Volume 24
Number 3
September – December 2012

ITALIAN JOURNAL OF APPLIED STATISTICS

Statistica Applicata

T. Di Battista,
S. T. Tamás
Co-Editors

Special Issue

Spatial statistics for business and services

ISSN: 1125-1964

RCE MULTIMEDIA

Statistica Applicata • Italian Journal of Applied Statistics   Special Issue   Volume 24 • Number 3 • 2012

Statistica Applicata - ITALIAN JOURNAL OF APPLIED STATISTICS

Vol. 24 n. 3 September – December 2012

Statistica Applicata – Italian Journal of Applied Statistics is a four-monthly journal published by the Associazione per la Statistica Applicata (A.S.A.), Largo Gemelli 1 – 20123 Milano, Italy (phone +39 02 72 34 29 04). Advertising: RCE Multimedia, Piazza Bagnoli 19 – 80124 Napoli, Italy (phone and fax +39 081 230 34 16); email: rcemultimedia@yahoo.it.

Rules for manuscript submission: http://sa-ijas.stat.unipd.it.

Subscription: yearly € 103.30; single copy € 40.00; A.S.A. associates € 60.00; supporting institutions: € 350.00.

Advertisement lower than 70%. Postal subscription Group IV, Milan. Forum licence n. 782/89, 31 October, 1989.

Editor-in-chief: Luigi Fabbris

Statistica Applicata – Italian Journal of Applied Statistics is associated to the following Italian and international journals:

QTQM – Quality Technology & Quantitative Management (http/web.it.nctu.ew/~qtqm/)

MICRO & MACRO MARKETING (www.mulino.it/edizioni riviste/issn/1121-4228)

SINERGIE – Rivista di Studi e Ricerche (www.sinergiejournal.it)

Published: July 2015

RCE Multimedia

Piazza Bagnoli, 19 – 80124 Napoli Phone - Fax +39 081 230 34 16

email: rcemultimedia@yahoo.it

www.rcemultimedia.com

Summary

Statistica Applicata
ITALIAN JOURNAL OF APPLIED STATISTICS
Volume 24 issue 3

DEMOGRAPHIC STRUCTURE AND SPATIALLY CONTIGUOUS AREAS IN TUSCANY

Federico Benassi[1]

Italian National Institute of Statistics, Rome, Italy

Alessia Naccarato

Department of Economics, University of Rome "Roma Tre", Rome, Italy

Received: 30th July 2013/ Accepted: 15th January 2014

***Abstract**. It is argued that the organization of services by local authorities should take into account both the demographic features of the usually resident population and the spatial dimension of the territory involved. Focusing on the Italian region of Tuscany, the paper seeks to identify homogeneous and spatially contiguous areas between the municipal and regional levels. Demographic indexes of usually resident population and spatial attributes of each Tuscan municipality are used as the input of a spatial clustering and regionalization model in order to obtain* n *areas that minimize the internal heterogeneity of the demographic structure under the constraint of spatial contiguity.*

***Keywords**: Spatial clustering, Regionalization, Spatial demography, Regional planning.*

1. INTRODUCTION

The study of spatial distribution constitutes an element of primary importance whenever the phenomenon of interest can be georeferenced. This holds all the more when the objective is the analysis of themes closely connected with populations of individuals that live in a specific area on a permanent basis and manifest their social habits, behaviours and interactions between subgroups (Voss, 2007). In this perspective, the planning of the resources required by a population in terms of public services cannot, in our view, dispense with careful study of how it is distributed over the territory with reference to the demographic characteristics that define it (De Castro, 2007).

The integration of spatial information into the context of social and especially demographic studies is a process broadly developed in the literature of the last few

[1] Federico Benassi, email: benassi@istat.it

decades (Goodchild and Janelle, 2010), not least due to the availability of computational processes capable of handling vast amounts of information and dedicated software for the mapping of spatial distribution.

Studying how a population is distributed over a territory can make it possible to define models for the allocation of resources in areas such as education (Elis-Williams, 1987) and health services (Lwasa, 2007) or for the planning and organization of economic resources in general (Linard et al., 2012) in order to take into account how the demographic characteristics of a population – which largely define the needs of individuals in terms of services – manifest themselves in the territory in question.

In Italy most of the public services are organized at the provincial level (a province is an area equipped with administrative structures within a region) and even the services not provided directly by provinces appear to require an intermediate level of management between the municipal and the regional. Given this situation and the plans put forward in the last few years for the reorganization of local government in Italy (regions, provinces and municipalities), the aim of this work is to ascertain – with reference to the region of Tuscany – whether the territorial distribution of population resulting from indicators of demographic structure corresponds to the present division of the region into provinces. If the population inside the province proves to be distributed homogeneously with respect to the primary demographic characteristics, the planning and provision of the services can continue to take place efficiently at the provincial level; if not, it will be necessary to adopt an area of reference other than the province.

This study is based on the idea that the spatial distribution of a phenomenon does not depend exclusively on topographical elements such as spatial proximity and/or Euclidean distance between individual locations, but also on other aspects lying outside the dimension strictly connected with spatial metrics (Anselin, 1995, 2003; Haining et al., 2010; Naccarato, 2012). It follows from this that the organization of services within a territory at the level of sub-regional areas must take into consideration not only the spatial proximity of the areas or the morphological characteristics of the territory, but also the demographic characteristics of the population of the areas in question. If a local authority is called upon to organize the provision of a particular service for a population distributed over various municipalities, it is advantageous both that they are close to one another, in order to comply with the constraints of costs and efficiency, and that they present no radical differences in terms, for example, of demographic structure. In the case of providing nursery schools, for example, the population interested in this service will presumably consist of young families, who are more likely to have children of pre-school age. The task is thus to provide the structures required in such a way that all of the younger population can make use of them. Given the need to comply with economic constraints, the optimal situation is for this population to live in neighbouring municipalities that can be served by a smaller number of nursery schools. If the young families are instead resident in municipalities distant from one another, it will be necessary either to provide more nursery schools to cater for their needs or to provide an efficient system of public transport.

