THE USE OF ROUGH SET AND SPATIAL STATISTIC IN EVALUATING THE PERIURBAN FRINGE

Beniamino Murgante, Giuseppe Las Casas , Anna Sansone,

University of Basilicata , 10, Viale dellAteneo Lucano, 85100 Potenza, Italy, murgante@unibas.it

Abstract
The distinction among urban, peri-urban and rural areas inside a territory represents a classical example of uncertainty in land classification. The transition among the three classes is not much clear and can be described with Sorites Paradox, considering the residential buildings and the settlements. Peri-urban fringe can be considered as a transition zone between urban and rural areas, as an area with its own intrinsic organic rules, as a built area without formal organisation or as an abandoned rural area contiguous to urban centres. In any case, concepts as density of buildings, services and infrastructures or the degree of rural, residential and industrial activities, will lead to uncertainty in defining classes, due to the uncertainty in combining some properties. One of the methods which can be utilized is the rough sets theory, which represents a different mathematical approach to uncertainty capturing the indiscernibility. The definition of a set is connected to information knowledge and perception about phenomena. Some phenomena can be classified only in the context of the information available about them. Two different phenomena can be indiscernible in some contexts and classified in the same way (Pawlak 83). The rough sets approach to data analysis hinges on two basic concepts, the lower approximation which considers all the elements that doubtlessly belong to the class, and the upper approximation which includes all the elements that possibly belong to the class. The rough sets theory furthermore takes into account only properties which are independent. This approach has been tested in the case of study of Potenza Province. This area, located in Southern Italy, is particularly suitable to the application of this theory, because it includes 100 municipalities with different number of inhabitants, quantity of services and distance from the main road infrastructures.

Keywords: kernel density, straight line distance, point pattern analysis, rough sets.

1 Introduction

In recent times the expression “periurban area” has been used frequently in planning documents. This expression, despite its large use, does not have a clear and unambiguous definition, which can provide precise indications to allow the exact edge of these zones.

The major reasons of this deficiency can be identified in the complexity of the phenomenon to be analyzed and in the huge variety of territorial contexts in which it can reveal. The expression “periurban fringe” evokes different meanings following the various disciplinary approaches to the problem.

For economists periurban areas represent a disorganized city expansion that involves an enormous increase of costs in services management and infrastructures maintenance. For ecologists periurban fringes are zones in which the natural resources have been irreparably damaged because of the large concentration of economic activities which provoke an huge deterioration of the environment. Agronomists take into account the agricultural loss of productivity due to constantly increasing presence of buildings.

Urban planners have two different approaches: the first considers the phenomenon from a theoretical point of view in comparison with the consolidated concepts of city and rural area, the other one takes into account the increase of the economic value of real estate due to the transformation.

The phenomenon is so complex to analyze, than even the local laws denote a certain superficiality, lacking of precise indications. Although local laws face correctly the problem giving different definitions, respecting different territorial peculiarities, on the other hand they consider only the proximity to the urban areas. It is obvious that contiguity condition alone is not enough to define such a complex phenomenon.

A clear definition of periurban fringe can be achieved considering this zone as an area with its own intrinsic organic rules, like the urban and the rural ones and not as a transition zone from urban to rural areas.

An accurate rule which defines the periurban fringe can be composed by the combination of other rules. For instance, proximity to urban areas, contiguity to road network, presence of utilities and urban services, density of population higher than in rural areas can generate a set of inclusive rules and if some of these rules are satisfied, the area can be included in periurban fringe. In the same way, exclusive rules considering archaeological sites, heritage areas, environmental preservation areas, deep slope terrains, landslide areas, erosion areas can be achieved. If even one of this rules is satisfied, the area cannot be included in periurban fringe.

The aim of this paper is to define more precise rules in order to describe the periurban phenomenon, using techniques of spatial statistic, such as point pattern analysis. This approach has been tested in the case of study of Potenza Province. This area, located in Southern Italy, is particularly suitable to the application of this method, because it includes 100 municipalities with different number of inhabitants, quantity of services and distance from the main road infrastructures.

