those points are referring to are located.

Figure 3: **Thematic map for the variable *% car_driver***

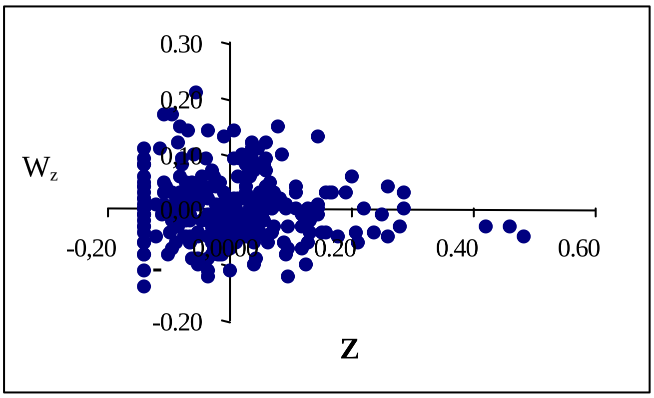

Figure 4: **Moran's scatterplot for the variable *% car_pass***

Through the map of Figure 5 one can see that a significant share of the zones in Q1 are in the central and southern parts of the city. There are also zones in the southeastern part of the urban area that belong to the same class, in which the percentage of trips as a car passenger is higher than the average value for the entire city. The zones in Q2, in which the percentage of car trips (as driver) is lower than the average percentage measured for the entire city, are spread all over the periphery, except in the south. Points in Q3 and Q4 are associated to zones spread all over the urban area.

Figure 5: **Thematic map for the variable *% car_pass***

For the variable *% non_mot*, Moran's I was equal to 0.162, which was the lowest spatial autocorrelation found in this study. Figure 6 and 7 show the distribution of the Z_i and W_z values for the variable and the map with the zone locations, respectively. In Figure 7 is noticeable that Q1 zones are spread all over the city, although slightly concentrated in the central and western parts of the city. There are also zones in the eastern part of the city in the same class, in which the percentage of trips by non-motorized modes is higher than the average value for the whole city. The zones in Q2 are mainly concentrated in the south of the city. In these second quadrant zones, the percentage of non-motorized trips is lower than the average percentage measured for the entire city.

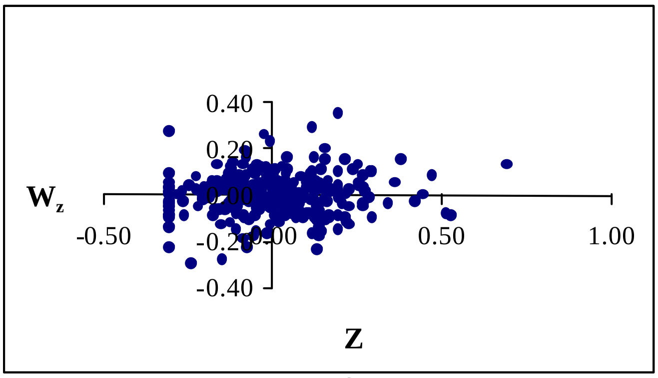

Figure 6: **Moran's scatterplot for the variable *% non_mot***

This result was somehow expected considering that the southern sector of the city had the highest percentages of car trips. Most points in Q3 and Q4 are located in the northern half of the city.