The study is divided into two phases. In the first one, the municipalities are divided into homogeneous clusters as regards selected demographic indicators with no constraints of territorial contiguity. The results obtained show a division of Tuscan municipalities into clusters that are internally homogeneous but differ from one another, which suggests a certain differentiation of demographic structure. Mapping of the individual units (municipalities) belonging to the various clusters shows that, in a certain number of cases, these differences follow clearly defined spatial patterns, in the sense that it is not infrequently possible to identify groups of municipalities that are similar in demographic profile and geographically close to one another. This bears out the first law of geography, namely that everything is connected with everything else but what is spatially close is more connected (Tobler, 1970). This regularity does not always occur, however, especially when constraints of territorial contiguity are imposed. This aspect emerges in the second phase, where the results obtained in the first phase and the imposition of a constraint of territorial contiguity are used for the purposes of regionalization in order to obtain a number *n* of contiguous areas at an intermediate level between the municipal and the regional. The results are not in line with the present division of the region of Tuscany into provinces and unsatisfactory as regards the primary objective of maintaining a good degree of internal homogeneity under the constraint of contiguity. The areas obtained therefore do not differ from one another in terms of demographic structure and the characterization found in the first phase of the work – where the spatial element, though present, was not imposed as a constraint – thus disappears.

This seems to indicate that the territorial patterns of demographic structures undergo a scale effect when they refer to provincial aggregates or in any case to areas intermediate between the local (municipal) level and the regional. The province would therefore seem to lose significance in terms closely connected with the demographic structures of the resident population. The geographical scale of reference appears to have an effect of a certain importance in this connection, as the same result is not obtained in the case of areas of a supra-municipal nature but less extended over the territory, such as metropolitan areas (White and Engelen, 2000; Mu and Wang, 2008; Benassi et al., 2013).

It is of course impossible to force the population to reside on the territory in such a way as to follow a homogeneous demographic structure that also complies with the constraint of contiguity. The solution is therefore to be sought at the administrative and managerial level, not confining the organization of local services within the present provincial boundaries, but rather identifying macro-aggregates of neighbouring municipalities that are homogeneous as regards the primary demographic characteristics of the resident population.

It should be noted that reference is made in this context to primary services meeting the needs of the community and generally provided by government bodies such as education, health care, road infrastructures and some social and welfare services.

The paper is organized as follows. Section 2 presents the method of spatial clustering and regionalization adopted. Section 3 describes the data used, the demographic indexes and the results obtained by applying the method outlined in Section 2. The concluding Section 4 presents some observations on the results and possible future developments.

2. SPATIAL DATA MINING FOR CLUSTERING AND REGIONALIZATION

While spatial data mining was regarded as the multidimensional equivalent of temporal data mining for several years (Roddick and Spiliopoulou, 1999), it is now largely seen by scholars as an independent approach to the analysis of data and the measurement of phenomena as confirmed by recent studies (Angayarkkani and Radhakrishnan, 2009; Behnisch and Ultsch, 2009, 2010; Jin and Guo, 2009; Moran and Bui, 2002; Yu Pan and Faloutsos, 2002).

It can be said in general terms that the difference between data mining and spatial data mining is the same as between classical statistical analysis and spatial statistical analysis. As is known, one of the most important assumptions of classical statistical analysis is that the sample data are generated independently. The spatial approach drastically eliminates this assumption on the grounds that the spatial location of the samples (data) is an element that cannot be ignored (Tobler, 1970).

In much the same way, data mining is closely connected with the concept of patterns and spatial data mining instead with spatial patterns. These theoretical differences obviously have important repercussions in operative terms. Adopting a spatial approach means in fact that the scale of large databases becomes still greater because the data in this case also comprise geographical and topological information. This necessitates access to more powerful techniques of data processing and knowledge extraction, new methodologies for the construction of maps capable of presenting the results obtained in readily comprehensible terms also to non-specialists (policy makers, local government officials, etc.), and finally flexible software that encourages the interaction of users with data (Koperski et al., 1996).

Spatial clustering and regionalization are directly connected with spatial data mining. The former is a process for grouping a set of spatial objects into meaningful subclasses (clusters) so that the members of a cluster are as similar to one another as possible, whereas members of different clusters differ as much as possible (Jiao and Liu, 2008). The role of spatial data mining is to scale a spatial clustering algorithm to deal with large geographical datasets. Spatial clustering algorithms and approaches can be separated into four general categories: partitioning method, hierarchical method, density-based method and grid-based method (Han et al., 2001, 2009). As defined by Guo (2008), regionalization is a process that divides a large set of spatial objects into a number of spatially contiguous regions while optimizing an objective function, normally a homogeneity (or heterogeneity) measure of the identified regions. Regionalization is therefore a special kind of spatial clustering where the condition of spatial contiguity between spatial objects plays a key role. As recalled by Bernetti *et al.* (2011), regionalization processes play an important part in many sectors of research, finding application in areas like climatic zoning (Fovell and Fovell, 1993; Jin and Guo, 2009; Wang et al., 2010), environmental analysis (Henderson, 2006; Romano et al., 2010), the analysis of units of landscape (Long et al., 2010), the interpretation and organization of census data (Openshaw and Rao, 1995) and public health data (Haining et al., 1994; Osnes, 1999), the analysis of socio-economic phenomena (Assuncão et al., 2006; Benassi and Ferrara, 2010; Benassi et al., 2010), and the analysis and interpretation of demographic and urban/regional dynamics (Behnisch and Ultsch, 2010; Benassi et al., 2013). In particular, the concept of regionalization hypothesized and applied to socio-economic entities by Openshaw (1977) results in the creation of geographic objects formed by combining contiguous elements sharing one or more characteristics and is closely connected with spatial statistics (Bernetti et al., 2011).

Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (RedCap), a method of spatial data mining recently put forward by Guo (2008) for operations of spatial clustering and regionalization, is employed in this work.

RedCap is based essentially on a set of six methods of regionalization given by the combination of three methods of agglomerative clustering – Single Linkage (SLK), Average Linkage (AVG), and Complete Linkage (CLK) – and two different strategies of spatial constriction, namely First-Order and Full-Order. Guo (2008) shows in his work that out of these six methods, the combination of Complete Linkage clustering with the Full Order constraining strategy (CLK-Full order) gives the best overall performance. Readers are referred to Guo et al. (2005) and

Guo (2008) for the technical and computational details.

RedCap consists of two basic steps. In the first, which is based on the interactive algorithms of the Self-Organizing Map (SOM) developed by Guo et al. (2005), the method identifies spatial clusters without imposing any spatial constraints, i.e. without imposing the constraint of territorial contiguity. The results of this first step are visually presented by the Unified Distance Matrix of the SOM, by the Parallel Coordinate Plot (PCP), and by an interactive map of the area under examination. In the second step, based on a matrix of spatial contiguity and a set of strategies of spatial constraints, the process of regionalization is completed. As stated, the results of these two steps are linked to and visually presented in interactive maps.

3. DATA AND RESULTS

3.1 INPUT VARIABLES

The population of reference is the one of 3,667,780 units resident on January 1, 2012 in the 287 municipalities of the region of Tuscany divided into the 10 provinces of Arezzo, Florence, Grosseto, Livorno, Lucca, Massa Carrara, Pisa, Pistoia, Prato, and Siena. In particular, we used data on usually resident population broken down by age and sex at municipalities level issued by the Istituto Nazionale di Statistica (Istat). The indexes of demographic structure then calculated for each municipality on the basis of this information were also employed in a recent study focused on a more specific territorial context, but still in the region of Tuscany, namely the Florentine Metropolitan Area (Benassi et al., 2013). The indexes calculated, which are described below, and the spatial attributes of each municipality constitute the input variables of RedCap (Guo, 2008).

The five indexes of demographic structure are: (a) the aging index, (b) the youth dependency index, (c) the old-age dependency index, (d) the structure of the population of active age index, and (e) the replacement of the population of active age index. The aging index (a) is an indicator of the aging of the population and consists of the ratio of the population aged 65 and over to the 0–14 age group. This is a dynamic index in which the aging of a population means both a decrease in the weight of the very young and an increase in the weight of the old; in other words, both terms of the ratio change in opposite directions (Livi Bacci, 1981). The youth dependency index (b) and the old-age dependency index (c) are dealt with separately here, even though they are part of the total dependency index, that is, a more general demographic index of a certain economic and social importance. These are the ratios of people presumed to be dependent – the 0–14 age group in the first case and the 65 and over age group in the second – to the people of active age (15–64), as these are theoretically responsible for supporting the other two sections of the population (the young and the old) through their work and economic activities (Livi Bacci, 1981). The structure of the population of active age index (d) is a measure of the degree of aging of the population aged between 15 and 64, and consists of the ratio of the 25 older generations (ages 40–64) to the 25 younger generations (15–39) contained within this segment of the population and destined to take their place. Finally, the replacement of the population of active age index (e) is the ratio of those about to exit the population of active age (the 60–64 age group) to those about to enter it (the 15–19 age group) (Livi Bacci, 1981).

While fairly elementary and also comparatively crude, these indexes are capable of revealing imbalances in terms of the demographic structure characterizing the population of a specific area. As is known, these structural imbalances are substantially attributable to the dynamics of fertility and mortality of the population over time as well as internal and international migration. All these events (births, deaths, emigration and immigration) have a substantial – albeit differentiated, especially in our day – impact on the age structure of a population.

3.2 SPATIAL CLUSTERING WITH NO CONSTRAINING STRATEGY

On the basis of the input variables, after standardization of the demographic indexes, RedCap identified 16 clusters of municipalities in the first phase. This numerosity is due to the dimension of the SOM, namely 4 x 4, which was chosen out of the various alternatives because it proved to ensure the best result in terms of cluster differentiation. The results obtained are presented by means of a multidimensional system composed of three elements: (a) multivariate mapping, (b) clustering with SOM, and (c) multivariate parallel coordinated plot (PCP). In each of these, the dimensions represented are multiple and refer to topographical characteristics of clustering, but also more strictly to dimensions that regard measurement of the individual input indicators in such a way as to permit the comparison both of clusters and of clusters with the values of the individual indicators calculated for the area under examination as a whole (the regional level in our case). It should be noted in this connection that the PCP is made up of as many axes as the indicators introduced into the model of spatial clustering and regionalization. Every axis is scaled using the nested means method, so that the central value of each corresponds to the value of the indicator in question for the regional territory as a whole.

