A DATA-DRIVEN APPROACH FOR THE DEFINITION OF METROPOLITAN REGIONS

Rui António Rodrigues RAMOS
Assistant Professor
Department of Civil Engineering
School of Engineering
University of Minho
Campus of Gualtar
4710-057, Braga – Portugal
Tel: +351 253 604720
Fax: +351 253 604721
E-mail: rui.ramos@civil.uminho.pt

Antônio Nélson RODRIGUES DA SILVA
Associate Professor
Department of Transportation,
São Carlos School of Engineering
University of São Paulo
Av. Trabalhador São-carlense, 400
13566-590, São Carlos – SP – Brazil
Tel: +55 16 273 9595
Fax: +55 16 273 9601
E-mail: anelson@sc.usp.br

Abstract: The objective of this paper is to present a data-driven approach for the definition of metropolitan regions. The proposed approach, which constitutes an option to avoid the endless confrontations that may be derived from the essentially subjective political criteria, explores two branches of Spatial Analyses: Spatial Statistics and Spatial Modeling. Spatial Statistics tools are used to identify the characteristics of local association and combined with Cellular Automata techniques in order to build prediction models. The analyses conducted with Exploratory Spatial Data Analyses (ESDA) tools and census data give a clear indication of clusters of zones with similar characteristics, which can be seen as uniform regions. Spatial dynamic models can then be used to foresee the global behavior of regions in terms of growth, although based on local (and historical) relationships among zones. The proposed approach is tested in a case study carried out in Portugal, where this is a timely issue.

Keywords: metropolitan regions, spatial statistics, spatial modeling, ESDA, urban management

1. INTRODUCTION

One of the challenges faced by urban managers nowadays is the fact that the administrative boundaries of the municipalities are often different from the limits of the urbanized areas in which they are comprised. In some cases, central governments take over the management of eventual conflicts produced by the excessive proximity of municipalities in that geographical context. In other cases, though, they try to transfer that responsibility to mid-level administrative divisions created only for that purpose, the so-called metropolitan regions. The definition of the boundaries of these new regions, however, is usually not an easy task, mainly due to various incompatible political and economical interests of the parties involved.

The focus of the problem discussed here is the definition of the metropolitan regions in spatial terms, considering that they are larger than any single clustered city but often smaller than the upper administrative divisions (e.g., province, state, or region). This is not a recent problem and therefore it is not difficult to find discussions about the issue and methods for such a definition in the literature, as in the works of NUREC (1994), Lacour and Puissant (1999), Ferreira and Rosado (1999), and Ferrão *et al.* (2002). A number of authors even carried out theoretical applications of some of the reported methods for defining boundaries of metropolitan regions in the main Portuguese regions, namely Lisbon and Oporto (Ferrão and Vala, 2001; and Ferrão *et al.*, 2002). They used the differences in terms of area, population, and employment resulting from the application of various methods to point out the main problems of the approaches investigated.

There is also literature about the topic available in the United States (e.g., Office of Management and Budget, 1990; Office of Management and Budget, 1998; Office of

Management and Budget, 1999; Metropolitan Area Standards Review Committee, 2000; and Office of Management and Budget, 2000), where the need of constant review of metropolitan area standards is often highlighted in order to ensure their continued usefulness and relevance. The current standard for defining metropolitan areas in the United States is essentially a commuting-based county-level approach (Office of Management and Budget, 2000), which was adopted after a discussion of four initial alternative approaches. They were: a commuting-based county-level approach, a commuting-based tract-level approach, a directional commuting tract-level approach and a comparative density county-level approach (Office of Management and Budget, 1998). It is important to observe the presence of the population density as an alternative to defining metropolitan areas, since reliable and up-to-date commuting data are very hard to find in some countries. This was also defended by the authors of Office of Management and Budget (1998), who stated that "*residential population density can serve as a surrogate for other measures of activity in the absence of nationally consistent and reliable data sets describing all daily and weekly movements of individuals*". As this is the case of most Portuguese urbanized areas, it is a good reason to consider the use of population density for defining metropolitan areas in that context.

The objective of this paper is to present a data-driven approach for the definition of metropolitan regions, which is based on the strength of spatial association of areas concerning specific attributes of the territory (e.g., population distribution). Hence, it constitutes a contribution to the discussion and another option to avoid the endless confrontations that may be derived from the essentially subjective political criteria. The proposed approach explores two branches of Spatial Analyses: Spatial Statistics and Spatial Modeling. In such a way, Spatial Statistics tools, which are used to identify the characteristics of local association, are combined with Cellular Automata techniques, in order to build dynamic models of the studied regions.

Exploratory Spatial Data Analyses (ESDA) tools allow the estimation of the strength of local association for certain attributes of the studied areas in discrete time periods that correspond to regular basis (usually decennial) census surveys. 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. In such a classification scheme, four possibilities are considered, which correspond to quadrants in the scatterplot. The analyses of both the points in the scatterplots and their associated locations in maps give a clear indication of clusters of zones with similar characteristics, which can be seen as uniform regions. The evaluation of these characteristics of the same region in different periods of time can then be used to build a prediction model. As it takes into account the future changes in the regions, this model can be used for delimiting metropolitan region, as suggested here.

The proposed approach is tested in a case study carried out in Portugal, where this is a very opportune issue. Although the country has two consolidated metropolitan regions, which are around the cities of Lisbon and Oporto, the application presented here is focused only on Oporto due to its continuous and strong expansion. The theoretical background for the proposed approach is discussed in part 2 of this paper, which is followed by the presentation of the application to Portugal in the period 1991-2011 in part 3, in which the specific analyses for the metropolitan region of Oporto are also included. The paper ends with some concluding remarks about the proposed method and suggestions for its improvement, just prior to the list of references.

