
A spatial analysis approach for the definition of metropolitan regions—the case of Portugal

Rui António Rodrigues Ramos

Department of Civil Engineering, School of Engineering, University of Minho, Campus Gualtar, Braga 4710-057, Portugal; e-mail: rui.ramos@civil.uminho.pt

António Nélson Rodrigues da Silva

Department of Transportation, São Carlos School of Engineering, University of São Paulo, 400 Avenida Trabalhador São-Carlense, São Carlos 13566-590, Brazil, e-mail: anelson@sc.usp.br

Received 8 October 2004; in revised form 10 March 2006

Abstract. The objective of this paper is to present a combined, two-step spatial analysis 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 analysis: spatial statistics and spatial modelling. Spatial statistics tools are used to identify the characteristics of local association and are combined with a neural network in order to build prediction models. The analyses conducted with exploratory spatial data analysis tools and census data give a clear indication of clusters of zones with similar characteristics, which can be seen as uniform regions. Spatial models can then be used to foresee the global behaviour of regions in terms of growth, albeit the basis of 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.

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 which they encompass. In some cases, central governments take over the management of eventual conflicts produced by the excessive proximity of municipalities in this geographical context. In other cases, however, 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 owing 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 (such as 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 Rosada (1999), and Ferrão et al (2002). A number of authors have 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; Ferrão et al, 2002; Ramos and Silva, 2003; Ramos et al, 2004). 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.

Literature about the topic is also available in the United States (for example, Metropolitan Area Standards Review Committee, 2000; Office of Management and Budget, 1990; 1998; 1999; 2000), where the need for 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. These alternatives 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, because reliable and up-to-date commuting data are very hard to find in some countries. This was also defended by the authors of “Alternative approaches to defining metropolitan and non-metropolitan areas” (Office of Management and Budget, 1998, page 70549), 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 for most Portuguese urbanized areas, it is a good reason to consider the use of population density to define metropolitan areas in that context.

The objective of this paper is to present a two-step spatial analysis approach for the definition of metropolitan regions, an approach which is based on the strength of the spatial association of areas concerning specific attributes of the territory (for example, population distribution). Hence, this paper 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 analysis: spatial statistics and spatial modelling. In such a way, spatial statistical tools, which are used to identify the characteristics of local association, are combined with spatial modelling techniques in order to analyse 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 a means to classify the behaviour of each particular zone regarding the value of the attribute under consideration and the mean value of the same attribute for neighbouring zones. Four possibilities are considered in such a classification scheme, which correspond to quadrants in the scatterplot. The analyses both of the points in the scatterplots and of 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 to delimit the metropolitan region, as suggested here.

The proposed approach was tested in a case study carried out in Portugal, where this is a very opportune issue. Similarly to the procedure for remote-sensing image classification, our first modelling approach used fixed transition rules, which were based essentially on the personal judgment of the modellers. It tried to replicate general trends rather than individual changes of density in the administrative subdivisions studied, and the results obtained were not easy to validate, for several reasons. A second modelling structure, which is the focus of this paper, shows an evolution of the first approach, achieved by the use of neural networks (NN) in the spatial modelling phase. This is supposed to eliminate most of the problems observed in the original method. 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 owing to its continuous and strong expansion. The theoretical background for the proposed approach is discussed in section 2 of this paper, which is followed by the presentation of the application to Portugal in the period 1991–2011 in section 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.

2 The proposed approach

The study explores two branches of spatial analysis to examine the dynamics of metropolitan regions and to establish their boundaries: spatial statistics and spatial modelling. 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, and therefore indicate 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 (1998), in a discussion of methodological and technical issues associated with the implementation of 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; 1999) defines ESDA as a collection of techniques to describe and visualize spatial distributions; to identify atypical locations or spatial outliers; to discover patterns of spatial association, clusters, or hot spots; and to 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 by which locational similarity (observations in spatial proximity) is matched to 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 for the present study and it consists of a fixed collection of discrete spatial locations (points or polygons).

Anselin (1998) 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 and 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 a means to classify the behaviour of each particular zone regarding the value of the attribute under consideration and the mean value of the same attribute for neighbouring 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:

- (1) Spatial proximity matrix (\mathbf{W}): a matrix of dimension $n \times n$, in which every w_{ij} element receives a value of 1 if zones i and j are neighbours and 0 otherwise. The matrix is normalized by the division of each element equal to 1 of a line i by the total sum of the same line.
- (2) Vector of deviations (\mathbf{Z}): each element of the vector is obtained by the subtraction of the total mean (μ) from the attribute value of each zone ($z_i = y_i - \mu$).
- (3) Vector of weighed averages (\mathbf{W}_z): the product of \mathbf{W} and \mathbf{Z} . Each element of the vector is the average of the deviations of zone i neighbours.