Particular importance is attributed to colour in the reading of these three elements. In the multivariate mapping (a), the units (municipalities) with the same colour belong to the same cluster. In the clustering with SOM (b), node hexagons with similar colours present a lower level of dissimilarity. This difference can also be seen in the PCP (c), where lines of very different colour refer to clusters that present divergent values of the input indicators. It should also be noted that the circumference of the node hexagons in (b) and the thickness of the lines in (c) are indicative of the size of the clusters.

Figure 1 shows that a certain degree of variability exists in terms of demographic structure (c) and that this variability appears to follow particular spatial patterns, at least in a certain number of cases. In actual fact, a fairly large number of municipalities with similar demographic structure, and therefore belonging to the same cluster or to clusters with a low level of dissimilarity, quite often prove contiguous in terms of spatial location (Fig. 1a).

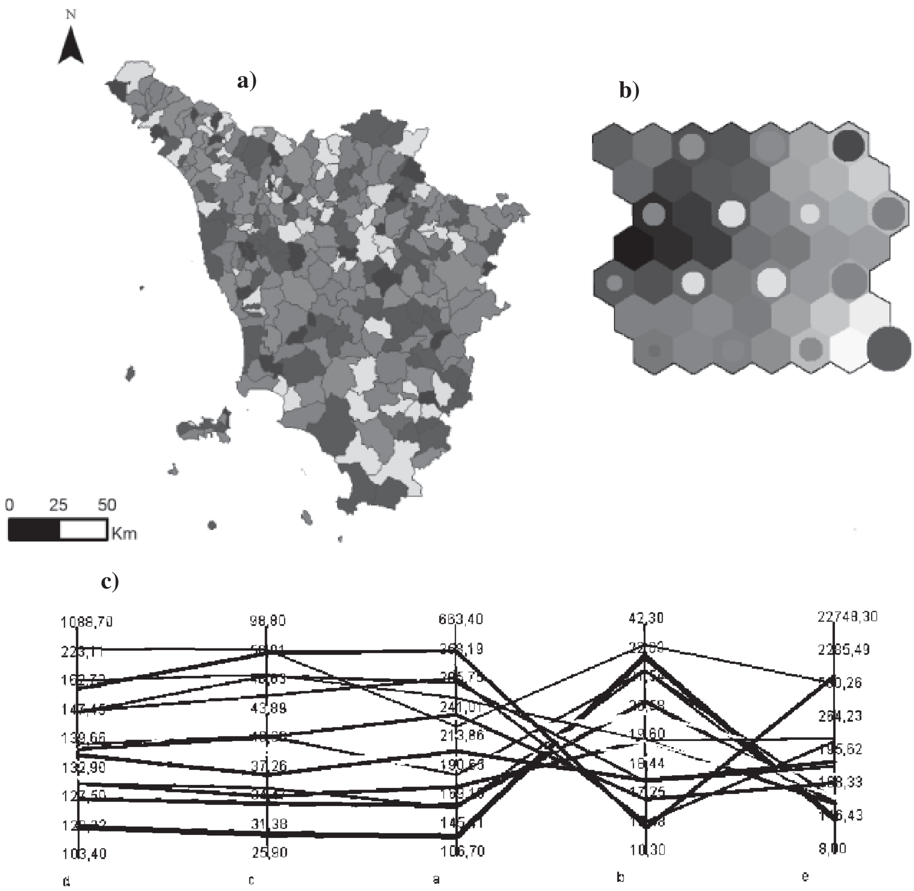

Figure 1: (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (parallel coordinate plot)

On the basis of these results, an exploratory analysis was carried out with the aim of further classification of the 16 clusters so as to identify 4 primary groups. The first is characterized by a very high proportion of old people with respect to a limited presence of young people, and is therefore called the old group (Fig. 2). It presents high values of (a) the aging index, (c) the old-age dependency index, and (d) the structure of the population of active age index.

The municipalities belonging to this first group present territorial contiguity in some cases. In terms of location, we have municipalities in the mountains and foothills in areas along the north-east border of Tuscany; municipalities in insular and coastal areas, albeit with no interruption of contiguity, and municipalities with a lower level of urbanization in the area of Chianti (Fig. 2 a).