2. THE PROPOSED APPROACH

The study carried out here 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.* According to Levine (1996), these measures describe the relationship among the different locations for a single variable, therefore indicating the degree of concentration or dispersion. In the particular case of this study, however, the interest is not only on global spatial autocorrelation measures, but also and more specifically on local spatial autocorrelation measures (Anselin, 1996; Serrano and Valcarce, 2000). As suggested by Anselin (1998a) when discussing methodological and technical issues associated with the implementation of Exploratory Spatial Data Analysis (ESDA) in a geocomputational environment, the emphasis is here on techniques that explicitly take into account the presence of spatial autocorrelation, such as visualization devices for spatial distributions and spatial association, and local spatial association.

Anselin (1995; 1998b) 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).

Anselin (1998a) presented four broad classes of ESDA techniques: visualizing spatial distributions, visualizing spatial association, local indicators of spatial association, and multivariate indicators of spatial association. 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. In such a classification scheme, four possibilities are considered, which correspond to quadrants in the scatterplot.

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 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;
- *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 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).

$$I = \frac{Z^t W_Z}{Z^t Z} \tag{1}$$

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 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.

In order to better observe those relationships the values of W_z x Z were plotted here in a graph. 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 values can be identified (Figure 1). The two lines divide the space in four quadrants, described here as Q1, Q2, Q3, and Q4.
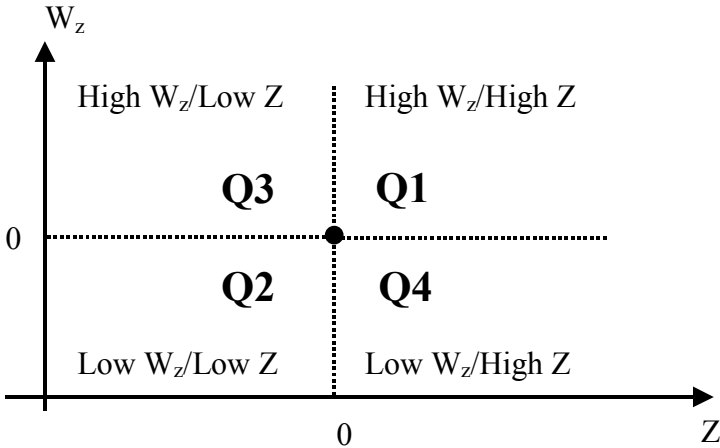

Figure 1 Interpretation of the Moran scatterplot

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 location in a map give a clear indication of clusters of zones with similar characteristics, which can be seen as uniform regions.

4

These homogeneous regions, however, do not necessarily represent, in the case considered, metropolitan regions, unless there is a previous knowledge of what is a positive aspect for the considered variable. In the case of population density, for instance, the zones likely to constitute a higher hierarchy region are those in which the value of the attribute in each zone is high and the mean value of the attribute to its neighbor zones is also high. Another potential problem of this simple view could come from a unique observation of the data in time. However, the regularity of census surveys offers a reliable alternative for long-term analyses, if carried out in the way aforementioned, because the analyses are not directly based on attribute values but on the association of zones according to their attribute relationships. Therefore the analyses become atemporal. This very same characteristic of the method is used to build the spatial dynamic model proposed here, which can be used to foresee the global behavior of regions, although based on local (and historical) relationships among zones.

Under a rigorous examination of the recent literature concerning CA theory (e.g., O'Sullivan and Torrens, 2000; Torrens, 2000; Torrens and O'Sullivan, 2001), our prediction model might not exactly fit into the class of Cellular Automata models, although we have borrowed from them some of the most important concepts applied here. The proposed models use, for example, the data coming from the Moran scatterplot to identify the strength of the spatial association among neighbors when building transition rules. This process is similar to what is applied in CA models in two aspects: in the use of the transition rules and in the implicit assumption that local relationships can generate global patterns. On the other hand, our model differs from ´traditional` CA models, for example, in the treatment given to the representation of space. In our case, instead of having the usual regular lattice of identical cells as the representation of the space we have used administrative areas as ´cells`. In addition, our model is not really dynamic, since it has been used to simulate only one time step ahead of the latest period used in its construction. However, it could be used (at least in theory) to produce more simulations. That is the reason why we called it a dynamic model, despite the difficulties in setting it up for sequenced simulations.

3. AN APPLICATION IN PORTUGAL

The application of the model proposed was divided in three phases. The first phase was the spatial analysis of the existing conditions in at least two past time periods based on the Moran scatterplot/map described in part 2. The outcome of the analysis (i.e., essentially the distribution of the areas in the quadrants and the quadrant changes in the two periods) provided the inputs to build the transition rules for the second phase, which was the construction of a prediction model and its application for generating a future scenario. Finally, the third phase was the comparison of the distribution of the zones in the quadrants in the future scenario with the legal boundaries of the subdivisions that have to be considered in the definition of the metropolitan region. The analysis carried out in the last step was based on the proportion of elements in quadrants 1 and 3 that have been observed in the first time period. These steps are described in the sequence, after a brief introduction about some particularities of the territory subdivisions in Portugal.

3.1 Territory subdivisions in Portugal

Before starting the description of the application, there are some important points concerning the territory subdivisions in Portugal that must be clarified. The definition of the actual boundaries of the metropolitan regions in Portugal has as a constraint the fact that they have to be coincident with the outer boundaries of the cluster of *Concelhos* they encompass. *Concelhos* are subdivisions of the country, which are one level below the *Distritos* and one level above the *Freguesias*. This is better understood by examining Figure 2, in which these

