

# A comparison of two methods for the definition of Regional Metropolitan Areas through an application in the North of Portugal

*Rui António Rodrigues Ramos*, Universidade do Minho - Escola de Engenharia, Braga, Portugal,

*Antônio Néelson Rodrigues da Silva*, Universidade de São Paulo - Escola de Engenharia de São Carlos - Departamento de Transportes, São Carlos, Brazil,

*Vasco Peixoto Miranda*, Universidade do Minho - Escola de Engenharia, Braga, Portugal

## ABSTRACT

The objective of this paper is to present a comparison of two methods for the definition of regional metropolitan areas, in a perspective of identification of urban territories with similar characteristics. In this study it is assumed that population density is, in the absence of another measure that describe daily or weekly individual movements, a variable that can be used to evaluate, although indirectly, the activity level and dynamics of the territory. The first method explored here uses ESDA (Exploratory Spatial Data Analyses) techniques for the definition of regional metropolitan areas. The second method is based on an index value that results from a rank of the variable under analysis in a spatial context, both at the local and national (or regional) levels. The methods are implemented in a case study in the north of Portugal, where this is a very opportune issue.

## 1. INTRODUCTION

One of the challenges nowadays faced by planners everywhere is the existence of urban agglomerations that do not respect the official, administrative boundaries. These settlements go beyond the limits of a single city and they often comprise several neighbor municipalities. Many problems arise from this kind of spatial arrangement, such as the definition of the boundaries of these so-called conurbations or metropolitan areas. The methods developed with this aim can vary significantly depending on the input data used for the calculation, what can produce distinct outputs.

Hence, the objective of this paper is to present a comparison of two of these methods for the definition of regional metropolitan areas, in a perspective of identification of urban territories with similar characteristics. Both methods use population data for defining clusters of neighbor census tracts with similar characteristics that are seen as uniform regions and therefore aggregated. In this study it is assumed that population density is, in the absence of another measure that describe daily or weekly individual movements, a variable that can be used to evaluate, although indirectly, the activity level and dynamics of the territory.

The first method explored here uses ESDA (Exploratory Spatial Data Analyses) techniques for the definition of regional metropolitan areas. An analysis of the location of census tracts in the geographical context and in each one of the four quadrants of a Moran's scatterplot makes possible to delimit regions that can be considered uniform regarding the variable under consideration.

The second method is based on an index also associated to each one of the census tracts. The index value results from a rank of the variable under analysis in a spatial context, both at the local and national (or regional) levels. The index value can vary from 0 to 1. Values close to the unit indicate that the area is relevant at both the local and national/regional level. The index values decrease if the area is not so important either at the local or at the national/regional level. Finally, neighbor regions with similar index values can be used to define uniform regions.

The methods are implemented in a case study in the north of Portugal, where this is a very opportune issue. The country has two consolidated major metropolitan regions, which are around the cities of Lisbon and Oporto, but there are certainly other regions that already meet the conditions to be included in the same category, as shown in the present study.

## 2. METHODOLOGY

The first method explores two branches of Spatial Analyses to examine the dynamics of metropolitan regions while establishing their boundaries: Spatial Statistics and Spatial Modeling. In the case of Spatial Statistics the focus is on *Measures of spatial autocorrelation*. Anselin (1995; 1998) defines ESDA as a collection of techniques to describe and visualize spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association, clusters or hot spots, and suggest spatial regimes or other forms of spatial heterogeneity.

As the concept of spatial autocorrelation or spatial association is central to these analyses it has to be clearly defined, as follows. It is the phenomenon where locational similarity (observations in spatial proximity) is matched by value similarity (attribute correlation). This can be observed in two types of spatial data: *geostatistical* data or *lattice* data (Cressie, 1993). The latter is the case of the present study and it consists of a fixed collection of discrete spatial locations (points or polygons).

The Moran scatterplot/map, which are the techniques used for the visualization of global spatial association in a *lattice* approach, are applied in the present study. Moran's scatterplots can be used as means to classify the behavior of each particular zone regarding the value of the attribute under consideration and the mean value of the same attribute for neighbor zones.

The first step of the analysis is the estimation of spatial autocorrelation values, which makes use of three basic elements:

- *Spatial proximity matrix (W)*: matrix of dimension  $n \times n$ , 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;
- *Vector of deviations (Z)*: each element of the vector is obtained by the subtraction of the total mean ( $\mu$ ) 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 global 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).

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

where the superscript  $t$  denotes transposed vector.

The 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;
- If the slope is positive, when the  $Z$  values grow, the  $W_Z$  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  $Z$  values grow, the  $W_Z$  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.

For each zone there is one among four possible classifications that correspond to quadrants of the scatterplot (Figure 1).