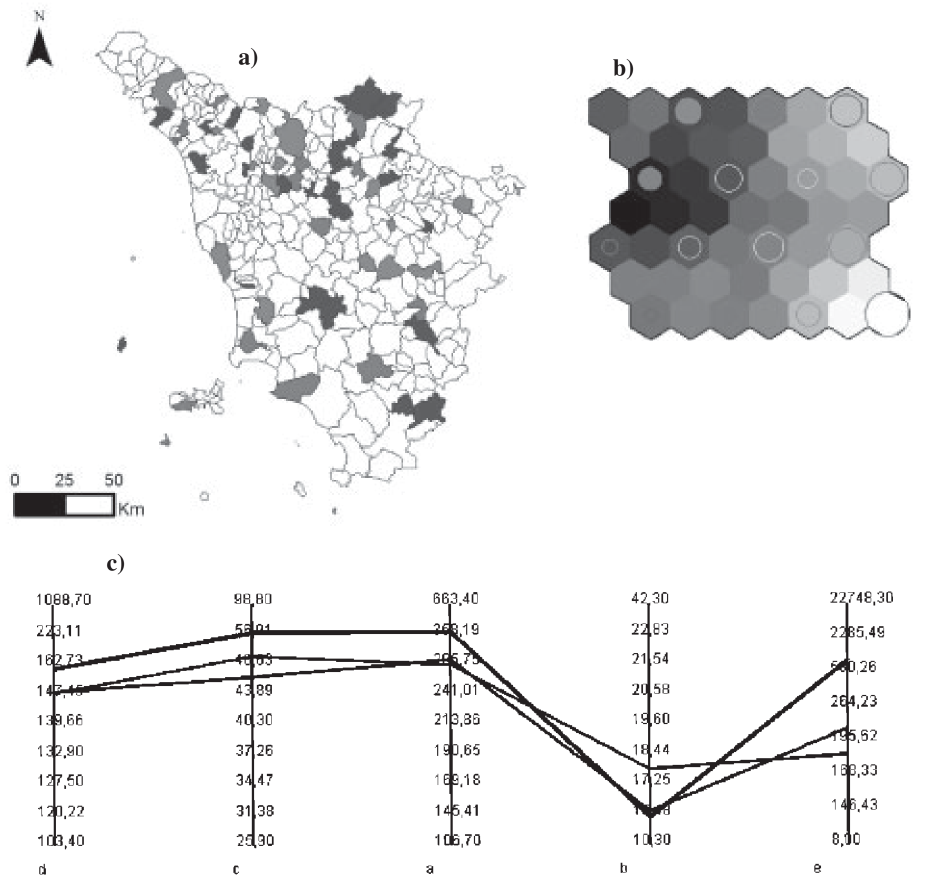

Figure 2: Group 1
– "Old" (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (parallel coordinate plot)

The second or young group consists of municipalities with the opposite age structure, i.e. a lower proportion of old people and a higher proportion of young. The demographic profile therefore proves significantly different, as shown in the PCP (Fig. 3 c). The values of (a) the aging index, (c) the old-age dependency index (d) the structure of the population of active age index, and (e) the replacement of the population of active age index prove far below the regional average, whereas the youth dependency index (b) is high (Fig. 3 c).

The municipalities belonging to this second group present a fair level of spatial contiguity. They are mostly located on the Tyrrhenian coast and in areas adjacent to the large urban and industrial agglomerations (Fig. 3 b).

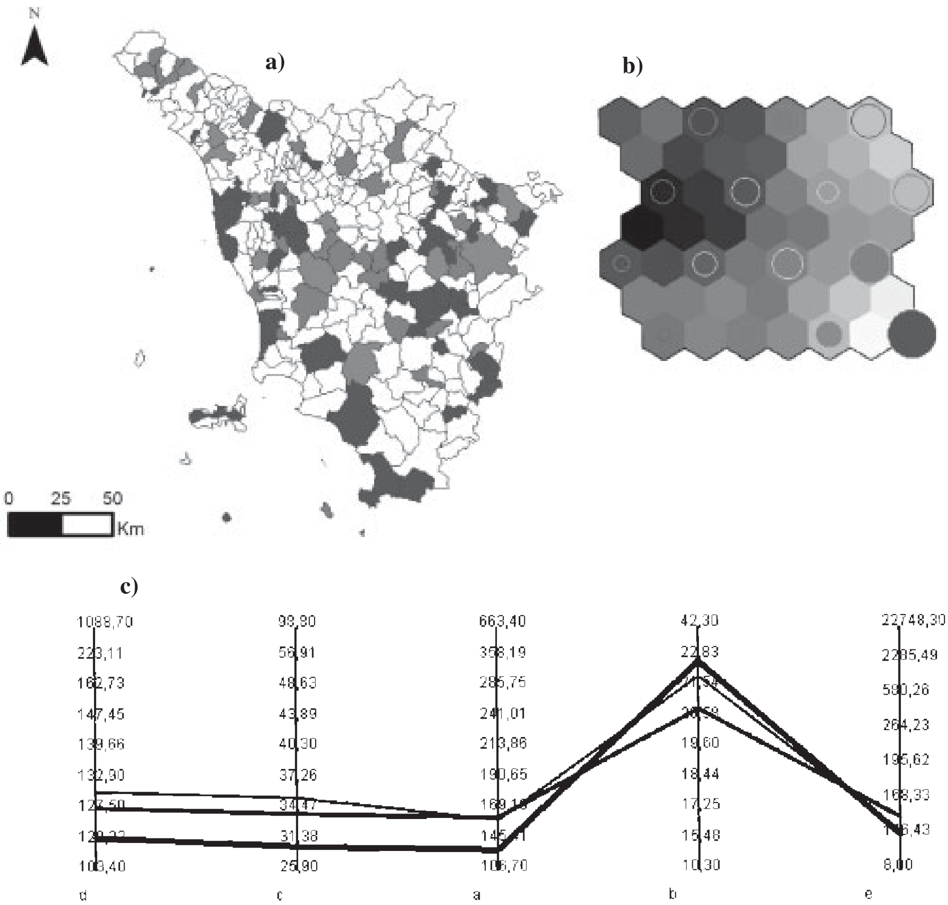

Figure 3: Group 2
– "Young" (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (parallel coordinate plot)

The third or medium group is characterized by a prevalence of municipalities with demographic profiles not far from the average regional levels located in the medium and large urban areas of Tuscany (Fig. 4 a, b and c). The fourth group ("no name") consists of municipalities that belong to none of the above groups due to their demographic characteristics and, in some cases, also their location. These residual municipalities are few in number and present no particular territorial patterns (Fig. 5 a, b and c).