2 Techniques of spatial analysis and rough set theory

In this study two techniques of spatial analysis have been used: kernel density and straight linens distance. These two approaches are call effects of first and second order and they consider respectively the amount of events observed per unit area and the distance among them.

All geographic themes achieved with these two techniques are combined with the rough set theory in order to define in “indiscernible” way the periurban fringe.

2.1 Kernel density

The kernel density is a technique of point pattern analysis that generates a grid from a point vector source. The numerical attributes associated to vector data can be considered as a weight increasing intensity of the phenomenon.

Bailey and Gatrell (1995) have developed a set of spatial analysis techniques applied in the field of spread of epidemics. These techniques are based on Waldo Tobler (1970) first law of geography: “Everything is related to everything else, but near things are more related than distant things”. Compared with classical statistical approaches, it is important to localize the data, considering the events as spatial occurrences of the considered phenomenon. Each event L_i is localized in an unambiguous way in the space from its coordinates x_i, y_i.

An event L_i is a function of its position and attributes which characterize it and quantify the intensity:

$$\textbf{Eq. 1}\quad L_i = (x_i, y_i, A_1, A_2, \ldots, A_n)$$

While the simple density function considers the number of events for each element of the regular grid that composes the study region R, the kernel density takes into account a mobile three-dimensional surface, which weighs the events according to their distance from the point from which the intensity is estimated (Gatrell *et al*., 1996).

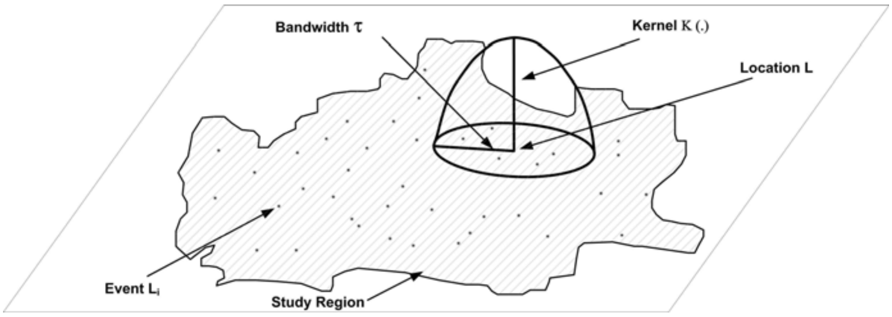

Fig. 1 Kernel Density Estimation (adapted from Bailey and Gatrell, 1995)

The density (intensity) of the distribution in point L can be defined by the following equation:

$$\textbf{Eq. 2} \quad \widehat{\lambda}(\mathrm{L}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{1}{\tau^2} k\left(\frac{L - L_i}{\tau}\right)$$

Where, λ(L) is the intensity of point distribution, quantified in the point L; Li is the event number i, k() represent the kernel function and τ the bandwidth. τ can be defined as the radius of a circle generated from the intersection between the surface and the plan which contains the study region R (fig. 1). Two major factors influence the results: the dimensions of the grid and the bandwidth (Batty et al. 2003). The bandwidth produces a three-dimensional surface, more or less corresponding to the phenomenon, allowing to analyze its distribution at different scales.

The choice of the bandwidth remarkably influences the surface of estimated density. If the bandwidth is high, the kernel density is closer to the values of the simple density.

With a narrow bandwidth, the surface will capture local events, with density close to zero for the elements of the grid far from each event (fig. 2).

The right bandwidth can be determined by estimating the phenomenon and, if it is important, by highlighting the peaks of distribution or a smooth spatial variation.