The global spatial-autocorrelation index applied here was Moran's I . This coefficient, which in the case of standardized forms of the Moran scatterplot varies from -1 to $+1$ and has an expected value approaching 0 for a large sample size in the absence of autocorrelation, is calculated through equation (1).

$$I = \frac{\mathbf{Z}^T \mathbf{W}_z}{\mathbf{Z}^T \mathbf{Z}}, \quad (1)$$

where the superscript T denotes a transposed vector.

A careful examination of equation (1) suggests that Moran's coefficient can be interpreted as the linear regression coefficient, with W_z viewed as the dependent variable and Z as the independent variable. In such a way, the index I could be thought of as the slope of the regression line adjusted to the set of pairs.

In order to observe the $W_z Z$ relationships better, their values were plotted 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 neighbouring 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 into four quadrants, described here as Q1, Q2, Q3, and Q4.

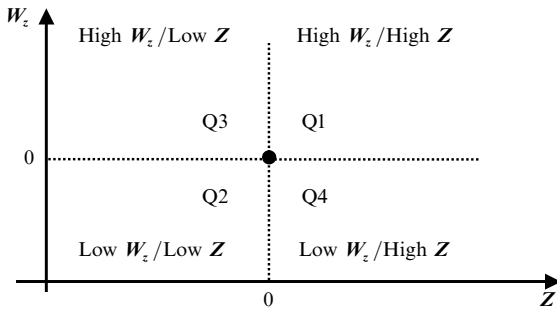


Figure 1. Interpretation of the Moran scatterplot. W_z is the dependent variable and Z is the independent variable.

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 neighbouring zones (positive value for the zone and positive average value for neighbours in Q1, and negative value for the zone and negative average value for neighbours 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 neighbouring zones (positive value for the zone and negative average value for the neighbours in Q4, and negative value for the zone and positive average value for neighbours 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, as the attribute values do not follow the pattern of the close neighbours.

Thematic maps based on the location of the point 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 it belongs to. In this way, one can visually identify the relationship between the attribute value measured for a particular zone and the average value obtained for its neighbours. The analyses both of the points in the scatterplots and of their associated location in a map give a clear indication of clusters of zones with similar characteristics, which can be seen as uniform regions.

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 neighbouring 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 model proposed here, which can be used to foresee the global behaviour of regions, although it is based on local (and historical) relationships among zones.

Under a rigorous examination of the literature concerning cellular automata (CA) theory (for example, O'Sullivan and Torrens, 2000; Torrens, 2000; Torrens and O'Sullivan, 2001), our prediction model does not fit into the class of CA models, although we have borrowed from them some of the 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 neighbours when building transition rules. This process is similar to that applied in CA models in two aspects: in the use of the transition rules (in the case of the present study, obtained with artificial NNs) and in the implicit assumption that local relationships can generate global patterns. However, 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 dynamic, because it has been used to simulate only one time step ahead of the latest period used in its construction although it could be used (at least in theory) to produce more simulations.

3 An application in Portugal

The application of the model proposed was divided into 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 and map described in section 2. The outcome of the analysis (that is, essentially the distribution of the areas in the quadrants and the quadrant changes in the two periods) provided the inputs for training the NN used to model the transition rules used in the second phase, which was the construction of a prediction model and its application for generating a future scenario. 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 Q1 and Q3 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 (over), in which these three spatial subdivisions are depicted for the region around the city of Oporto. Only the *concelhos* and *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 in figure 2, which is the actual, 'official' metropolitan region of Oporto. It is recognizable in figure 2(b) as a cluster of gray-shaded *concelhos*.

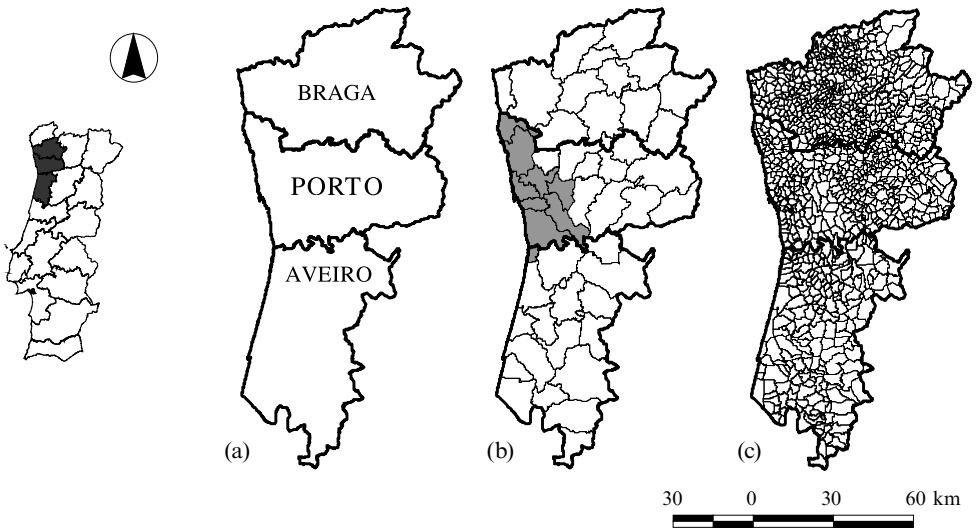


Figure 2. Spatial subdivisions in Portugal: (a) distritos, (b) concelhos, and (c) freguesias. The shaded concelhos mark the 'official' metropolitan region of Oporto.