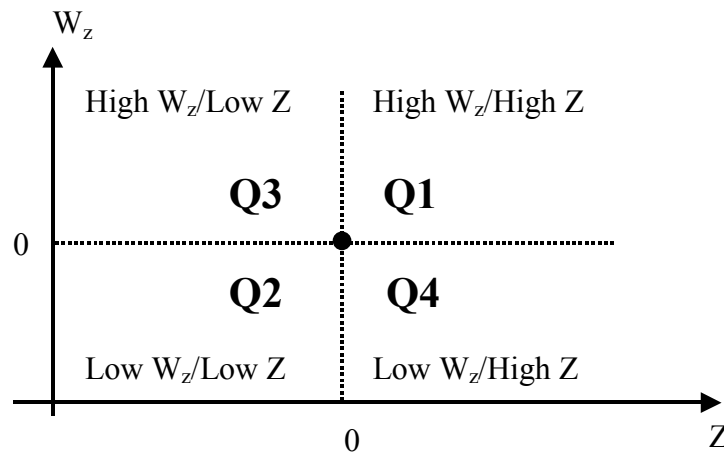


Figure 1 - Interpretation of the Moran's scatterplot

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. 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 Q4, and negative value for the zone and positive average value for neighbors in Q3). 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 can also be drawn. This kind of representation allows the identification of each element according to its classification ( $Z$  and  $W_z$  values), directly by the identification of the quadrant they belong to. In that way, one can visually identify the relationship between the attribute value measured for a particular zone and the average value obtained for its neighbors. The analyses of both the points in the scatterplots and their associated locations in a map give a clear indication of clusters of zones with similar characteristics, which can be seen as uniform regions.

The second approach was originally presented by Office of Management and Budget (1998). This method results from the hierarchization of census tracts, according to its occupation intensity at both the national and regional levels.

The implementation of this method produces an index  $I$  that represents the land parcels ranking at both levels. There is a regional ratio (the order of importance regarding the number of regional parcels) and a national ratio (the relative importance of the land parcel at the national level). Both ratios can vary from 0 (for the land parcel with the lowest order) to 1 (for the land parcel with the highest order).

The index number assigned to any given census tract is determined by multiplying its population density ranking ratio at the regional level by its ranking ratio at the national level. This provides a relative measure of activity intensity for comparative purposes nationwide by taking into account both the national and regional contexts.

Concisely, the calculation of the index  $I$  involve the following steps:

- i. The overall residential population density for each census tracts is calculated by the division of its total population by its total land area;
- ii. All census tracts within a given region are ranked according to population density. The highest-density census tract is assigned the rank N, where N equals the number of census tracts in the region. The second highest-density area is assigned the rank N - 1; third-highest, N - 2; and so forth. For example, if there are 100 census tracts in a region, then the census tract with the highest population density has a rank of 100; the census tract with the second highest population density is 99; etc.
- iii. The region ranking ratio (RRR) of each census tract is calculated by dividing the rank of the census tract by the total number of census tracts in the region, using the following equation:

$$RRR = N [N - 1, N - 2, \dots] / N \quad (2)$$

- iv. After assigning each census tract a ranking ratio within the region, steps 1, 2, and 3 are repeated at the national level. In this iteration, N will represent the number of census tracts within the country;
- v. Each census tract is assigned an index number (I) by multiplying its regional ranking ratio (RRR) and the national ranking ratio (NRR) using the following equation:

$$I = RRR \times NRR \quad (3)$$

This produces an index value that can be used to classify and compare census tracts in terms of population density, and thus relative social and economic importance. The land parcels of higher order at the regional and the national levels present an index value around 1. This value decreases if the order of the land parcels is higher at the regional level and lower at the national level. Finally, land parcels of lower order at both levels presents an index value around 0.

The classification of the land parcels according to their index values and the subsequent visualization of their spatial distribution can be used to identify continuous areas. These areas correspond to patterns of equal territorial occupation (from higher occupation to lower occupation).

### 3. AN APPLICATION IN PORTUGAL

The application of the proposed approaches was carried out in two stages. In the first stage both methodologies were implemented in the entire continental territory of Portugal, while in the second stage just the northwestern part of the country was analyzed. The main sources of information for the analysis here presented were the latest two decennial Censuses conducted in Portugal (INE, 1992; 2002), from which only data about population by *Freguesia* were taken. *Freguesia* is the lower political-administrative level in the country and it also matches the census tracts.

In a first phase, the 1991 dataset was reorganized in order to have the same boundaries of the 4037 *Freguesias* officially registered in 2001. The analyses were developed using built-in tools of the Geographical Information System software ArcView (ESRI, 1996) along with the Spacestat extension (Anselin and Bao, 1997; Anselin and Smirnov, 1998). Spacestat made possible to obtain the *Freguesias* Spatial Proximity Matrix, which was thereafter used in a regular spreadsheet in order to finish the calculations. After the conversion of the absolute data of total population per *Freguesia* into population densities and following the proposed steps for both approaches mentioned in the methodology, the Moran's quadrant and the index I for every *Freguesias* were obtained.