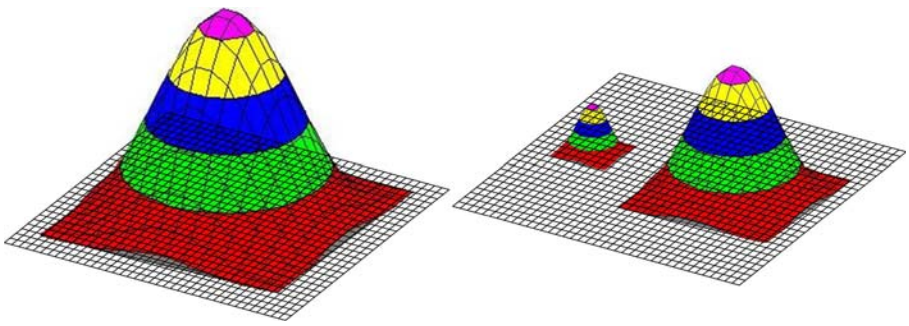

Fig. 2 Relation between the bandwidth dimension and the study area

2.2 Straight line distance

The straight line distance is a function which measures the distance between every cell and the nearer source. The source can be in vector or grid format. In the case of grid format some cells will contain information about the source and the others will not have values, while in the case of a vector theme it will be necessary a previous transformation in grid before determining the distance. The output of straight line distance is in grid format and the distance is measured between the barycentre of the cells. Also, in this case it is important to estimate some factors such as the maximum distance within which one has to assess measures and dimensions of the cells.

2.3 Rough set theory

An ontology of uncertainty in spatial information is reported in figure 3. The attention will be focused on the poorly defined data. Some data are said to be vague if their meaning is not clear, whereas ambiguous data can have at least two specific meanings. For instance, ambiguity of meanings leads to a discordance in data classification due to a different perception of the phenomenon. Vagueness, which is contrary to the Boolean concept, widely diffuse in the representation of geographical information, can be addressed by multi-valued logic and applications of fuzzy set theory. Inaccuracy produces uncertainty in the case of low quality of data, due to a high percentage of errors.

Rough set theory (Pawlak, 1982) is certainly less famous than the others outlined in figure 3, and it classifies elements in indiscernible way, using the available information. In other words the elements of the universe can be seen in a certain way in the context of an available information about them. The direct consequence is that two different elements can be indiscernible in some circumstances and in other contexts they can belong to different classes.

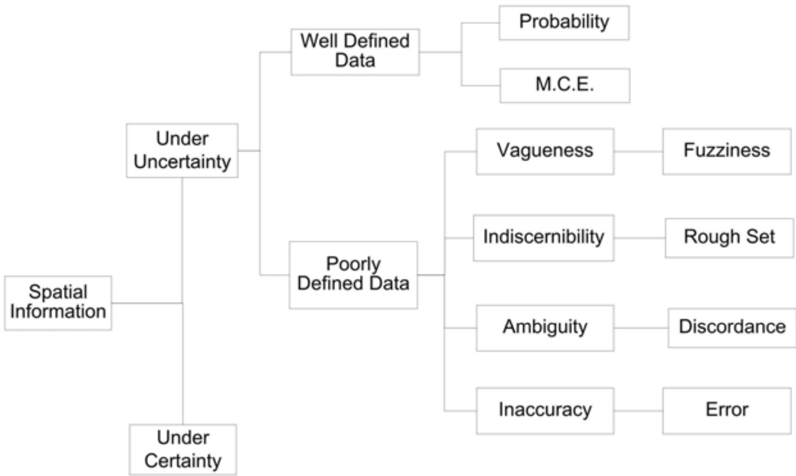

Fig. 3 An ontology of uncertainty in spatial information (adapted from Las Casas and Murgante, 2004).

This methodology is based on the concepts of indiscernibility relation, upper and lower approximation and accuracy of the approximation.

If we consider a set of defined objects U= $X_1, X_2, \ldots, X_n$ and a set of attributes A= $a_1, a_2, \ldots, a_n$, for each set of attributes B⊂A it is possible to define the following indiscernibility relation: two elements Xi and Xj are indiscernible by the set of attributes B in A if b(Xi)=b(Xj) for each b⊂B. The equivalence class of Ind(B) *B* is called elementary set. Lower and upper approximation (figure 4) are defined respectively as the elements that are contained with certainty in the set and as the objects which probably belong to the set. The difference between upper and lower approximation defines the boundary of set X.