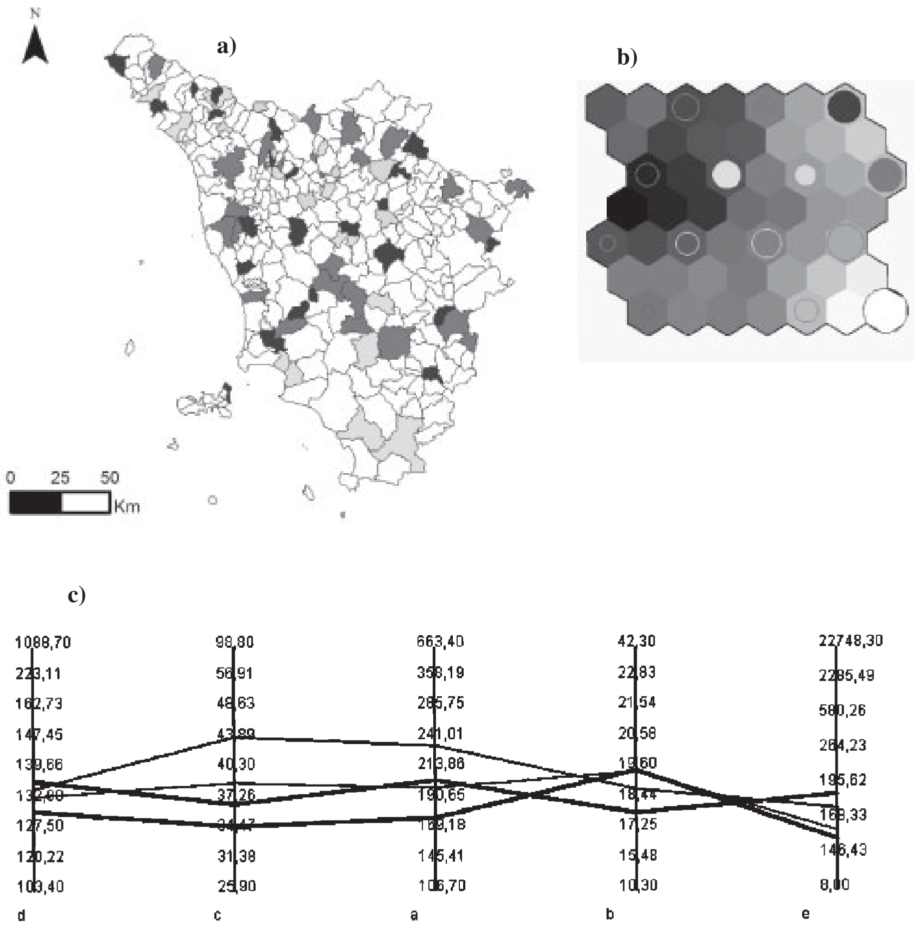

Figure 4: Group 3
– "Medium" (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (parallel coordinate plot)

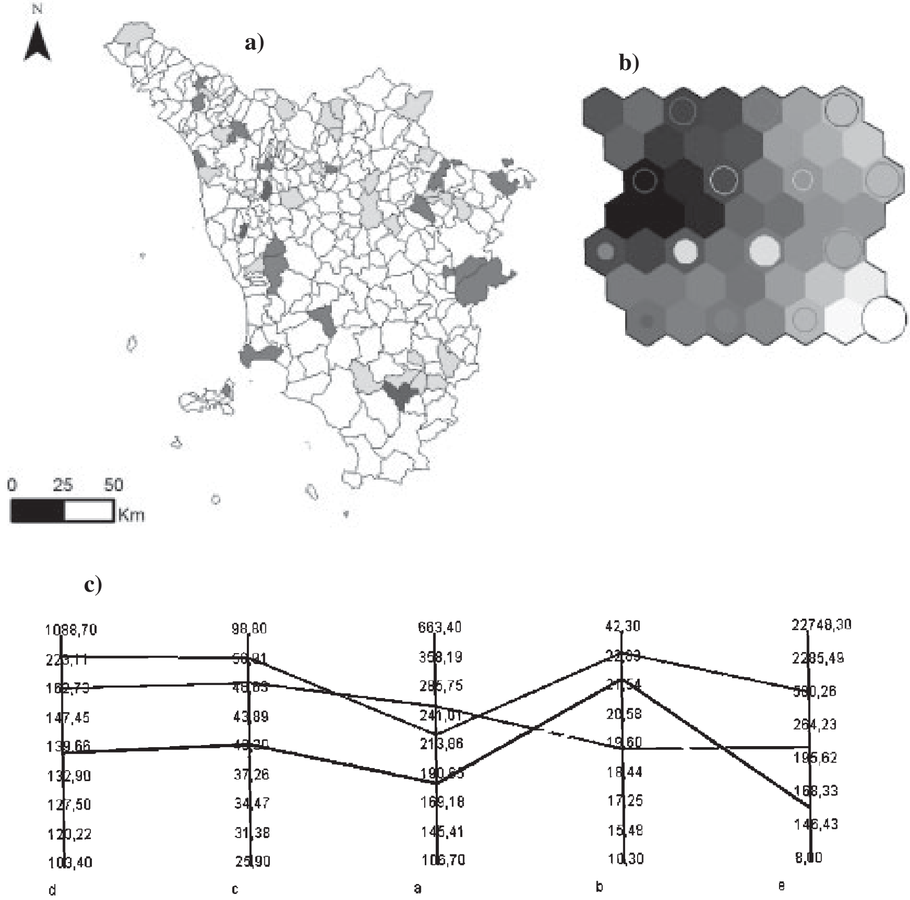

Figure 5: Group 4
– "No name" (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (parallel coordinate plot)

3.2 REGIONALIZATION (SPATIAL CLUSTERING WITH SPATIAL CONSTRAINING STRATEGY)

As suggested by Guo (2008), a Complete Linkage Clustering method combined with a Full Order Constraining Strategy (CLK - Full Order) was applied in order to identify a cluster of municipalities that are both homogeneous in terms of demographic structure and spatially contiguous. The constraint was imposed of a maximum of ten intermediate areas between the municipal and regional levels and a minimum of about 370,000 for the population resident in each area.