In the maps shown in Figure 2 one can observe the spatial distribution of the population density by *Freguesia*. The red zones in the maps represent values above the national average (555.91 inhabitants for km<sup>2</sup> in 1991 and 522.04 inhabitants for km<sup>2</sup> in 2001) and the grey

zones values below the national average. The map classes highlight the well-known concentration of population around the two more important urban centers of the territory under study, Lisbon (more to the south) and Oporto (more to the north). It shall be noticed, however, the existence of other zones with high values of density (i.e., higher than 1500 inhabitants for km<sup>2</sup>) distributed by the whole territory. In addition, the northwestern part of the territory stands out with an extensive area where the population density is well above the average in both years.

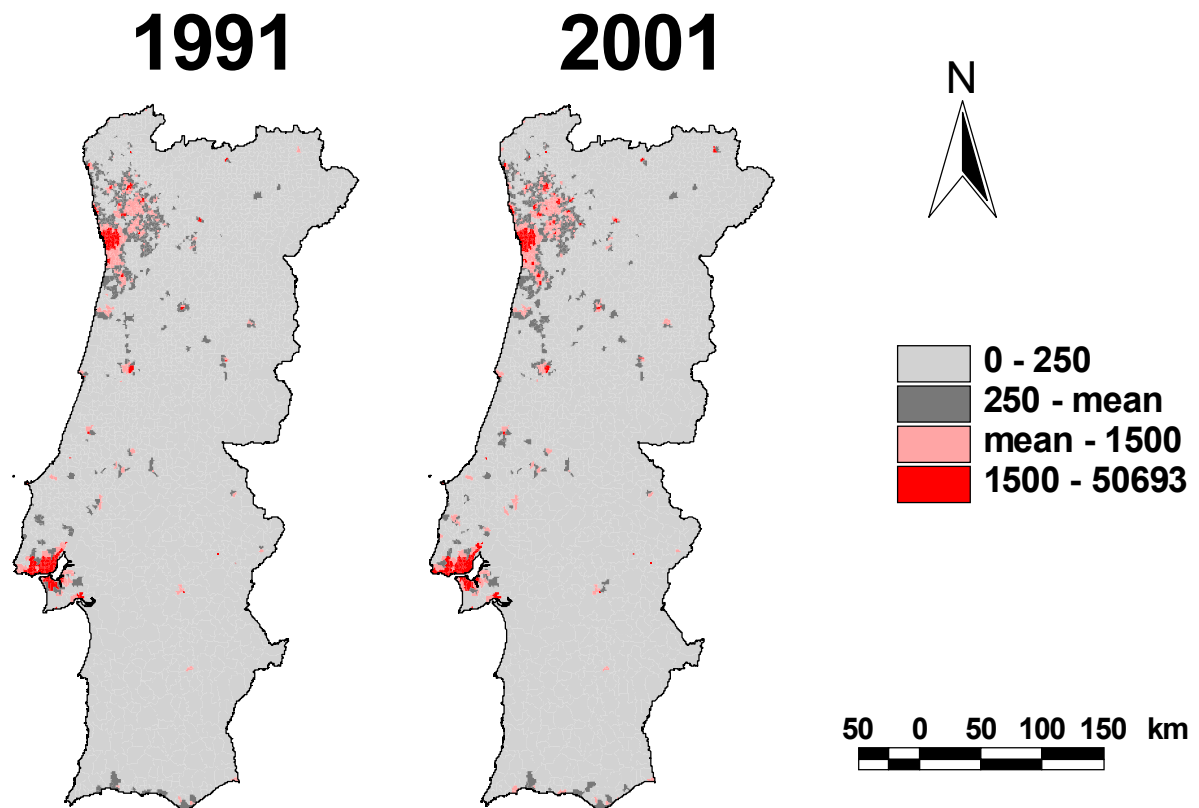


Figure 2 – Thematic maps showing the distribution of the population density variable by *Freguesia* in the years 1991 and 2001.

The maps of Figure 3 show the spatial distribution of the *Freguesias* according to the quadrant classification of the variable population density in the Moran's scatterplots. By the analysis of the maps one can conclude that most *Freguesias* belonging to quadrant 1 are located in two well-defined territorial areas, which are the cities of Lisbon and Oporto (and their surroundings). In those areas a positive correlation exists between the attribute of the zone and the average value of the attribute in neighboring zones, in both cases with the density values in the zones higher than the average value obtained for the whole territory. In addition, the zones belonging to quadrant 2, which are scattered throughout the country, constitute most of the *Freguesias*. Again, there is a pattern of similarity between the density values in the zones and the average density values of neighboring zones. In this case, however, both values are below the average of the whole territory.

It is interesting to analyse the spatial distribution of the zones corresponding to quadrants 3 and 4. Close attention should be given to the zones belonging to quadrant 3, those where the attribute value is below the average value of the entire territory, although the average of their neighboring zones is just the opposite (i.e., higher than the zone value). As a consequence, those zones are likely to move to quadrant 1 in the future due to the "pressure" imposed by the neighboring zones.

An analysis of the pattern changes that took place along the years shows an accentuated growth in all red zones located in the northeast part of Oporto.