three spatial subdivisions are depicted for the region around the city of Oporto. The region shown in more detail three times is an enlarged view of the area highlighted in the map of the country placed in the upper part of Figure 2. The *Distritos* can be seen on the lower left side of Figure 2 (marked with *a* and containing 3 subdivisions: Braga, Porto e Aveiro). Next, the *Concelhos* are marked with *b* and placed in the center of the figure. Finally, the *Freguesias* are on the lower right side of Figure 2 and marked with *c*. Only the *Concelhos* and the *Freguesias* are political–administrative levels, in which the *Concelhos* are the municipalities even though they are not always entirely urban. There is also another important point to be highlighted on Figure 2, which is the actual Metropolitan region of Oporto. That is easily recognized in the lower middle part of the figure (part *b*) as a cluster of gray-painted *Concelhos*.
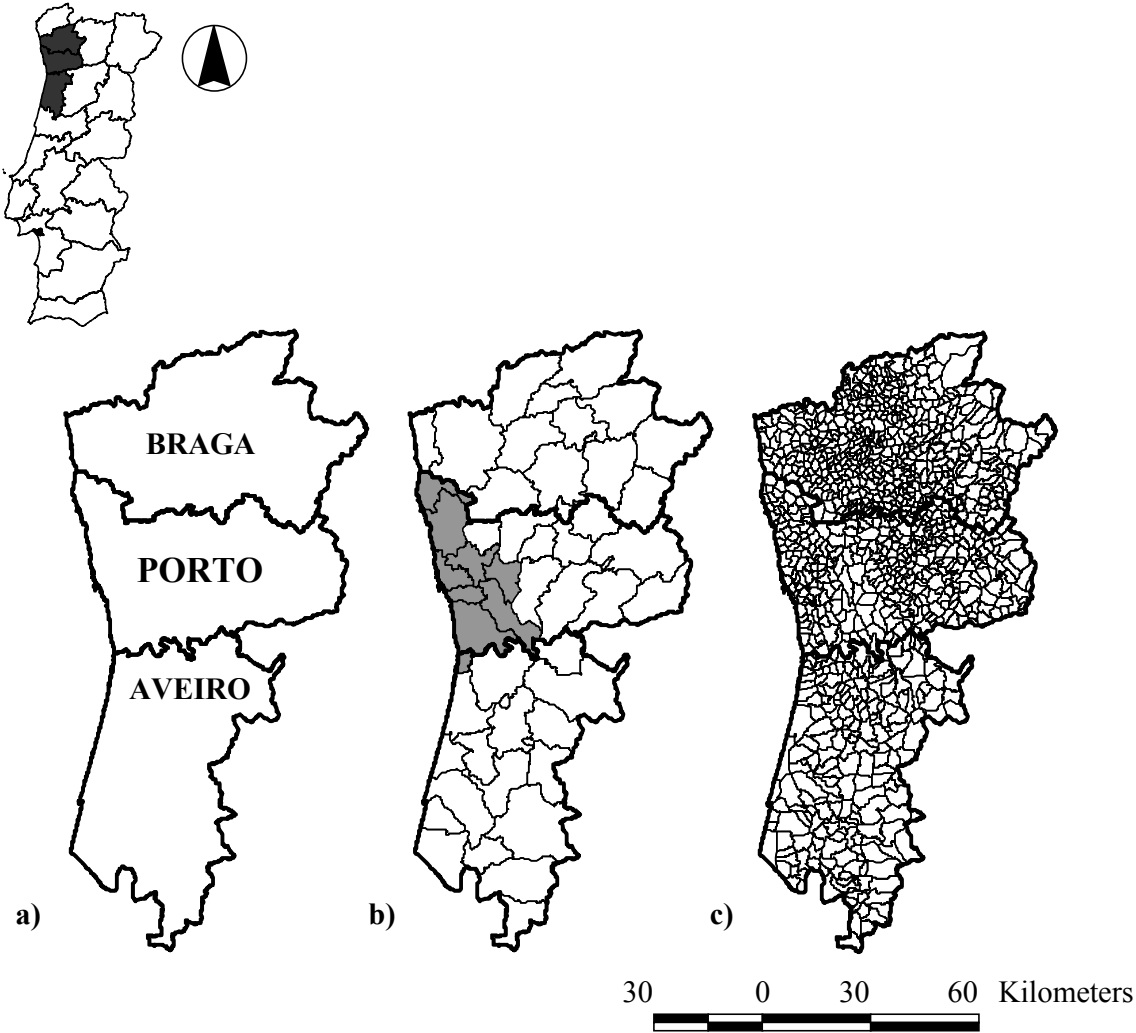
Figure 2 Spatial subdivisions in Portugal: a) *Distritos*; b) *Concelhos*; and c) *Freguesias*

3.2 Spatial Analyses

The main sources of data for this study were the decennial Censuses conducted in Portugal in the years 1991 and 2001 (INE, 1992; 2002), from which only data about population was taken. At first, the 1991 dataset was regrouped in order to match the 4037 *Freguesias* registered in 2001, so that the spatial divisions became coincident in both time periods. The analyses were carried out with the tools available in the GIS software package *ArcView* (ESRI, 1996) along with the *Spacestat* extension (Anselin and Bao, 1997; Anselin and Smirnov, 1998) - for

6

computing the spatial proximity matrix, and a regular spreadsheet. After converting the total values of population per *Freguesia* into population densities and following the steps described in section 2, the results obtained were summarized in two graphs and two maps as shown in Figures 3 and 4, respectively.