Eq. 3 $LX = \{x_i \in U[x_i]_{ind(B)} \subset X\}$ Lower approximation

Eq. 4 $UX = \{x_i \in U[x_i]_{ind(B)} \cap X \neq 0\}$ Upper approximation

Eq. 5 $BX = UX - LX$ Boundary

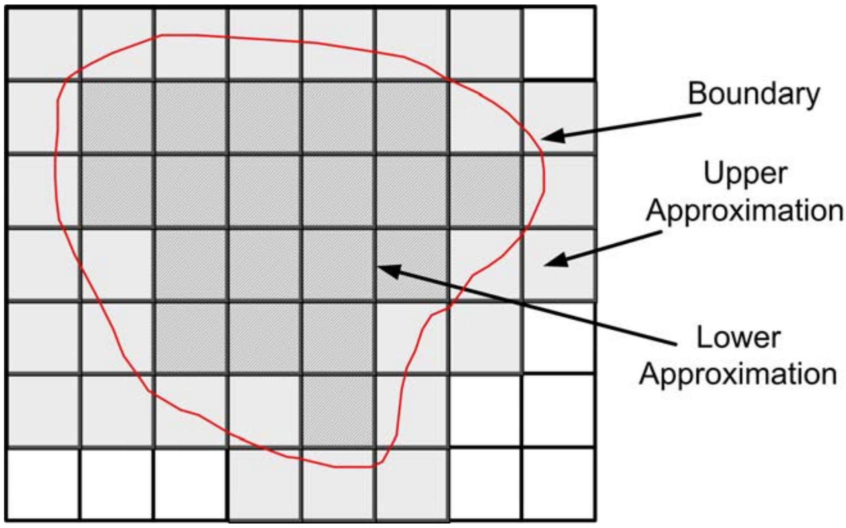

Fig. 4 Boundary, Upper and Lower Approximation of a set X.

The accuracy is defined as the ratio of the cardinality of the lower and the upper approximation:

$$\textbf{Eq. 6} \quad \mu_B(X) = card(LX)/card(UX)$$

The result must be included between 0 and 1.

3 The case study

Potenza Province is an area in Southern Apennine with a low settlement density and a population of 400,000 inhabitants, distributed on 650,000 hectares.

The main town of the province is Potenza with 70,000 inhabitants, while the demographic consistency of the other municipalities can be classified in three groups. Twelve towns have more or less 12,000 inhabitants, twenty municipalities have a population around 5,000 inhabitants, the population of the remaining 68 municipalities varies from 700 to 2,000 inhabitants (fig. 5).

One of the phenomena more frequent in western European countries is the spreading of scattered settlements. Generally this phenomenon is not likely to be found in territories with low density of population. Small Municipalities, in spite of depopulation, present a huge diffusion of scattered buildings due to the abandonment of old town centres. In bigger towns, diffusion of settlements in rural areas is determined by the high cost of flats in urban areas. An accurate analysis of the phenomenon has been carried out using all the potentialities of Geographical Information Systems. All the polygons which represent the buildings have been converted in points. All these data have been updated and integrated using digital orthophotos. These analyses have highlighted that the phenomenon begins immediately outside the areas planned for future settlements by masterplans.

The cartographic update has allowed to estimate the effects of such diffused phenomenon on environment. The conversion procedure in vector points has been useful in order to use kernel density function. This function has had multiple uses, from the epidemics localization (Gatrell et al. 1996) to the studies on spreading of the city services (Borruso and Schoier 2004), while has not been used enough in the field of territorial planning.