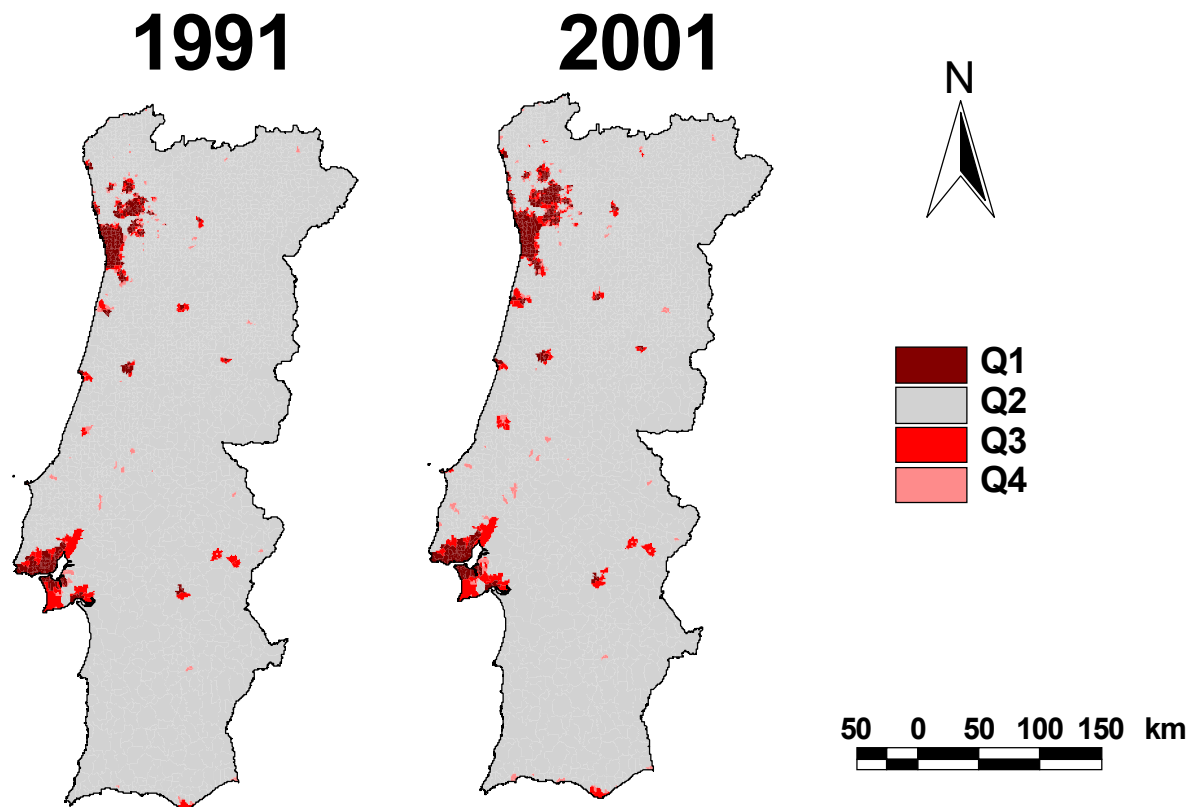


Figure 3 – Thematic maps showing the spatial distribution of the points in Moran's scatterplots based on the population density of the *Freguesias* in 1991 and 2001

In the maps of Figure 4 the spatial distribution of the index defined in the second approach mentioned in the methodology is presented. As the calculated index reflects the importance of each zone at both the national (in this case considering the entire continental territory of the country but excluding the insular areas) and regional levels (in this case the NutsIII statistical subdivisions), there is a larger number of areas in red (i.e., the ones ranking higher). In this analysis, these are the zones in the capitals of district and other *Freguesias* with high density values relative to the regional context. The method also generates, though, an extensive area in the Northwest of the country extending from to the south of Oporto to the East and Northeast. It includes the strongly industrialized areas in the south of Oporto and the valleys of the rivers Ave and Cávado. Other areas ranking high in Figure 4 (i.e., values higher than 0.60) are found along the coast in the central part of the country, as well as in the hinterland. The analysis of the maps presented in Figures 2, 3, and 4 led to a second stage of the case study. Some particularities of the northwestern part of the territory that have been highlighted in the analysis justified an extension of the study in order to look to that region with a deeper level of detail. This sort of more detailed territorial analysis can be used not only to demonstrate the potentiality of the methodologies for identifying uniform urban areas but also to characterize the region under study. This new region encompasses seven NutsIII: Ave, Baixo Vouga, Cávado, Entre Douro e Vouga, Grande Porto, Minho-Lima and Tâmega.