The general Moran´s I index obtained for the variable population density in Portugal in the years 1991 and 2001 with equation (1) was respectively 0.7359 and 0.7271. The values indicate a high spatial autocorrelation for that variable in both periods. Figure 3 shows the distribution of the Z and W_z values in the two years, 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. In those quadrants (particularly in Q2, where more than 3000 points are clustered very close to the crossing lines in the middle of both graphs) were most of the points for both periods investigated. This can be seen by the numbers placed in the quadrants of the graphs in Figure 3. The location of all regions to which the points in the scatterplot are referring to can be seen in the thematic map of Figure 4.
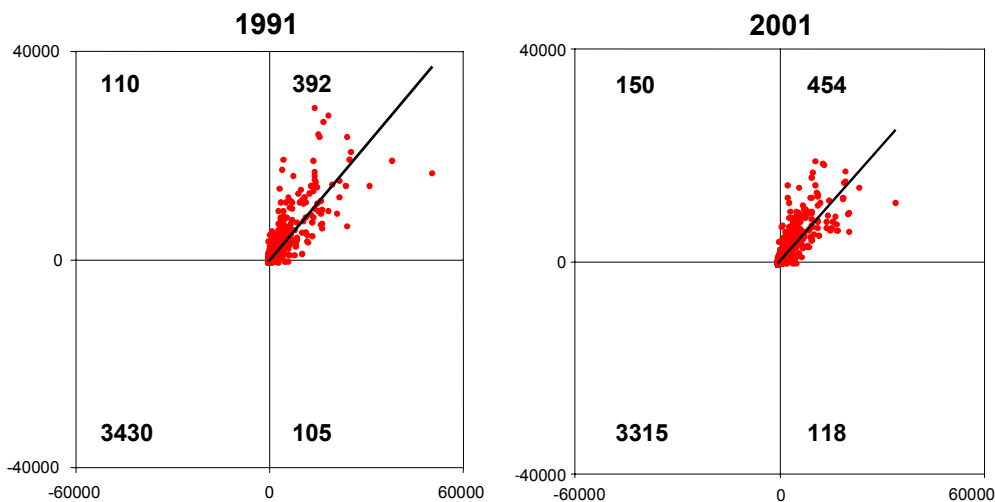

Figure 3 Moran`s scatterplots for the variable population density in Portugal in the years 1991 and 2001

In the maps shown in Figure 4 one can observe that most regions associated to points in quadrant 1 were geographically located in two areas close to the west shore. These are exactly the regions where the existing metropolitan regions of Lisbon and Oporto are located. In those zones there was 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 was similar to the neighborhood average value. For zones in quadrant 1 the deviation value per zone and the average of deviations were both positive, what indicated that the population density was higher than the average value measured for the entire country. Hence, it was not really a surprise that most points in quadrant 1 were found within the boundaries of the official metropolitan regions, which are also depicted in Figure 4. Points in Q2 constituted the majority of the results and they were spread all over the country, although highly concentrated on the graphs. Again, there was a 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 were both negative. Thus, the population density was lower than the average value measured for the entire country.

It is interesting to observe, also in Figure 4, the spatial distribution of points in quadrants 3

and 4. Moreover, there was a special interest here for points in Q3, in which the attribute value of the zone was smaller than the average value of the surrounding area. Due to its propensity to change to quadrant 1, zones in quadrant 3 were particularly important for the prediction model, as shown in the next step.
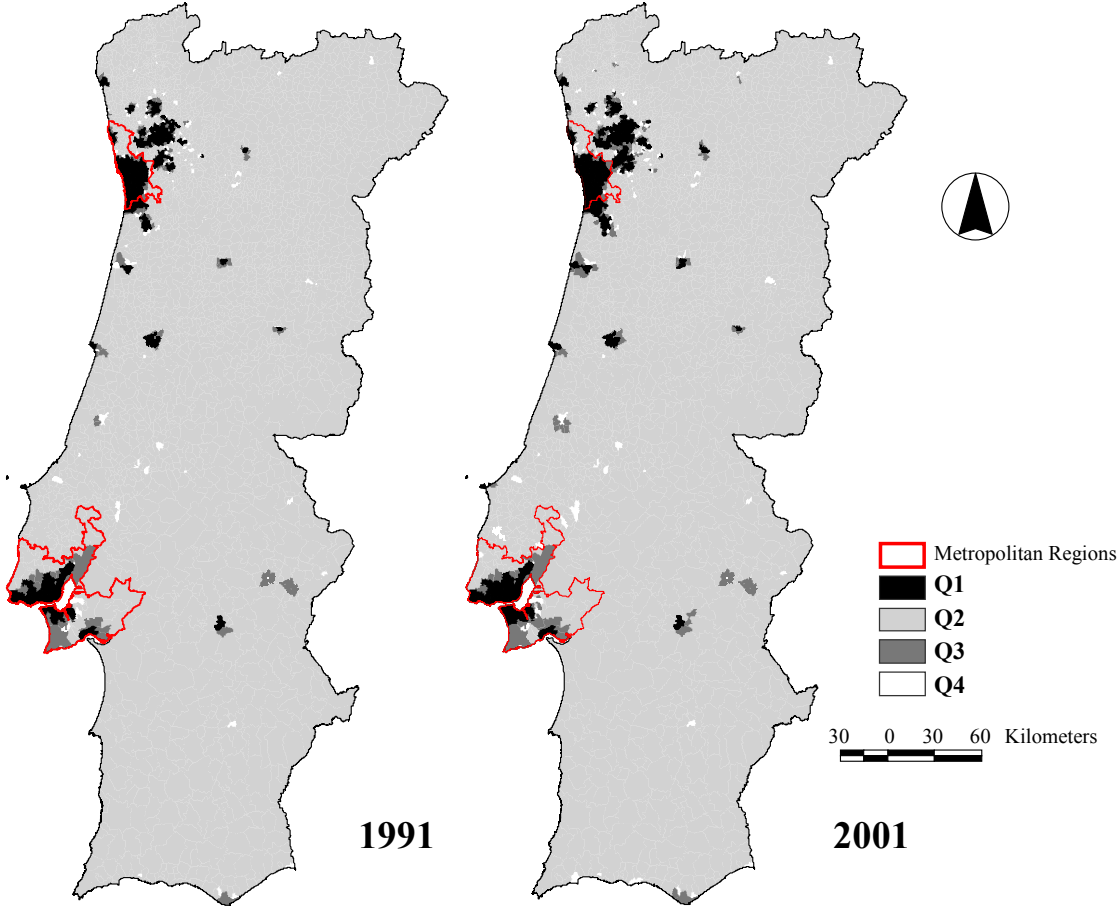

Figure 4 Thematic maps showing the distribution of points in the Moran scatterplot based on the variable population density in Portugal for the years 1991 and 2001

3.3 The Prediction Model

The prediction model was based on the transitions observed in the two past periods analyzed. These transitions through time can be seen as quadrant changes, as summarized in the transition matrix shown in Table 1. Next, the transition classes were identified by a code composed of numbers associated to the origin quadrant and the destination quadrant. In the case of the 11 zones that changed from Q2 to Q1 (Table 1), for example, they have been identified with the code 21. The number of neighbors in each quadrant was subsequently identified for each of these transition classes. These figures, along with the population density value and the average value of the same attribute in immediate neighbors, where used to build the transition rules for the prediction model.