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 by the Portuguese census bureau (INE, 1992; 2002), from which only data about population were 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 match between the boundaries was based on an official report of the changes produced from 1991 to 2001. The changes result from the subdivision of some 1991 freguesias. The 1991 data were then distributed by the new territorial subdivision, with the proportionality verified in 2001 maintained. The analyses were carried out with the tools available in the geographic information system software package ArcView (ESRI Inc., Redlands, CA) along with the Spacestat extension (Anselin and Bao, 1997; Anselin and Smirnov, 1998)—to compute 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. It is important to emphasize that two alternatives were tested in the generation of the weight matrix: rook and queen neighbourhoods. As the alternatives did not produce significantly different results, owing to the irregular configuration of the boundaries, the neighbourhood matrix considered was based on queen values.

The general Moran's I indices obtained for the population-density variable in Portugal in the years 1991 and 2001 with equation (1) were 0.7359 and 0.7271, respectively. 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 neighbouring zones. In those quadrants (particularly in Q2, in which more than 3000 points are clustered very close to the crossing lines in the middle of both graphs) were located 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 large number of zones with values in Q2 is a consequence of the population migration trend

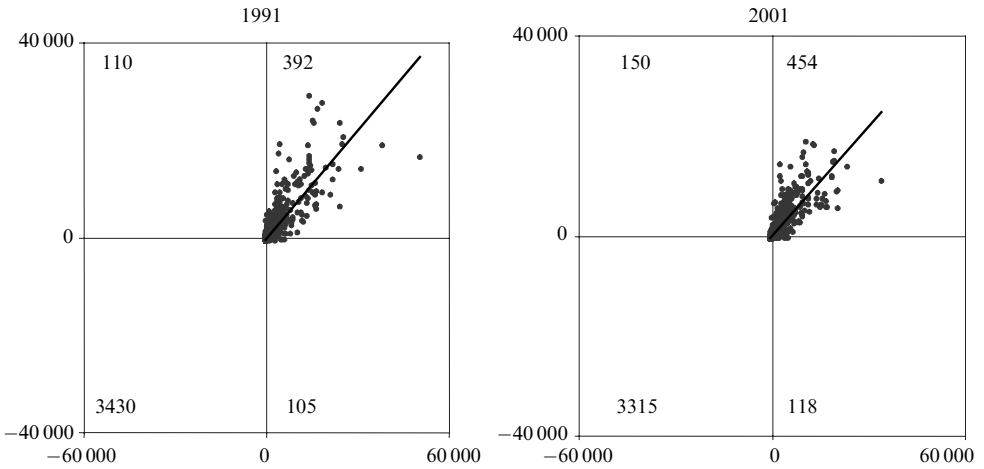


Figure 3. Moran's scatterplots for the population-density variable in Portugal in the years 1991 and 2001. Refer to figure 1 for an explanation of the quadrants.