On the basis of these parameters, RedCap identifies seven intermediate areas between the municipal and the regional level (Fig. 6). The result that appears most

Figure 6: Regions obtained

evident is that when the constraints indicated above are imposed, the areas identified lose internal homogeneity in terms of demographic structure, which indicates that the constraint of territorial contiguity has a substantially negative effect in this sense. This prompts reflection on the ensuing implications as regards the organization of services on the part of local authorities. It appears in fact that the province should not necessarily be seen in terms of territorially contiguous areas, as this condition impairs the homogeneity of the population in terms of demographic structure and thus leads to greater fragmentation of the demand for services.

Above and beyond any considerations of an economic and political character, it would make no sense, at least in Tuscany, to take the province as a basis for organizing the provision of services in relation to the demographic characteristics of the population, i.e. the characteristics that can be plausibly regarded as determining the demand for the services in question. It would instead make sense, as shown in the first phase of municipal clustering, to target a sub-provincial grouping with urban and non-urban areas kept separate. In point of fact, the clustering of municipalities with no constraints of contiguity makes it possible to identify territories that are demographically homogeneous and sufficiently contiguous. Imposing the constraint of contiguity instead gives us territories of very little homogeneity that are therefore more difficult to organize in terms of the provision of services.

4. CONCLUDING REMARKS

The results show that the Tuscan municipalities can be grouped homogeneously with respect to the demographic structure of the resident population at a level higher than the municipal, but not corresponding precisely to the provincial. It appears in fact that the provincial territory does not present areas that are particularly homogeneous in demographic terms, which instead emerge for less extended territories made up of neighbouring municipalities that can also belong to different provinces. Given that the composition by age of the resident population and its spatial distribution are elements to be taken into consideration in the planning of policies and the organization of service for a specific territory, the operative spheres of local government bodies should be identified as areas that are not too large and do not correspond to the existing provinces. The results obtained through application of the same demographic indexes and the same method of investigation to the municipalities of the Florentine Metropolitan Area proved instead satisfactory in terms of construction of homogeneous and contiguous supra-municipal areas. This suggests that if the provinces are abolished, the right approach could be either to operate at the municipal level or to identify local super-areas around metropolitan areas (the actual existence of which is recognized, as they were instituted in 1990 by a national law that has yet to be implemented).

This naturally leads us to conclude that the aggregations of municipalities on a regional basis probably present still less demographic homogeneity than the provincial aggregations. In this sense, a future development of the study presented should involve municipalities belonging to a number of neighbouring regions. Clustering organized on a multiregional basis would probably result in aggregations of homogeneous municipalities equal in size to the provinces but made up in many cases of municipalities belonging to different regions.

Another observation regards the marked internal variability of the provinces in terms of demographic structure found in the case of Tuscany. While this is unquestionably related to the demographic indexes used for the purposes of spatial clustering, it should in any case be pointed out that these are the ones known to be used for the purpose of planning or redefining public projects.

A further development of this study, above all in terms of comparison at the regional level, regards the impact of the usually resident foreign population. While their presence is certainly reflected in the age structure of the population, closer analysis could unquestionably lead to more detailed observation.

Finally, it should be pointed out that the results of any kind of regionalization process also depend on the type and number of the variables adopted. While the demographic indexes used here are those best known and most commonly applied in order to describe the demographic structure of a given population, a system of weighting could help to pinpoint the relative importance of individual variable. For example, if services for the elderly have a greater economic impact for local government than primary schools, the regionalization could focus more attention on the older population than the younger.

REFERENCES

Angayarkkani, K. and Radhakrishanan, N. (2009). Efficient forest fire detection system: A spatial data mining image processing based approach. *International Journal of Computer Science and Network Security*, 9: 100-107.

Assuncão, R.M., Neves, M.C., Câmara, G. and Da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Sciences*, 20: 797-811.

Anselin, L. (1995). Local indicators of spatial sssociation – LISA. *Geographical Analysis*, 27: 93-115.

Anselin, L. (2003). Spatial externalities, spatial multipliers and spatial econometrics. *Regional Science Review*, 26: 153-166.

Behnisch, M. and Ultsch, A. (2010). Are there clusters of communities with the same dynamic behaviour?. In H. Locarek-Junge and C. Weihs, editors, *Classification as a Tool for Research*, 445-453. Springer-Verlag, Berling-Heidelberg.

Behnisch, M. and Ultsch, A. (2009). Urban data mining: spatiotemporal exploration of multidimensional data. *Building Research & Information*, 37: 520-232.

Benassi, F. and Ferarara, R. (2010). Regionalization with dynamically constrained agglomerative clustering and partitioning. An application on spatial segregation of foreign population in Italy at regional level. *Proceedings of the 45th Scientific Meeting of the Società Italiana di Statistica*, Padova.

Benassi, F., Ferrara, R. and Strozza, S. (2010). Verso l'individuazione di aree sovra-comunali per contiguità spaziale e omogeneità delle caratteristiche degli stranieri residenti in Campania. *Rivista Italiana di Economia, Demografia e Statistica*, 64: 47-54.

Benassi, F., Bocci, C. and Petrucci, A. (2013). Spatial data mining for clustering: an application to the Florentine Metropolitan Area using RedCap. In A. Giusti, G. Ritter, and M. Vichi, editors, *Classification and Data Mining*, 157-164. Springer Heidelberg, New York, Dordercht, London.