An example of the data used in the definition of the transition rules can be seen in Table 2, in which only the transition classes with origin in Q2 are represented. The absolute numbers of neighbors per quadrant were converted to percentages of the total number of neighbors for each transition class. Another change was the combination of the population density values and the average values of neighbors in a ratio, as shown in the last column of Table 2.

Table 1 Transition matrix in the period 1991-2001

	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Quadrant 1	392	--	--	--
Quadrant 2	11	3314	70	35
Quadrant 3	30	--	80	--
Quadrant 4	21	1	--	83

Table 2 Conditions observed for zones moving from Q2 in 1991 to all quadrants in 2001

Transition Class	% of Neighbors in Q1	% of Neighbors in Q2	% of Neighbors in Q3	% of Neighbors in Q4	Average Pop. Density of Neighbors/Pop. Density of Zone
21	0.1452	0.5000	0.1613	0.1935	1.1196
22	0.0046	0.9714	0.0096	0.0144	1.1353
23	0.1540	0.5868	0.0929	0.1663	1.6219
24	0.0544	0.7826	0.0978	0.0652	0.6726

A careful examination of the conditions described in Table 2 for all transition classes resulted in a set of transition rules for zones with origin in all four quadrants. In the case of zones moving from Q2, for example, the transition rules adopted were the following:

If % of Neighbors in Q2 > 79%
Then 22 (i.e., zones moving from Q2 to Q2)
Else **If** Average Pop. Density of Neighbors/Pop. Density of Zone < 1.0
Then 24 (i.e., zones moving from Q2 to Q4)
Else **If** Average Pop. Density of Neighbors/Pop. Density of Zone < 1.5
Then 21 (i.e., zones moving from Q2 to Q1)
Else 23 (i.e., zones moving from Q2 to Q3)

The first transition rule resulted from the analysis of the third column (% of Neighbors in Q2) of Table 2. The zones that remained in Q2 had on average a higher percentage value (0.9714) than the others (0.5000, 0.5868, 0.7826). The threshold value for that transition was selected by experimentation having the value 79% as a lower limit because of the value 0.7826 also found in the third column of Table 2.

The second transition rule resulted from the analysis of the last column (Average Pop. Density of Neighbors/Pop. Density of Zone) of Table 2. The value concerning the transition class 22 was no longer relevant after the implementation of the first rule. Thus, among the remaining values attention was drawn to the value associated to the transition class 24 (0.6726), which was significantly lower than the other values. The strategy to select the threshold values for the transition from Q2 to Q4 was then similar to what was previously

done with the transition class 22. In that case, though, the upper limit was found to be 1.0. The third transition rule resulted from the analysis of the same column of Table 2 (Average Pop. Density of Neighbors/Pop. Density of Zone). The values concerning the transition classes 22 and 24 became then irrelevant after the implementation of the first and second rules. Therefore, zones whit values smaller than 1.5 were supposed to change from Q2 to Q1 while all other remaining zones leaving Q2 would go to Q3.

The combination of the distinct sets of transitions rules applied to all quadrants (and not only those mentioned in the example above described, for zones that were originally classified as Q2) resulted in a model able to match 95% of the quadrants found in 2001. The model was then used to build a scenario for the year 2011. The changes projected from 2001 to 2011 are summarized in the transition matrix of Table 3 and the resulting scenario is presented as a map in Figure 5. Again, the boundaries of the actual metropolitan regions are left in the map for comparison with the distribution of zones in the quadrants.
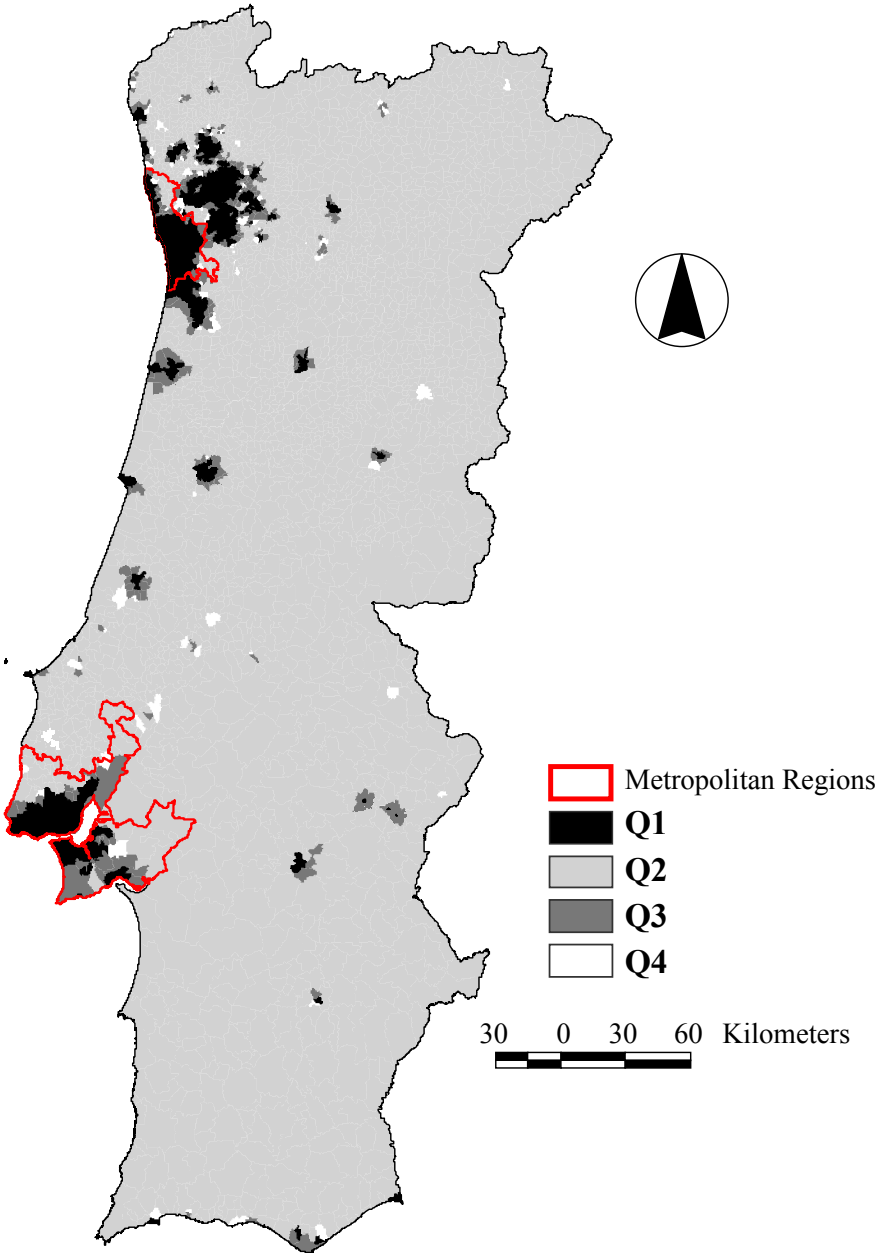