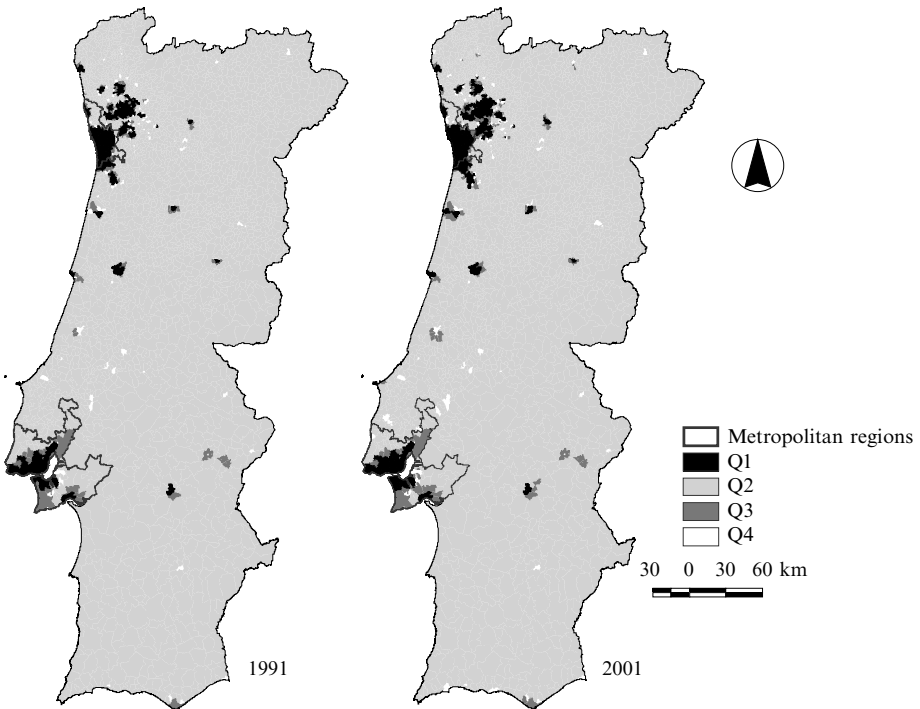


Figure 4. Thematic maps showing the distribution of points in the Moran scatterplot on the basis of the population-density variable in Portugal for the years 1991 and 2001. Q1, Q2, Q3, and Q4 correspond to the quadrants shown in figure 1.

from hinterland to major urban areas. The location of all regions to which the points in the scatterplot are referring can be seen in the thematic map of figure 4.

In the maps shown in figure 4 one can observe that most regions associated with points in Q1 were geographically located in two areas close to the west shore. These are exactly the regions in which 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 neighbouring zones—that is, the attribute value was similar to the neighbourhood average value. For zones in Q1 the deviation value per zone and the average of deviations were both positive, which 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 Q1 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 they are 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 neighbouring zones. In this 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 Q3 and Q4. 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. Owing to the propensity to change from Q3 to Q1, zones in Q3 were particularly important for the prediction model, as shown in the next step.

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 with the origin quadrant and the destination quadrant. In the case of the eleven zones that changed from Q2 to Q1 (table 1), for example, the associated code is 21. The number of neighbours 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 neighbours, were used to build the fixed transition rules used in our first prediction model.

A careful examination of the conditions identified for all transition classes resulted in a set of transition rules for zones with an origin in all four quadrants, as usually performed in the analysis and classification of images (for example, in remote sensing). The essence of such an approach is to classify the zones according to their relative positions in relation to one or more thresholds by means of 'if ... then ... else' rules. In this case, the combination of the distinct sets of transitions rules applied to all quadrants resulted in a model able to match 95% of the quadrants found in 2001. This apparently good result was, however, only based on the overall number of elements in the quadrants. Therefore, it could be used only to predict future quadrants, which can accommodate a reasonably large range of population densities. In addition,

Table 1. Transition matrix in the period 1991–2001. Q1, Q2, Q3, and Q4 correspond to the quadrants shown in figure 1.

	2001			
	Q1	Q2	Q3	Q4
1991				
Q1	392	–	–	–
Q2	11	3314	70	35
Q3	30	–	80	–
Q4	21	1	–	83

the model was not suitable for validation. These were the reasons why we decided to change the approach for modelling the transitions, as presented in the sequence. One of the main advantages of using NNs is the possibility of having population-density values as the output of the models instead of quadrants. In such a way, the quadrants can be subsequently calculated on the basis of the predicted density values, a process which certainly provides a more robust model. Another advantage of the NN method is the possibility of having parts of the dataset for validation and testing. In order to do so, we randomly selected 25% of the data for each of these steps, in addition to 50% used for training the network. The kind of NN used was a multilayer perceptron with a standard backpropagation algorithm. The superior performance of the NN model for predicting the population densities can be confirmed by the results presented in figure 5, in which graphs with validation and test data are displayed side by side. A visual analysis of the charts in which predicted values were plotted against actual values showed that there were no points extremely far from the 45° reference line (figure 5).

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 2 and the resulting scenario is presented as a map in figure 6. Again, the boundaries of the actual metropolitan regions are left in the map for comparison with the distribution of zones in the quadrants. The results of the first modelling approach are kept in table 2 side by side with the summary of the quadrants obtained through the NN prediction, in order to make easy a direct comparison of the aggregate results cell by cell.