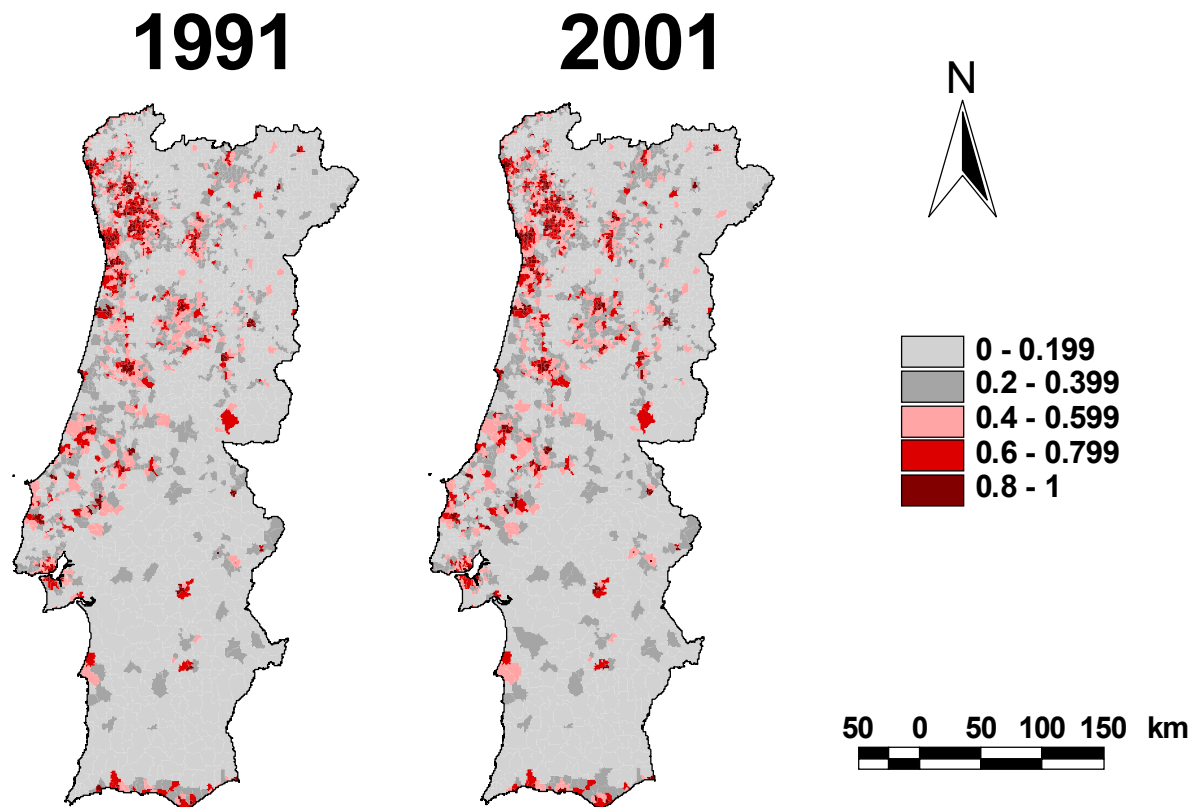


Figure 4 – Thematic maps representing the spatial distribution of the index I based on the variable population density by *Freguesia* in 1991 and 2001

In this phase, the administrative subdivision *Concelho* become the unit of analysis. *Concelhos* are the subdivisions of the country there are just one level above the *Freguesias* in the administrative hierarchy, i. e., every *Concelho* is necessarily formed by one or more contiguous *Freguesias*. This option was taken at this point of the study in order to better reproduce the legal conditions for the aggregation of administrative subdivisions in Portugal. Due to regulation constraints, although two (or more) *Concelhos* can be combined to form a region, it is not possible to join together only *Freguesias* of different *Concelhos*. Accordingly, the next step of the analysis in the case of the first approach was the evaluation of the percentage of *Freguesias* belonging to each one of the four quadrants in the *Concelhos*. This was done only in the restricted study area, as displayed in Figure 5. Similarly, the average value of the index I of the *Freguesias* was calculated in the case of the second method (Figure 7). The values were calculated in both cases for the sixty-five *Concelhos* that constitute the seven NutsIII analysed, which are shown at the right hand side of Figures 5, 6, and 7.

In 1991, the *Concelhos* that belong to the *Grande Porto* NutsIII subdivision were essentially the ones that had more than 60 % of their *Freguesias* in quadrant 1. In 2001, however, the red areas have extended to the northeast, beyond the west border of the *Tâmega* region. In addition, a few isolated *Concelhos* had more than 80 % of *Freguesias* in quadrant 1 in both periods. São João da Madeira (located in the *Entre Douro e Vouga* NutsIII subdivision) in 1991 and 2001, and Vizela (located in the *Ave* NutsIII subdivision) just in 2001.

The two maps show that the main urban agglomeration of the study area is located in the *Grande Porto* subdivision, although it has grown to the neighbor municipalities located in the east, northeast and southeast of it. This area contrasts with the *Minho-Lima* region, where none of the *Concelhos* had more than 20 % of *Freguesias* in quadrant 1. The same is happening in a significant part of the *Baixo Vouga* and *Cávado*, and in the interior municipalities of *Ave*, *Tâmega* and *Entre Douro e Vouga*.