Figure 5 Thematic map showing the distribution of points in the Moran scatterplot based on the variable population density in Portugal for the year 2011

Table 3 Transition matrix in the period 2001-2011

	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Quadrant 1	454	--	--	--
Quadrant 2	--	3128	142	45
Quadrant 3	28	--	122	--
Quadrant 4	64	--	--	54

3.4 The Definition of the Metropolitan Region

The final step of the method proposed for the definition of the metropolitan regions was an analysis of the changes that took place during the periods under consideration in the areas that are contiguous to the existing metropolitan regions. In the case studied here, the analysis focused on the metropolitan region of Oporto, given that the Lisbon metropolitan region was since 1991 large enough to incorporate nearly all Q1 and Q3 areas in 2001 and 2011. On the other hand, the conditions around the metropolitan region of Oporto are exactly the opposite. In the case of Oporto, the official metropolitan region did not include several adjacent areas that had a strong spatial association with it already in 1991 and 2001, as can be seen in the detailed views of Figure 6. The picture is even more evident when looking to the projected scenario in the 2011, when several areas in quadrants 1 and 3 appear clustered together around the initial focus area. The problem here lies then in the identification of the *Concelhos* that should be incorporated to the metropolitan region when adjusting its boundaries to reflect the changes through time.
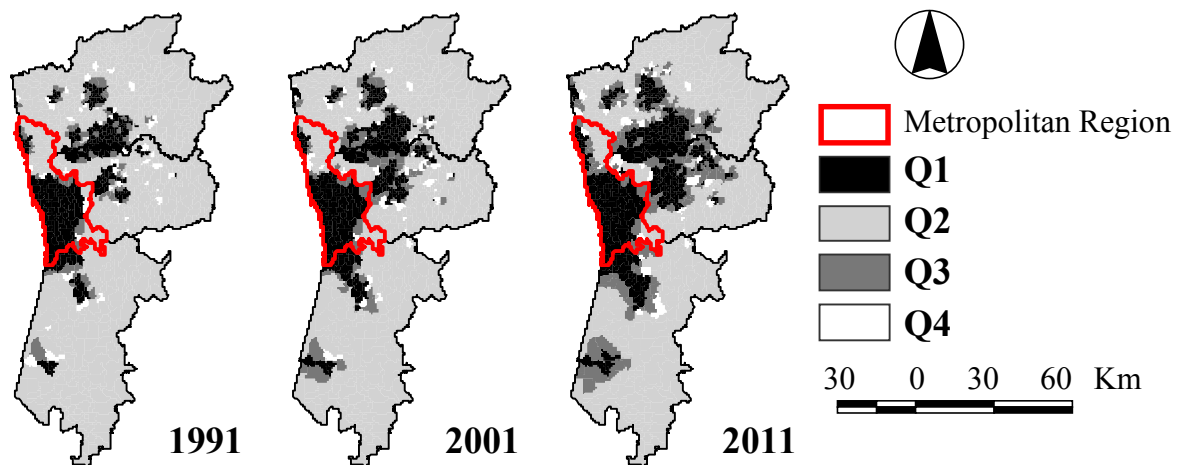

Figure 6 Thematic maps showing the distribution of points in the Moran scatterplot based on the variable population density in Portugal for the years 1991, 2001 and 2011

The process of selecting the *Concelhos* that should become part of the metropolitan region was set next. It was based on the proportion of Q1 and Q3 areas that were part of the original region in 1991, as shown in the darkest bar on the right hand side of Figure 7. The other two bars correspond to the proportions of elements in each quadrant in the periods of 2001 and 2011 put in the graph just for comparison. The aggregation of *Concelhos* that led to values of Q1 and Q3 close to the 70 % found in 1991 is represented in Figure 8, accordingly to the values shown in Figures 9 and 10 for the years 2001 (gray ´1 + 3` bar) and 2011 (white

11

´1 + 3` bar), respectively.