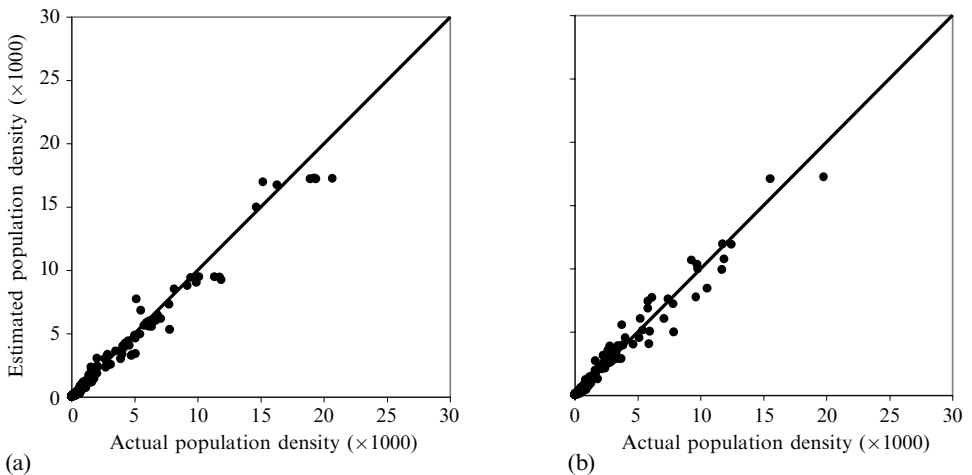


Figure 5. Actual versus estimated population density in 2001 for (a) validation data and (b) test data.

Table 2. Transition matrix in the period 2001 – 11. Values in parentheses are results from the first modelling approach. Q1, Q2, Q3, and Q4 correspond to the quadrants shown in figure 1.

	2011			
	Q1	Q2	Q3	Q4
2001				
Q1	439 (454)	–	14 (–)	1 (–)
Q2	–	3 271 (3 128)	34 (142)	10 (45)
Q3	21 (28)	8 (–)	121 (122)	–
Q4	16 (64)	26 (–)	–	76 (54)

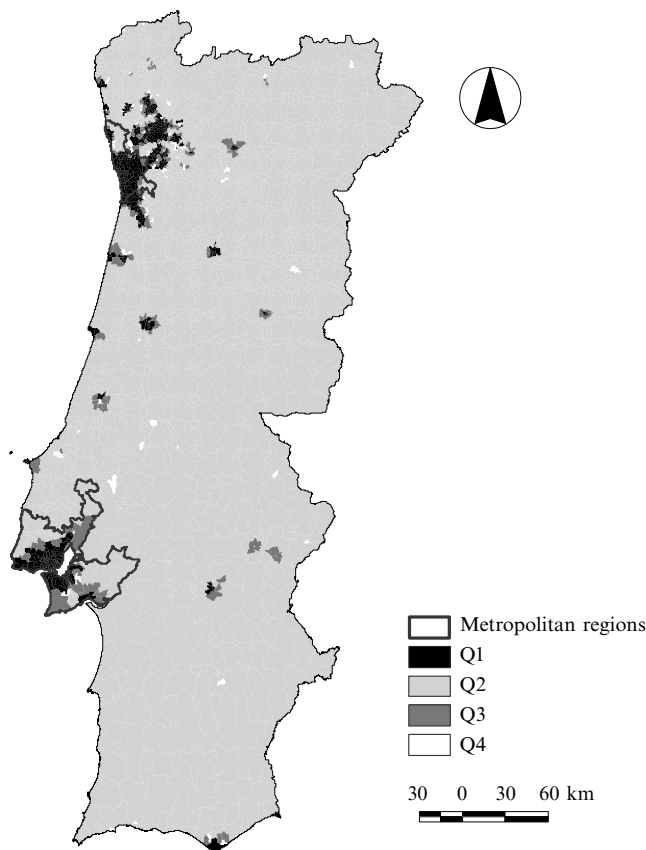


Figure 6. Thematic map showing the distribution of points in the Moran scatterplot on the basis of population-density variable in Portugal for the year 2011. Q, Q2, Q3, and Q4 correspond to the quadrants shown in figure 1.

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, after 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 7. The picture is even more evident when looking at the projected scenario in 2011, in which 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 into the metropolitan region when adjusting its boundaries to reflect the changes through time.

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 (70%) on the right-hand side of figure 8. The other two bars correspond to the proportions of elements in each

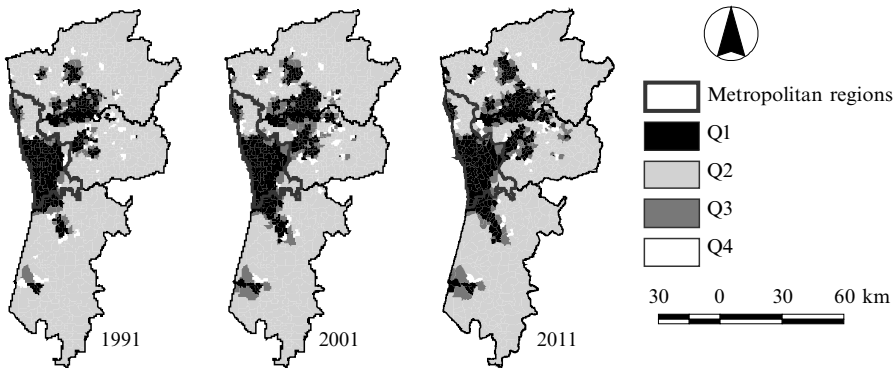


Figure 7. Oporto metropolitan region—for the concelhos of Aveiro, Braga, and Oporto, showing the distribution of points in the Moran scatterplot on the basis of the population-density variable in Portugal for the years 1991, 2001, and 2011. Q1, Q2, Q3, and Q4 correspond to the quadrants shown in figure 1.