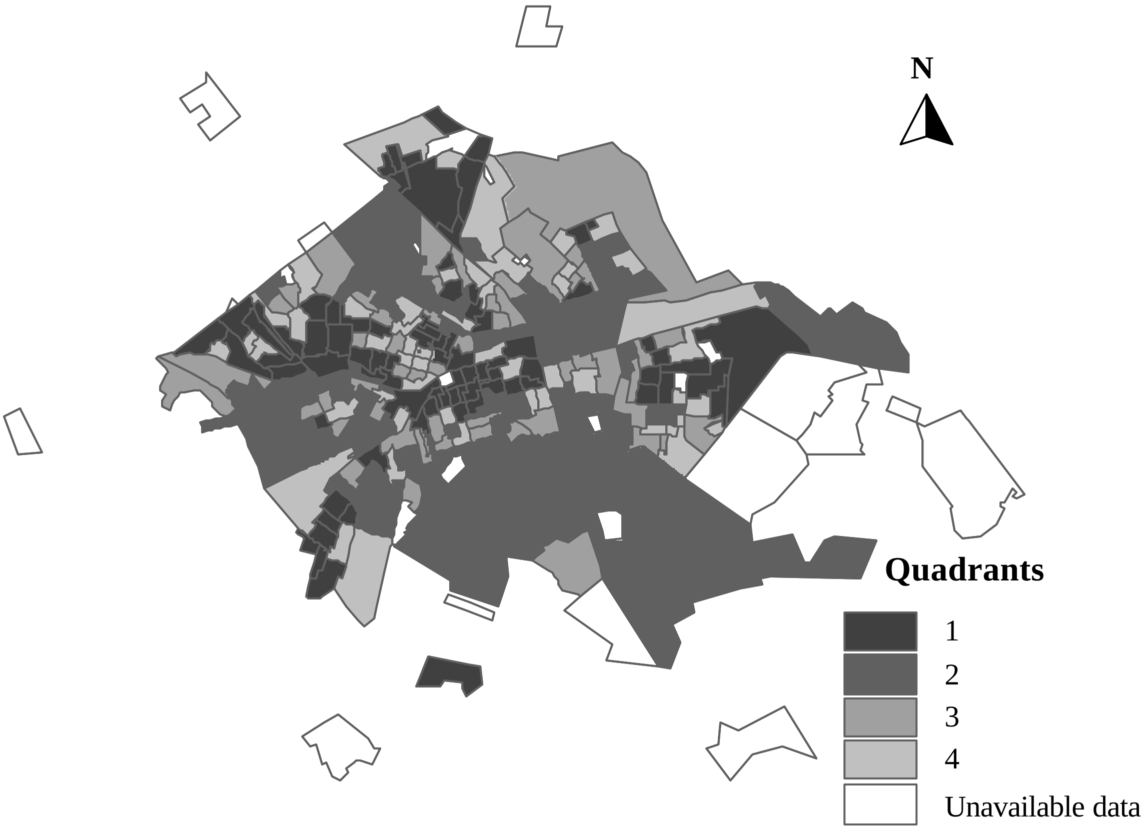

Figure 7: **Thematic map for the variable *% non_mot***

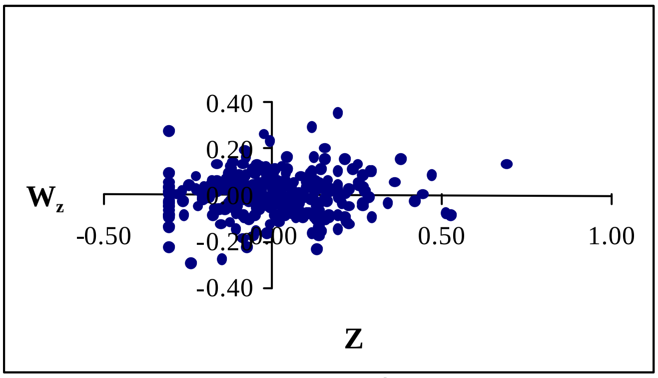

Figure 8: **Moran's scatterplot for the variable *% bus***

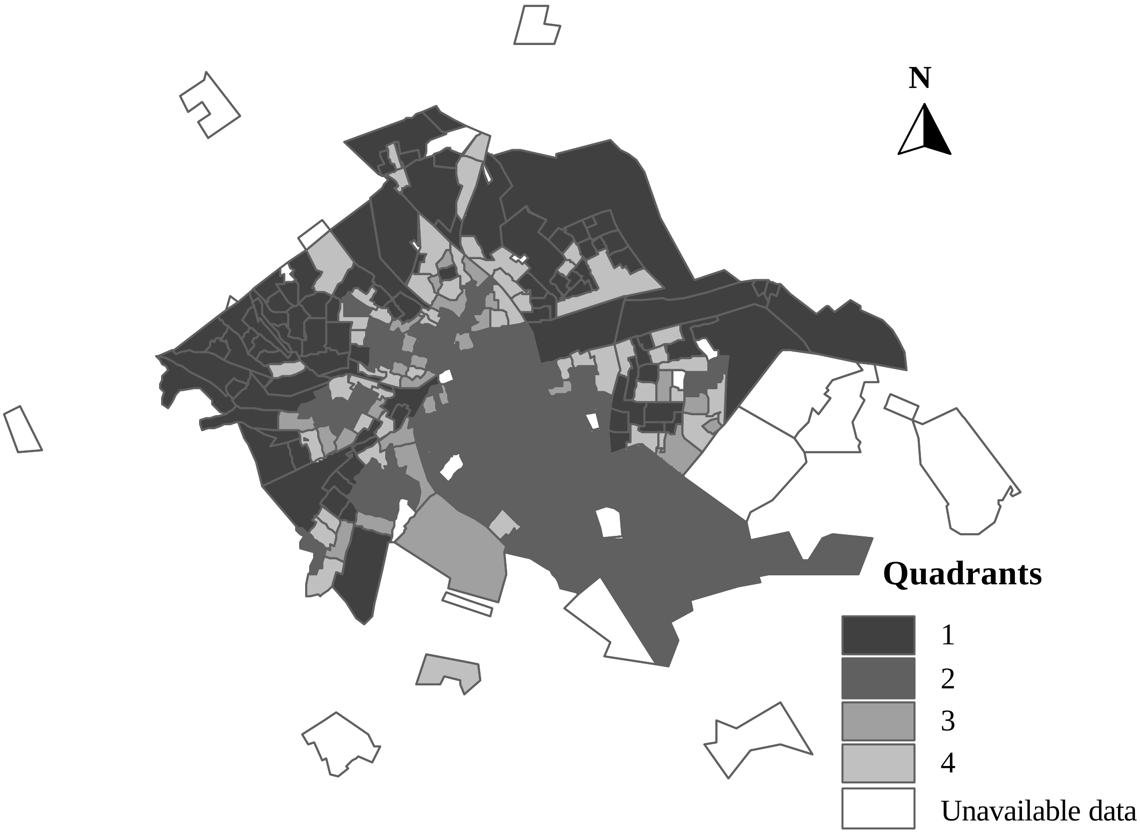

Figure 9: **Thematic map for the variable *% bus***

Finally, the Moran's I value for the variable *% bus* was equal to 0.361. That was also a relatively low spatial autocorrelation. Figure 8 shows the distribution of the Z_i and W_z values for that variable, and Figure 9 is the map in which the zones are located according to the classes they belong to. In the latter, one can see that zones in quadrant 1, in which the percentage of bus trips is higher than the average value for the entire city, are concentrated in the northern, eastern and western parts of the city. On the other hand, Q2 zones, which are concentrated in the central and southeastern parts of the urban area, have a lower percentage of bus trips than the average percentage measured for the entire city. The extreme cases, i.e., points in quadrants 3 and 4, are related to zones spread all over the city, although rarely in peripheral sectors.

5 CONCLUSIONS

In order to better understand the influence of location on transportation accessibility, data of four different transportation modes have been analyzed with spatial statistics tools in a GIS environment in this case study. At first, the global Moran's I index of spatial autocorrelation was estimated for all modes. The results were: 0.375 for *% car_driver*, 0.189 for *% car_pass*, 0.162 for *% non_mot*, and 0.361 *for % bus*. The values found were all low positive spatial autocorrelation indicators. That means that the attribute values measured for several city zones were not similar to the values found for their immediate neighbors. This statistic, however, does not make a distinction of the zones that have this particular condition, working only as a global indicator for the entire city.