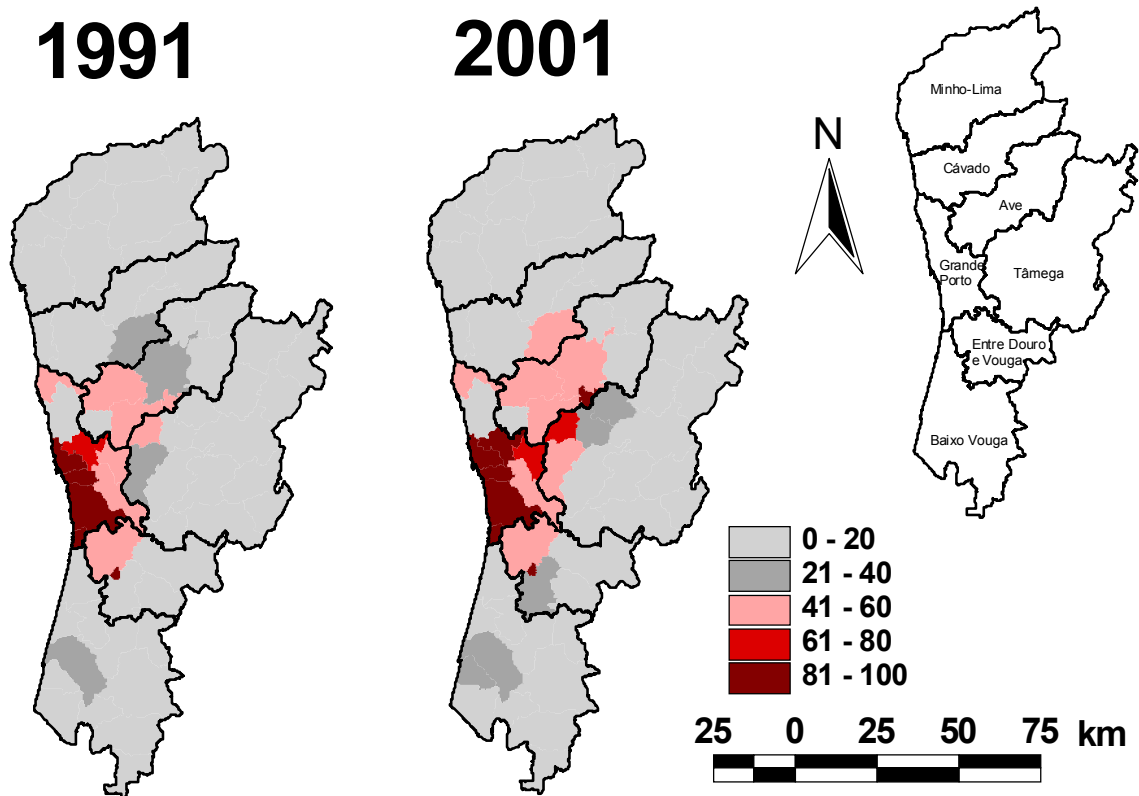


Figure 5 – Thematic maps showing the percentage of *Freguesias* in quadrant 1 in the *Concelhos* located in the northwestern part of Portugal

Furthermore, considering that the *Freguesias* belonging to quadrant 3 constitute potential areas of urban growth, those *Freguesias* were added to the figures represented in Figure 5. New maps were then generated in order to show the proportions of *Freguesias* in quadrants 1 and 3 for each *Concelho*, in terms of their spatial distribution, see figure 6. The subsequent analysis is in everything similar to the previous one done for the data displayed in Figure 5. An important difference in this case is the fact that *Ilhavo*, which is one of the *Concelhos* located in *Baixo Vouga*, had more than 60 % of its *Freguesias* in quadrants 1 or 3 in 2001. Also interesting is that *Ilhavo* is not adjacent to the urban area of *Grande Porto*. Conversely, it is located 30 km in the south direction.

In the analysis done for the index I, the average of the values obtained in the *Freguesias* was assumed as the representative value for each *Concelho*. The spatial distribution of those values is represented in the maps of Figure 7, where five main urban areas are evident: *Porto* and *Matosinhos* in *Grande Porto*, *Braga* in *Cávado*, the west part of *Tâmega* and *Vizela* in *Ave*, and the west part of *Entre Douro e Vouga* and the north parts of *Baixo Vouga* and *Ílhavo* in *Baixo Vouga*. Although these five areas are not contiguous, they are somehow connected by other *Concelhos* with I values ranging from 0.4 to 0.6.