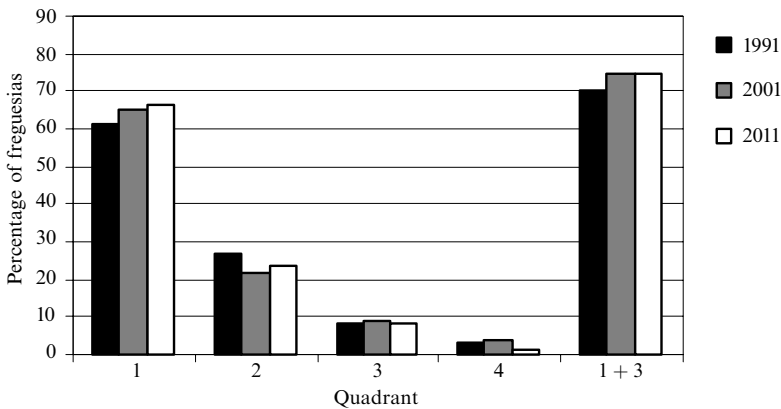


Figure 8. Distribution of freguesias that were part of the original metropolitan region around Oporto in 1991 per quadrant, along with comparisons for 2001 and 2011.

quadrant, given the 1991 boundary, in the years of 2001 and 2011, put in the graph just for comparison. Different groups of concelhos could be added to the metropolitan regions in order to reach the percentage target, preferably the ones with Q1 areas but also with Q3 elements. The aggregation of concelhos that led to values of Q1 and Q3 close to the 70% found in 1991 is represented in figure 9; also shown in figure 9 are the aggregations for the years 2001 and 2011, corresponding to the values shown in figures 10 (gray ‘1 + 3’ bar) and 11 (white ‘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. Rather, in the aggregation of 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 patterns 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.

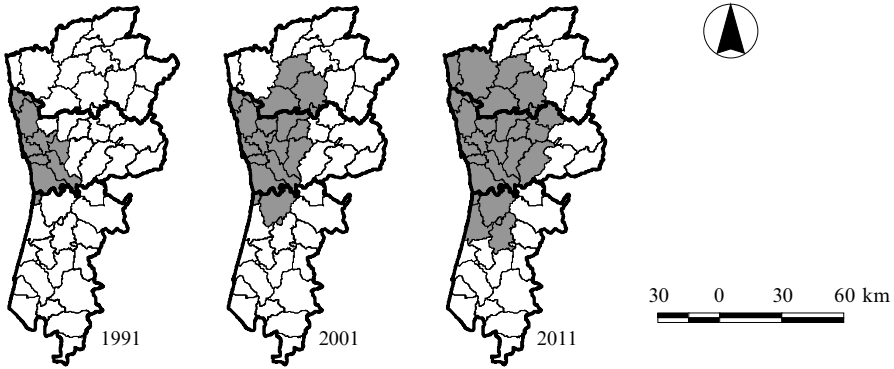


Figure 9. Concelhos that define the actual metropolitan region around Oporto (marked as 1991) and the proposed boundaries for the years 2001 and 2011.

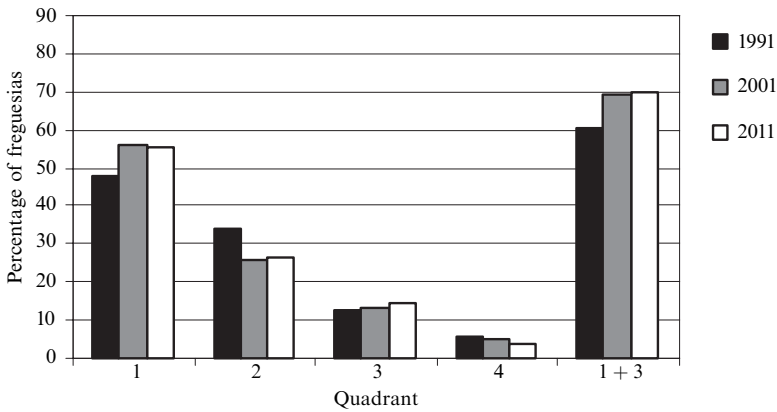


Figure 10. Distribution of freguesias that were part of the proposed metropolitan region around Oporto in 2001 per quadrant, along with comparisons for 1991 and 2011.