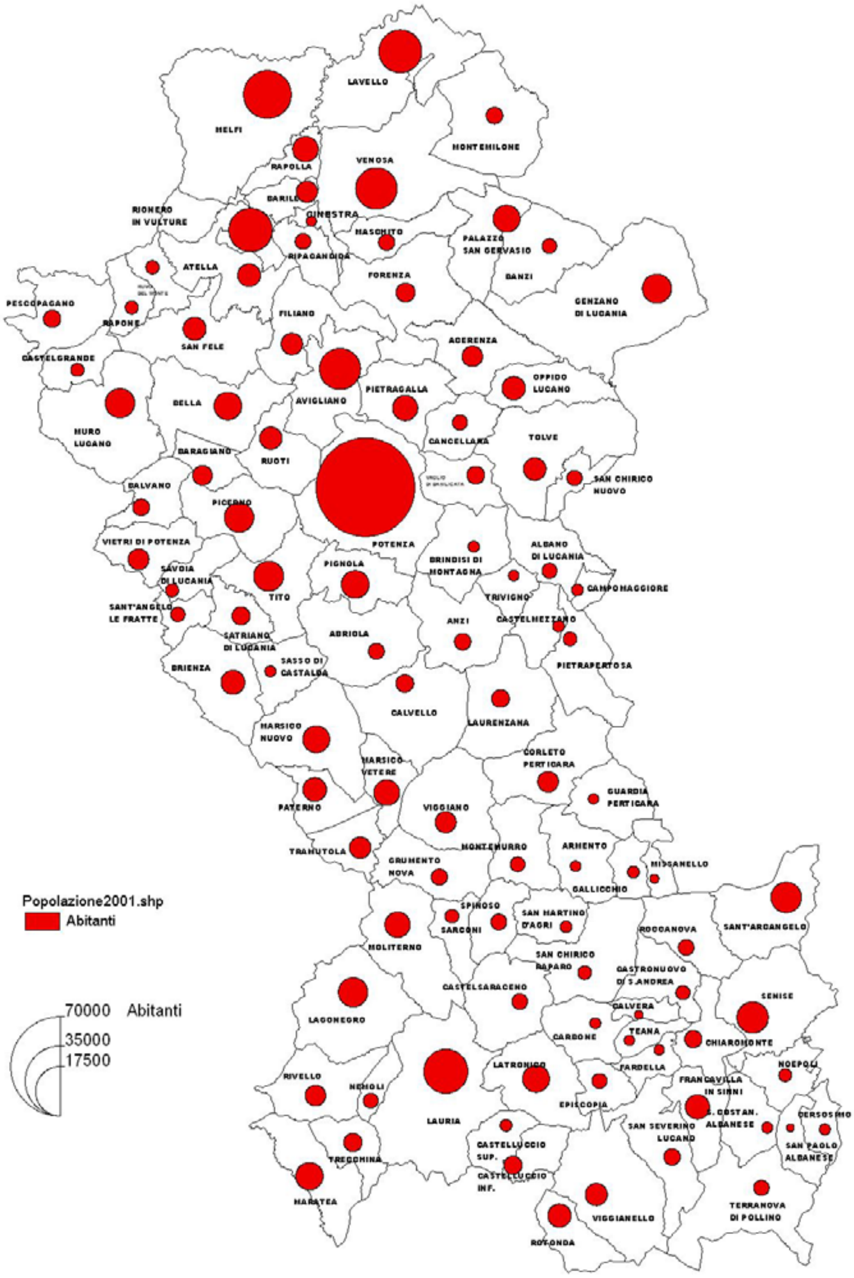

Fig. 5 Demographic dimension of Potenza Province Municipalities (2001)

The point theme of the buildings has been compared with the census zone. In urban areas each census zone corresponds to one building, while in rural areas a census zone coincides with a small rural centre; another census zone contains the scattered buildings outside the small rural centres. Considering a homogenous typology of buildings inside each census zone it has been possible to associate the number of dwelling-houses to each building. This attribute of the point theme has been used as a weight in kernel function, in order to understand the impact of the number of dwelling-houses on a particular area. In the case study it has been used a value of the bandwidth of 400 m and a cell dimension of the grid of 10 m.