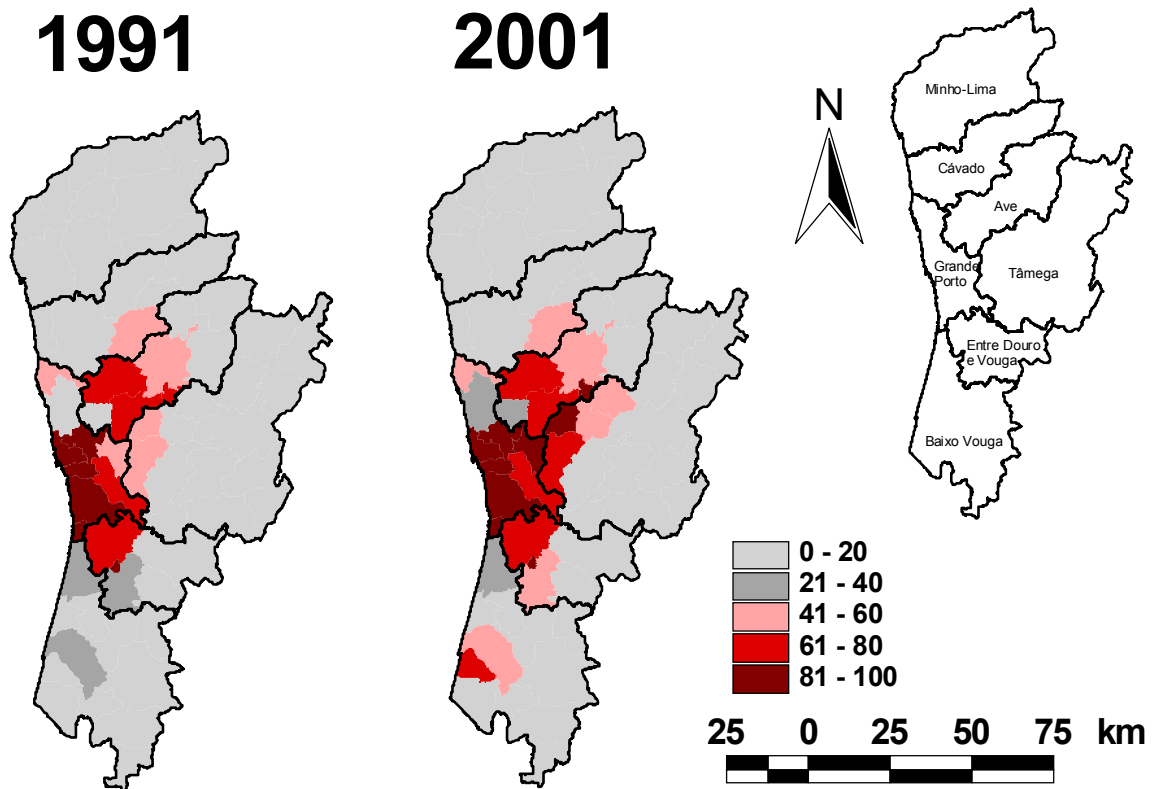


Figure 6 – Thematic maps showing the percentage of *Freguesias* in quadrants 1 or 3 in the *Concelhos* located in the northwestern part of Portugal