Bernetti, I., Ciampi, C., Saccelli, S. and Marinelli, A. (2011). La pianificazione di distretti agroenergetici. Un modello di analisi per la regione Toscana. *L'italia forestale e Montanta/Italian Journal of Forest and Mountain Enviroments*, 66: 306-320.

De Castro, M.C. (2007). Spatial demography: An opportunity to improve policy making at diverse decision levels. *Population Research and Policy Review*, 26: 477-509.

Elis-Williams, D.G. (1987). The effect of spatial population distribution on the coast of delivering local services. *Journal of Royal Statistical Society A*, 150: 152-166.

Fovell, R.G. and Fovell, M.Y.C. (1993). Climate zones of the conterminous United Stases defined using cluster analysis. *Journal of Climate*, 6: 2103-2135.

Goodchild, M.F. and Janelle, D.G. (2010). Toward critical spatial thinking in the social sicences and humanities. *GeoJournal*, 75: 3-13.

Guo, D., Gahegan, M., MacEachren, A.M. and Zhou, B. (2005). Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Sciences*, 32: 113-132.

Guo, D. (2008). Regionalization with dynamically constrained agglomerative clustering and partitioning. *International Journal of Geographical Information Sciences*, 22: 801-823.

Han, J., Kamber, M. and Tung, A.K.H. (2001). Spatial clustering methods in data mining: a survey. In H.J. Miller, and J. Han, editors, *Geographic Data Mining and Knowledge Discovery*, 1-29. Taylor & Francis, London.

Han, J., Lee, J.G. and Kamber, M. (2009). An overview of clustering methods in geographic data analysis. In H.J. Miller, and J. Han, editors, *Geographic Data Mining and Knowledge Discovery*, 2nd edition, 149-170. Taylor & Francis: London.

Haining, R.P., Kerry, R. and Oliver, M.A. (2010). Geography, spatial data analysis and geostatistics: An overview. *Geographical Analysis*, 42: 7-31.

Haining, R.P., Wise, S.M. and Blake, M. (1994). Constructing regions for small area analysis: Material deprivation and colorectal cancer. *Journal of Public Health Medicine*, 16: 429-438.

Henderson, B. (2006). Exploring between site in water quality trends: A functional data analysis approach. *Envirometrics*, 17: 65-80.

Jiao, L. and Lin, Y. (2008). Knowledge discovery by spatial clustering based on self-organizing feature map and composite distance measure. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37: 219-224.

Jin, H. and Guo, D. (2009). Understanding climate change patterns with multivariate geovisualization.

In *Proceedings of the Interntaional Conference on Data Mining Workshops*, IEE Pres: 217-222.

Koperski, K., Adhikany, J. and Han, J. (1996). Knowledge discovery for spatial database: progress and challenges. In *Proceedings of the Workshops on Research Issues on Data Mining and Knowledge Discovery,* Montreal: 55-70.

Linard, C., Gilbert, M., Snow, R.W., Noor, A.M. and Tatem A.J. (2012). Population distribution, settlement patterns and accessibility across Africa in 2010. *PlosOne*, 7: 1-8.

Livi Bacci, M. (1981*). Introduzione alla demografia*. Loescher Editore, Torino.

Long, J., Nelson, T. and Wulder, M. (2010). Regionalization of landscape pattern indices using multivariate cluster analysis. *Enviromental Management*, 46: 134-142.

Lwasa, S. (2007). Geospatial analysis and decision support for health services planning in Uganda. *Geospatial Health*, 2: 29-40.

Moran, C.J. and Bui, E.N. (2002). Spatial data mining for enhanced soil map modeling. *International Journal of Geographical Information Sciences*, 6: 533-549.

Mu, L. and Wang, F. (2008). A scale-space clustering method: Mitigating the effect of scale in the analysis of zone-based data. *Annals of the Association of American Geographer*, 98: 85-101.

Naccarato, A. (2012). Spatio-temporal weights with simultaneous effect for enviromental data. *Working Paper n. 157,* Department of Economics, Roma Tre University.

Openshaw, S. (1977). A geographical solution to scale and aggregation problems in region building, partitioning and spatial modeling. *Transaction of the Institute of British Geographers*, 2: 459-472.

Openshaw, S. and Rao, L. (1995). Algorithms for reeingineering 1991 census geography. *Enviromental and Planning*, 27: 425-446.

Osnes, K. (1999). Iterative random aggregation of small units regional measures of spatial autocorrelation of cluster localization. *Statistics in Medicine*, 18: 707-725.

Roddick, J.S. and Spiliopoulou, M. (1999). A bibliography of temporal, spatio and spatio-temporal data mining research. *SIGKDD Explorations*, 1: 34-38.

Romano, E., Balzanella, A. and Verde, R. (2010). A new regionalization method for spatially dependent functional data based on local variogram models: An application on enviromental data. *Proceedings of the 45th Scientific Meeting of the Società Italiana di Statistica*, Padova.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46: 234-240.

Voss, P.R. (2007). Demography as a spatial social science. *Population Research and Policy Review*, 26: 457-476.

Wang, H., Zhang X., Li, S. and Song, X. (2010). Spatial clustering for the regionalization of maize cultivation in China and its outlier analysis. *WSEAS Transactions on Information Science and Applications*, 7: 860-869.

White, R. and Engelen, G. (2000). High-resolution integrated modelling of the spatial dynamics of urban and regional system. *Computers, Environment and Urban System,* 24: 383-400.

Yu Pan, J. and Faloutsos, C. (2002). "GeoPlot": Spatial data mining on video libraries. *Proceedings of the Eleventh International Conference on Information and Knowledge Management* (CIKM'02): *405-412.*