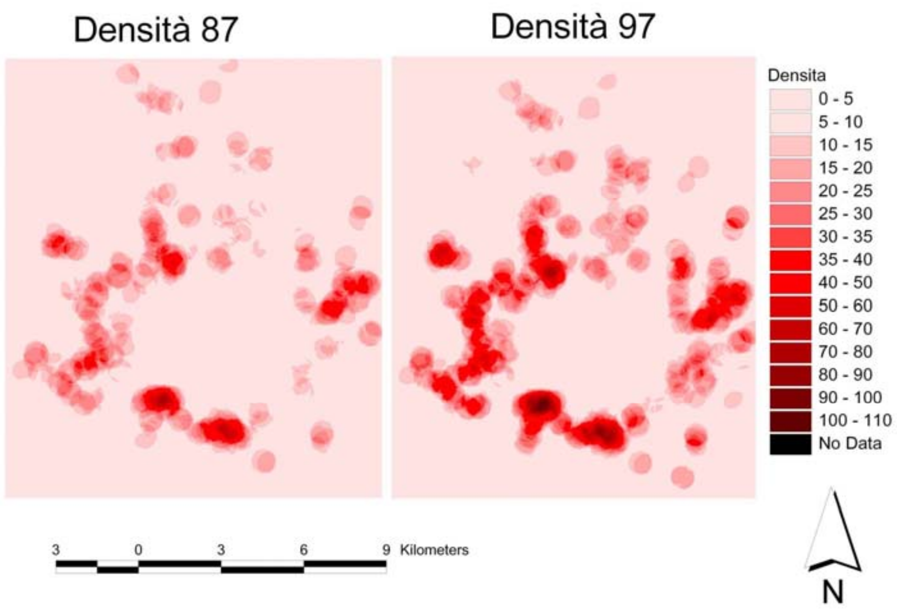

Fig. 6 Scattered settlements in 1987 and 1997

An interesting result has been obtained comparing scattered settlements in 1997 to those in 1987. It was possible to modify the point theme from orthophotos taken in two different dates(fig. 6). This comparison has allowed to characterize more precisely zones where the phenomenon of settlement dispersion has had a huge growth.

After the localization of areas where the phenomenon is more considerable, it is important to understand the factors which could increase it.

Considering the relationship between scattered buildings and areas in landslide and in high slope, it can be noticed, that these factors do not discourage settlement growth. Rural buildings, in fact, have been realized indiscriminately on landslides and on high slopes.

Comparing the theme achieved with the kernel density with other themes (fig. 7) it is possible to give further interpretations of the phenomenon of settlement dispersion. Scattered settlements can be localized as crown around the urban area or along the main line of road network. From figure 5, it appears more clear that the greater increase of settlement dispersion has occurred in north western and south zones, mostly situated on the mountains. In north eastern zones there is a completely different situation: scattered buildings are considered as a threat for intensive agricultural production.

In the study case the kernel density has been considered according to the following classes:

- it is reasonable to classify a region as rural if the presence of houses is less than one per hectare;
- from 1 to 5 houses per hectare, it is possible to define the periurban class;
- urban features are predominant beyond 5 houses per hectare.