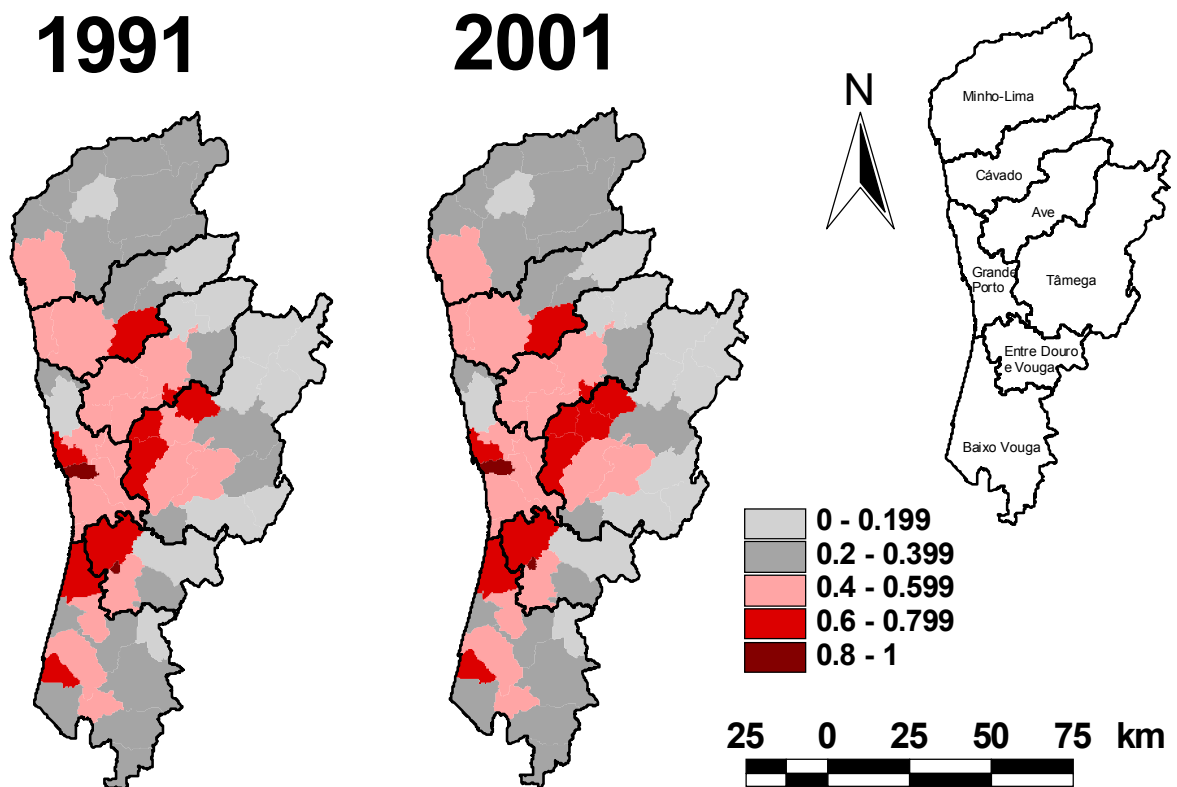


Figure 7 – Thematic maps representing the spatial distribution of the average value of the index I found in the *Concelhos* located in the northwestern part of Portugal

#### 4. CONCLUSIONS

Unlike the usual political options, the methodologies here proposed privilege a context of territorial administrative construction in that the space continuity is one of the most relevant aspects. Both methodologies favor the analysis of the spatial arrangement of the variable under investigation. The maps showing the spatial distribution of the zones in specific quadrants and of the indexes obtained for the zones translate with clarity the form of occupation of the territory and enhance the larger or smaller homogeneity and space continuity, although in this case limited to only one variable, which was population density.

The application of the methodologies to the Northwest part of Portugal led to interesting results. A comparison of the results obtained with the need of defining new borders for the metropolitan areas in the Northwest region of Portugal can suggest two possible alternatives of action. The first one maintains the current metropolitan area of *Oporto* and aggregates contiguous urban municipalities located in the west parts of *Cávado*, *Ave*, *Tâmega* and *Entre-Douro e Vouga* and in the north part of *Baixo Vouga*. The second one results in the definition of multiple metropolitan areas, one in *Grande Porto*, another one in the *Ave and Cávado* regions, other in the west part of *Tâmega* and a last one in the west part of *Entre-Douro e Vouga* and the north region of *Baixo Vouga*. The first solution, which was already proposed by Ramos and Silva (2003), seems to be the most interesting in territorial terms, because it will allow the creation of a large area with capacity to answer to the paradigm of the great metropolitan areas. That means that the area is able to answer not only to the internal needs of the country, but also and preferably it contains the elements for an effective relationship with other metropolitan areas at a global world scale (Ascher, 1995).

Finally, it is important to highlight that the proposed methodologies are promising options for the intended objective of having the integrated and holistic approach necessary for the analysis of territory variables.

#### ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the Brazilian agencies CAPES (Post-Graduate Federal Agency), FAPESP (Foundation for the Promotion of Science of the State of São Paulo), and CNPq (Brazilian National Council for Scientific and Technological Development), and to the Portuguese agency GRICES (Office for International Relations in Science and Higher Education of the Portuguese Ministry of Science), which have supported our efforts for the development of this work in different ways and periods.

#### REFERENCES

- Anselin, L. (1995) Local indicators of spatial association - LISA, *Geographical Analysis*, Vol. 27, 93-115.
- Anselin, L. (1998) Interactive techniques and exploratory spatial data analysis. In P. Longley, M. Goodchild, D. Maguire and D. Rhind (eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*. Geoinformation International, Cambridge.
- Anselin, L. e Bao, S. (1997) Exploratory spatial data analysis linking SpaceStat and ArcView. Em M. Fisher and A. Getis (eds.), *Recent Developments in Spatial Analysis*. Springer-Verlag, Berlin.
- Anselin, L. e Smirnov, O. (1998) *The SpaceStat extension for ArcView 3.0*. Regional Research Institute, West Virginia University, Morgantown.
- Ascher, F. (1995) *Métapolis: ou l'avenir des villes*. Éditions Odile Jacob, Paris,

- Cressie, N. (1993) *Statistics for Spatial Data*. Wiley, New York.
- ESRI (1996) *ArcView GIS, The Geographic Information System for everyone, Using ArcView GIS*. Environmental Systems Research Institute, Redlands-CA.
- INE (1992) *Recenseamento da População e da Habitação (Portugal) – Censos 1991*. Instituto Nacional de Estatística, Lisbon.
- INE (2002) *Recenseamento da População e da Habitação (Portugal) – Censos 2001*. Instituto Nacional de Estatística, Lisbon.
- Levine, N. (1996) Spatial statistics and GIS: Software tools to quantify spatial patterns, *Journal of the American Planning Association*, Vol. 62, No. 3, 381-392.
- Office of Management and Budget (1998) Alternative Approaches to Defining Metropolitan and Non-metropolitan Areas, *Federal Register*, Vol. 63, No. 244, December 21.
- Ramos, R.A.R. e Silva, A.N.R. (2003) A Data-driven Approach for the Definition of Metropolitan Regions, *CD-ROM Proceedings of the 8<sup>th</sup> International Conference on Computers in Urban Planning and Urban Management - Reviewed Papers*. Sendai, Japan.