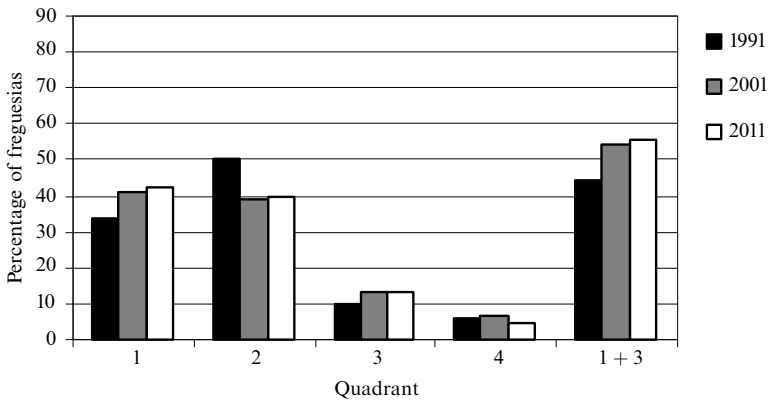


Figure 11. Distribution of freguesias that were part of the proposed metropolitan region around Oporto in 2011 per quadrant, along with comparisons for 1991 and 2001.

The effects of the conditions applied can be seen by the variance among the zones in terms of the global values considered. The cutoff procedure produced variance values inside the metropolitan area of 10 284 118 in 1991 and 3 147 591 in 2001. A direct comparison of the 1991 and 2001 variance values is possible, but care should be taken when looking to the 2011 variance values, given the implicit assumption of densification associated with the NN output. That is the reason why we did not present the 2011 variance values here. Instead, we tried to extend the comparison of 1991 and 2001 values in different ways. Table 3 helps to understand how each concelho individually performs in terms of the cutoff and how the data are distributed in 1991 and 2001. Although some of the concelhos also had freguesias classified in Q2 and Q4, the percentages shown only consider Q1 and Q3 to sum to 100%.

Table 3. Percentage of Q1 and Q3 zones in the concelhos which constitute the 1991 and 2001 metropolitan areas. Values in bold correspond to concelhos in the metropolitan area for the given year.

Concelhos	1991		2001	
	Q1	Q3	Q1	Q3
Braga	73.5	26.5	80.6	19.4
Espinho	100.0	0.0	100.0	0.0
Gondomar	87.5	12.5	87.5	12.5
Guimarães	77.1	22.9	87.8	12.2
Maia	75.0	25.0	82.4	17.6
Matosinhos	100.0	0.0	100.0	0.0
Paredes	50.0	50.0	68.4	31.6
Paços de Ferreira	100.0	0.0	78.6	21.4
Porto	100.0	0.0	100.0	0.0
Póvoa de Varzim	71.4	28.6	85.7	14.3
Santa Maria de Feira	73.7	26.3	73.9	26.1
Santo Tirso	75.0	25.0	76.5	23.5
São João da Madeira	100.0	0.0	100.0	0.0
Trofa	100.0	0.0	50.0	50.0
Valonga	100.0	0.0	80.0	20.0
Vila do Conde	40.0	60.0	37.5	62.5
Vila Nova de Famalicão	71.0	29.0	65.7	34.3
Vila Nova de Gaia	95.5	4.5	95.5	4.5
Vizela	80.0	20.0	100.0	0.0

4 Concluding remarks

In this paper we have demonstrated that the combination of tools and techniques of spatial statistics and spatial modelling could be used to build 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 could be promising for prediction models. The spatial statistics approach which was applied brought to the model important spatial relationships of the variable under consideration. In the use of the number of neighbours 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 define 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. Though the latter has a population of 2 682 687 inhabitants, an area of 3219 km², and a population density of 833.5 inhabitants/km², the 2011 proposed delimitation for Oporto would count also in 2001: 2 593 461 inhabitants, an area of 3334 km², and a population density of 777.9 inhabitants/km². Together, the two areas would constitute 53.5% of the national population, excluding the insular territories.

The results found with the prediction model were certainly improved by the replacement of the single transition rules originally set, which were based on aggregate values, by a NN model that worked with all records of the dataset and looked for patterns of change. One of the main advantages here is the fact that the NN model allowed the prediction of population density values (and therefore Z) instead of the simultaneous prediction of Z and W_z carried out before. Another point that could be further explored now is the replacement of the population-density variable 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. 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. Finally, we thank the referees for their valuable comments.