It is important to notice that this is not simply the definition of clusters of regions with similar population density values. Differently, when aggregating zones that belong to the same quadrants, the spatial relationships beyond the immediate vicinity of the individual zones are implicitly taken into account. There may be a problem, however, caused by the size of the spatial units under consideration. Local variations of the population distribution patters within the spatial units may not be revealed if the spatial units are too large. Smaller spatial units would reduce that problem, but their use can make the entire calculation process more cumbersome if the number of elements for calculation becomes too large.
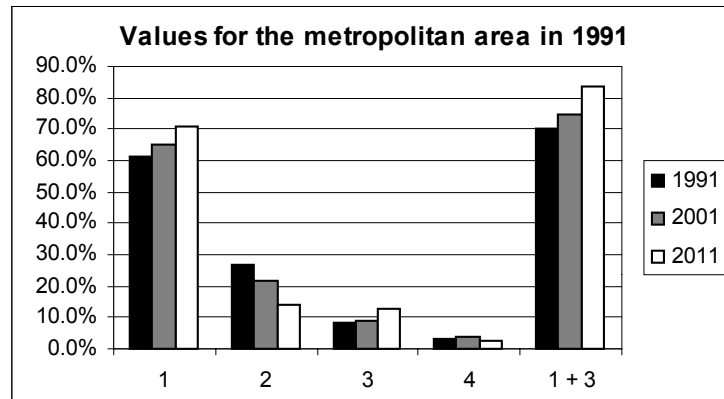

Figure 7 Distribution of *Freguesias* that were part of the original metropolitan region around Oporto in 1991 per quadrant
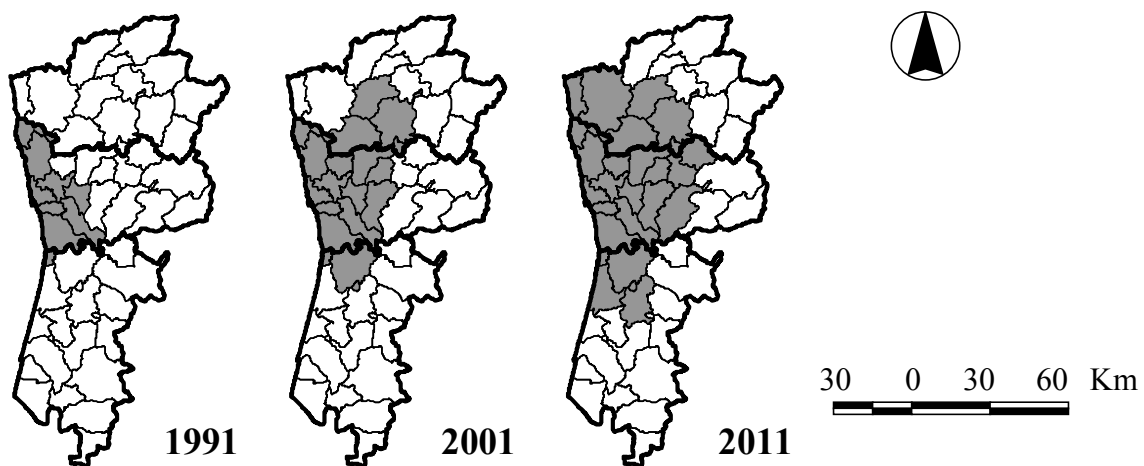

Figure 8 *Concelhos* that define the actual metropolitan region around Oporto (marked as 1991) and the proposed boundaries for the years 2001 and 2011
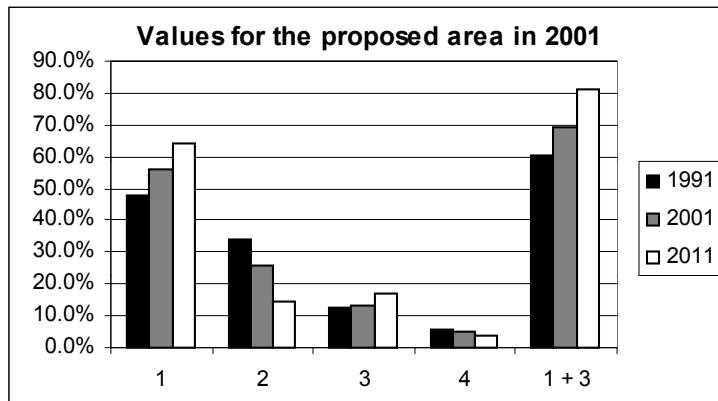
Figure 9 Distribution of *Freguesias* that were part of the proposed metropolitan region around Oporto in 2001 per quadrant
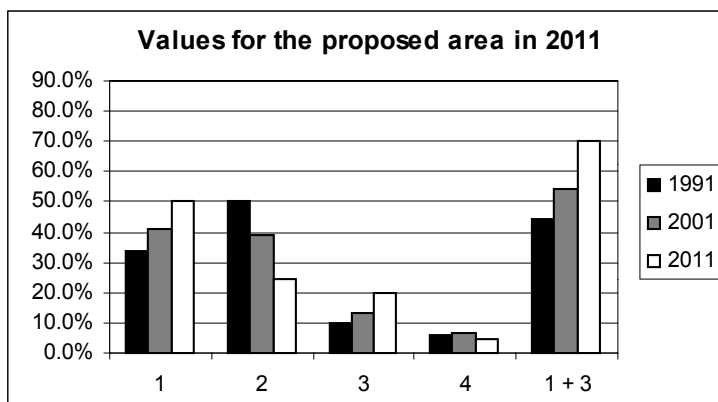

Figure 10 Distribution of *Freguesias* that were part of the proposed metropolitan region around Oporto in 2011 per quadrant

4. CONCLUDING REMARKS

In this paper we demonstrated that the combination of tools and techniques of Spatial Statistics and Spatial Modeling could be used for building models that can support the definition of metropolitan regions. The results found with the method proposed here show that the kind of structure used in the model construction, although not very complex, can be promising for prediction models. The spatial statistics approach applied brought to the model important spatial relationships of the variable under consideration. When using the number of neighbors per quadrant as one of the elements in the definition of the transition rules, for example, the model was able to take into account the spatial relationships beyond the immediate vicinity of each zone. This is certainly an advantage of the proposed approach and it may be, to a certain extent, what makes it superior to the traditional approaches (mainly those dealing with measures such as population density to defining metropolitan areas).