The analyses of Moran's scatterplots and of the maps produced in the sequence allowed some particularized conclusions about the use of the transportation modes in the different city zones. The general conclusion of this work is that trips originated in the peripheral zones are predominantly made by motorized (car and bus) modes. This can be seen in Figures 3 and 9, which somehow complement each other. The relevant fact here is that one can possibly think about competition between modes, given that a more intense use of one mode is almost directly reflected in a lower usage of the other. In other words, it seems that the residents at the periphery of the city under analysis have a reasonable accessibility level, apparently because they can afford motorized modes. A more complete analysis, however, should compare the results obtained here with the income distribution of the population of the distinct city zones, in order to particularly understand how low income people behave in terms of the transportation mode used. One of the interesting outcomes of the application was the identification of regions with particular dynamics, which go against the pattern found in the overall urban area.

Despite the fact that spatial statistics tools are not limited to those used here, the application conducted seems to suggest that their combination with GIS is indeed promising for the analysis of transportation accessibility. This work has also shown that, although one can find commercial software with many of these tools already implemented, this is not really necessary if the GIS package has tools for working with matrices and the user has some basic programming knowledge.

6 ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Brazilian agencies FAPESP (Foundation for the Promotion of Science of the State of São Paulo), CAPES (Post-Graduate Federal Agency), and CNPq (Brazilian National Council for Scientific and Technological Development), which have supported this research in different phases.

7 REFERENCES

Bullen, N. (1997) Linking GIS and multilevel modelling in the spatial analysis of house prices, in Craglia, M and H. Couclelis (eds.), *Geographical Information Research: Bridging the Atlantic*. Taylor & Francis, London, pp. 412-426.

Câmara, G.; V. Correa; J.A. Paiva; A.M.V. Monteiro; M.S. Carvalho; C.C. Freitas; F.R. Ramos and M.C. Neves (1999) *Geoprocessamento: Teorias e aplicações*. http://www.ptr.Usp.br/Cursos/AnaliseEspacial/Livro/cap13-estatespacial.pdf.

Fischer, M. and A Getis (1996) *Recent developments in spatial analysi*s. Springer, New York.

Forkenbrock, D.J.; S. Benshoff and G.E. Weisbrod (2001) *Assessing the social and economic effects of transportation projects*. National Cooperative Highway Research Program, Transportation Research Board, National Research Council, Washington, D.C.

Goto, M.; A.N.R. Silva and J.F.G. Mendes (2001) Uma análise de acessibilidade sob a ótica da eqüidade – O caso da Região Metropolitana de Belém, Brasil. *Engenharia Civil/Civil Engineering*. **10**, pp. 55-66.

Haining, R. (1995) Spatial analysis and GIS, in Fotheringham, A.S. and P. Rogerson (eds.), *Designing spatial data analysis modules for Geographical Information Systems*. Taylor & Francis, London, pp. 45-64.

Levine, N. (1996) Spatial statistics and GIS: software tools to quantify spatial patterns, *Journal of the American Planning Association*. **62** (3), pp. 381-392.

Patkar, V.N. (1997) On integration of analytical techniques and GIS for urban management, *Proceedings of the 5th International Conference on Computers in Urban Planning and Urban Management*, Mumbai, India, December 16-19, 1997, v. 1, pp. 257-268.

Reyes, D.E.L. (1999) Um procedimento para determinação da matriz origem-destino para diferentes modos: método indireto baseado em modelo de escolha discreta. Ph.D. Thesis, Department of Transportation, São Carlos School of Engineering, University of São Paulo, São Carlos, Brazil.

Silva, A.N.R.; A.A Raia Jr. and P. van der Waerden (2000) A comparison of accessibility values and actual travel patterns, *Proceedings of the 14th Congresso de Pesquisa e Ensino em Transportes*, Gramado, Brazil, November 22-24, 2000, v. 1, p. 165-174.

Silva, A.N.R.; N.C.M. Brondino and R.S Lima. A comparison of accessibility measures in a medium sized city of a developing country, *Proceedings of the 7^{th} International Conference on Computers in Urban Planning and Urban Management*, 7, Honolulu, HI, EUA, July 18-20, 2001, in CD-ROM.

Silva, A.N.R.; R.S. Lima; A.A. Raia Jr. and P. van der Waerden (1998) Urban transportation accessibility and social inequity in a developing country, in Freeman, P. and C. Jamet (eds.) *Urban transport policy - A sustainable development tool.* Balkema, Rotterdam, pp. 709-714.