References

- Anselin L, 1995, "Local indicators of spatial association—LISA" *Geographical Analysis* **27** 93–115
- Anselin L, 1996, "The Moran scatterplot as an ESDA tool to assess local instability in spatial association", in *Spatial Analytical Perspectives on GIS* Eds M Fischer, H Scholten, D Unwin (Taylor and Francis, London) pp 111–125
- Anselin L, 1998, "Exploratory spatial data analysis in a geocomputational environment", in *GeoComputation, a Primer* Eds P Longley, S Brooks, B Macmillan, R McDonnell (John Wiley, New York) pp 77–94
- Anselin L, 1999, "Interactive techniques and exploratory spatial data analysis", in *Geographical Information Systems: Principles, Techniques, Management and Applications—Volume 1* Eds P Longley, M Goodchild, D Maquire, D Rhind (John Wiley, New York) pp 251–264
- Anselin L, Bao S, 1997, "Exploratory spatial data analysis linking SpaceStat and ArcView", in *Recent Developments in Spatial Analysis—Spatial Statistics, Behavioural Modelling and Computational Intelligence* Eds M Fisher, A Getis (Springer, Berlin) pp 35–59
- Anselin L, Smirnov O, 1998 *The SpaceStat Extension for ArcView 3.0* Regional Research Institute, West Virginia University, Morgantown, WV
- Cressie N, 1993 *Statistics for Spatial Data* (John Wiley, New York)
- Ferrão J, Vala F, 2001, "Delimitação das aglomerações metropolitanas de Lisboa e Porto com base no critério de continuidade de espaço construído" [Delimitations of Lisbon and Oporto metropolitan regions based on the continuity of built space] *Revista de Estudos Regionais—Região Lisboa e Vale do Tejo* **2** 7–35
- Ferrão J, Rodrigues D, Vala F, 2002 *As regiões metropolitanas portuguesas no contexto ibérico* [The Portuguese metropolitan regions in the Iberian context] Direcção-Geral do Ordenamento do Território e Desenvolvimento Urbano, Campo Grande 50, Lisbon
- Ferreira M J, Rosada A R, 1999, "As grandes áreas urbanas portuguesas. Conceitos e delimitação espacial" [The large Portuguese urban areas. Concepts and space delimitation], in *Grandes Áreas Urbanas. Reorganização institucional e territorial. O caso da área Metropolitana de Lisboa* Ed. M Pereira (Centro de Estudos de Geografia e Planeamento, Universidade Nova de Lisboa, Lisbon) pp 38–50
- INE, 1992 *Recenseamento da População e da Habitação (Portugal)—Census 1991* [Portugal 1991 census of population and households] (Instituto Nacional de Estatística, Lisbon)
- INE, 2002 *Recenseamento da População e da Habitação (Portugal)—Census 2001* [Portugal 2001 census of population and households] (Instituto Nacional de Estatística, Lisbon)

-
- Lacour C, Puissant S, 1999 *La Métropolisation. Croissance, Diversité, Fractures* [Metropolization. Growth, diversity, fractures] (Anthropos, Paris)
- Levine N, 1996, "Spatial statistics and GIS: software tools to quantify spatial patterns" *Journal of the American Planning Association* **62** 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* **65** 51060 – 51077
- NUREC, 1994 *Atlas of Agglomerations in the European Union Network on Urban Research in the European Union*, Duisburg, <http://www.uni-duisburg.de/duisburg/fatlas.htm>
- Office of Management and Budget, 1990, "Revised standards for defining metropolitan areas in the 1990: notice" *Federal Register* **55** 12154
- Office of Management and Budget, 1998, "Alternative approaches to defining metropolitan and non-metropolitan areas" *Federal Register* **63** 70526 – 70561
- 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 defining metropolitan areas" *Federal Register* **64** 56628 – 56644
- Office of Management and Budget, 2000, "Standards for defining metropolitan and micropolitan statistical areas" *Federal Register* **65** 82228 – 82238
- O'Sullivan D, Torrens P M, 2000, "Cellular models of urban systems", in *Theoretical and Practical Issues on Cellular Automata* Eds S Bandini, T Worsch (Springer, Berlin) pp 108 – 117
- Ramos R A R, Silva A N R, 2003, "Um contributo para a delimitação da área metropolitana do noroeste de Portugal" [A contribution for the delimitation of the metropolitan area in the northwest of Portugal] *Revista Portuguesa de Estudos Regionais* **3** 61 – 82
- Ramos R A R, Silva A N R, Miranda V, 2004, "A comparison of two methods for the definition of regional metropolitan areas through an application in the north of Portugal", in *Proceedings of the 44th Congress of the European Regional Science Association* Oporto University, Oporto (CD-Rom)
- Serrano R M, Valcarce E V, 2000 *Técnicas Económicas para el Tratamiento de Datos Espaciales: La Econometría Espacial* [Econometric techniques for the treatment of spatial data: the spatial econometrics] (Edicions Universita de Barcelona, Barcelona)
- Torrens P M, 2000, "How cellular models of urban systems work", WP 28, Centre for Advanced Spatial Analysis, University College London, London
- Torrens P M, O'Sullivan D, 2001, "Cellular automata and urban simulation: where do we go from here?" *Environment and Planning B: Planning and Design* **28** 163 – 168