The application of the method in a case study for the redefinition of the Oporto metropolitan region in Portugal produced interesting results. Looking to the delimitation produced for the year 2011 and based on the census data for 2001, for example, the figures for the Oporto region visibly approach those of the actual Lisbon metropolitan region. While the latter has a population of 2,682,687 inhabitants, an area of 3,219 km^2 and a population density of 833.5 inhabitants/km^2, the 2011 proposed delimitation for Oporto would count also in 2001:

2,593,461 inhabitants, an area of 3,334 km^2 and a population density of 777.9 inhabitants/km^2. Together, the two areas would sum up 53.5 % of the country population, excluding the insular territories.

Despite the interesting results found with the prediction model, however, there is still room for improvement. A potential alternative for that is the replacement of the single transition rules set so far, which were based on aggregate values such as those shown in Table 2, by neural networks models that could work with all records of the dataset while looking for patterns of change. It would allow the prediction of population density values (and therefore Z) instead of the simultaneous prediction of Z and W$_Z$ carried out here. This option in already being investigated and the results shall be soon available for comparison. Another point that could be investigated is the replacement of the variable population density by any other variable that could play an important role in the definition of metropolitan regions. There is also the possibility of combining variables in the analyses.

ACKNOWLEDGEMENTS

REFERENCES

Anselin, L. (1995) Local indicators of spatial association - LISA, **Geographical Analysis, Vol. 27**, 93-115.

Anselin, L. (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In M. Fischer, H. Scholten and D. Unwin (eds.), **Spatial Analytical Perspectives on GIS**. Taylor & Francis, London.

Anselin, L. (1998a) Exploratory spatial data analysis in a geocomputational environment. In P. Longley, S. Brooks, B. Macmillan and R. McDonnell (eds.), **GeoComputation, a Primer.** Wiley, New York.

Anselin, L. (1998b) 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. and Bao, S. (1997) Exploratory spatial data analysis linking SpaceStat and ArcView. In M. Fisher and A. Getis (eds.), **Recent Developments in Spatial Analysis.** Springer-Verlag, Berlin.

Anselin, L. and Smirnov, O. (1998) **The SpaceStat extension for ArcView 3.0**. Regional Research Institute, West Virginia University, Morgantown.

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.

Ferrão, J. and Vala, F. (2001) Delimitação das aglomerações metropolitanas de Lisboa e Porto com base no critério de continuidade de espaço construído, **Revista de Estudos Regionais - Região Lisboa e Vale do Tejo, No. 2**, 7-35.

Ferrão, J., Rodrigues, D. and Vala, F. (2002) **As regiões metropolitanas portuguesas no contexto ibérico**. DGOTDU, Lisbon.

Ferreira, M.J. and Rosado, A.R. (1999) As Grandes Áreas Urbanas portuguesas. Conceitos e delimitação espacial. In M. Pereira (ed.), **Grandes Áreas Urbanas. Reorganização institucional e territorial. O caso da Área Metropolitana de Lisboa**. Centro de Estudos de Geografia e Planeamento, Universidade Nova de Lisboa, Portugal.

INE (1992) **Recenseamento da População e da Habitação (Portugal) – Censos 1991.** Instituto Nacional de Estatística, Lisboa.

INE (2002) **Recenseamento da População e da Habitação (Portugal) – Censos 2001.** Instituto Nacional de Estatística, Lisboa.

Lacour, C. and Puissant, S. (1999) **La métropolisation. Croissance, diversité, fractures.** Anthropos, Paris.

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.

Metropolitan Area Standards Review Committee (2000) Final Report and Recommendations From the Metropolitan Area Standards Review Committee to the Office of Management and Budget Concerning Changes to the Standards for Defining Metropolitan Areas, **Federal Register**, **Vol. 65**, **No. 163**, August 22.

NUREC (1994) **Atlas of agglomerations in the European Union**. Network on Urban Research in the European Union, Duisburg.

O'Sullivan, D. and Torrens, P.M. (2000) Cellular models of urban systems. In S. Bandini and T. Worsch (eds.), **Theoretical and Practical Issues on Cellular Automata**. Springer-Verlag, London.

Office of Management and Budget (1990) Revised Standards for Defining Metropolitan Areas in the 1990's; Notice, **Federal Register**, **Vol. 55**, **No. 62**, Friday, March 30.

Office of Management and Budget (1998) Alternative Approaches to Defining Metropolitan and Non-metropolitan Areas, **Federal Register**, **Vol. 63**, **No. 244**, December 21.

Office of Management and Budget (1999) Recommendations From the Metropolitan Area Standards Review Committee to the Office of Management and Budget Concerning Changes to the Standards for De-fining Metropolitan Areas, **Federal Register**, **Vol. 64**, **No. 202**, October 20.

Office of Management and Budget (2000) Standards for Defining Metropolitan and Micropolitan Statistical Areas, **Federal Register**, **Vol. 65**, **No. 249**, December 27.

Serrano, R.M. and Valcarce, E.V. (2000) **Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial**. Edicions Universita de Barcelona, Spain.

Torrens, P.M. (2000) **How cellular models of urban systems work - WP-28**. Centre for Advanced Spatial Analysis (CASA), University College London, England.

Torrens, P.M. and O'Sullivan, D. (2001) Cellular Automata and urban simulation: Where do we go from here? **Environment and Planning B: Planning and Design, Vol. 28**, 163-168.