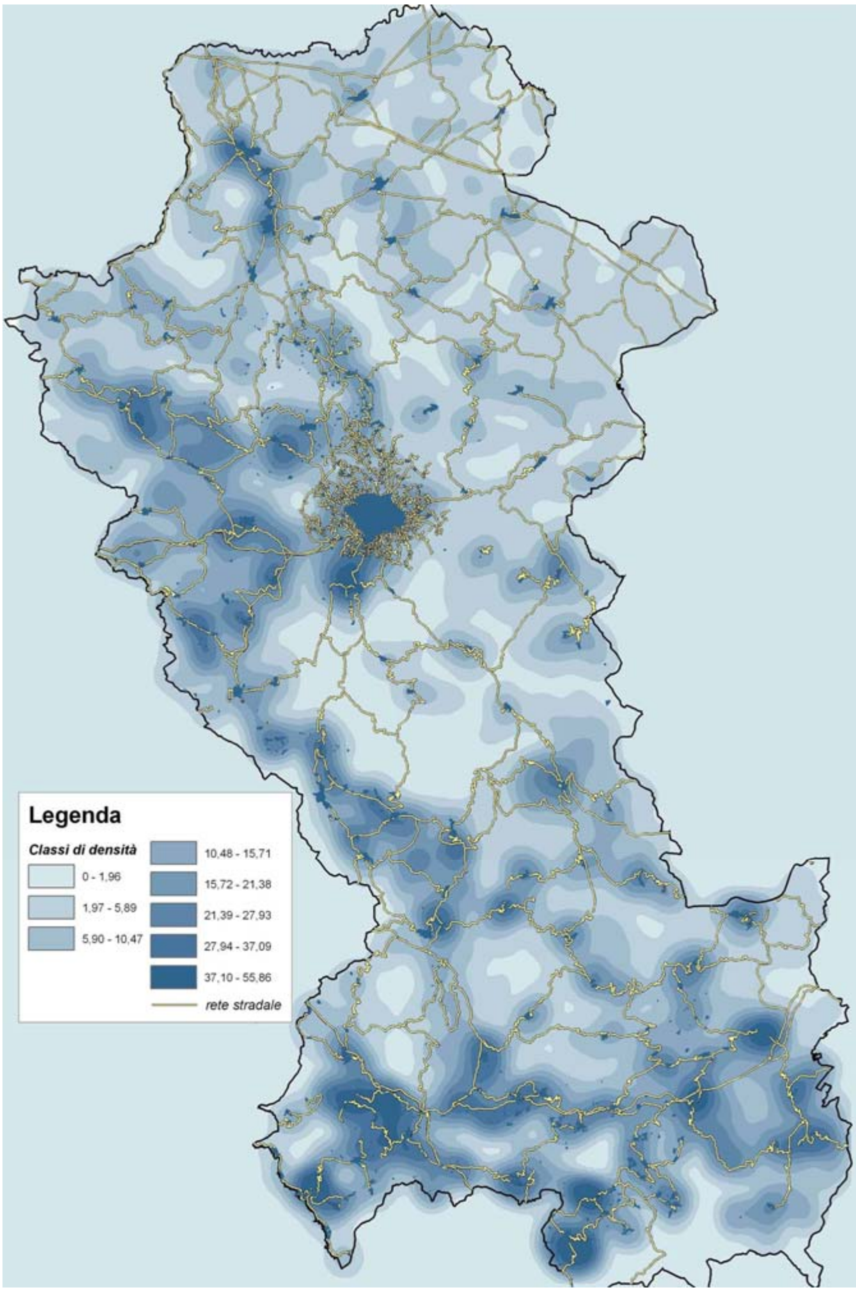

Fig. 7 Density of scattered settlements compared to road network

Depth of contiguity zone, for each centre, has been localized using a shape index for the boundary of the urban area.

The shape index is the ratio between the perimeter of the urban area and the perimeter of the circle that inscribes it. It is obvious that such an index can assume values greater than one. The more the value is greater than one, the more the shape of the settlement will be jagged, long and narrow.

A good level of compactness corresponds to a shape index comprised between 1 and 1.6, a medium level to values between 1.61 and 2.4, a low level if to an index greater than 2.4. In the following table, the 100 municipalities of Potenza Province are grouped in three classes according to compactness rate.

	Index value	Number of Centres
Good compactness	1-1,6	12
Medium compactness	1,61-2,4	68
Poor compactness	2,41-4,81	20

Tab. 1 Centre classes based on compactness rate

Table 1 shows the low level of compactness of the urban areas of Potenza province: settlement dispersion is spread along the main roads without following the morphological features. Two zones have been considered for the contiguity. The first is the ratio between area and perimeter of the urban region, the second is the ratio between area and perimeter of the circle inscribing the urban region. Figure 8 shows as the ratio area/perimeter depends on shape and dimensions. Proximity to the road network has been determined with the straight line distance assigning to each cell a distance value. The phenomenon of settlement dispersion has been observed in various municipalities with different dimensions and this study indicates that new buildings are completely within a distance of 200 m from the road networks. This distance has been considered between the inclusive rules. The following exclusive rules have been considered: area included within a distance of 150 m from rivers and streams, slope higher than 35%, Nature 2000 sites, hydro-geological risk zones.

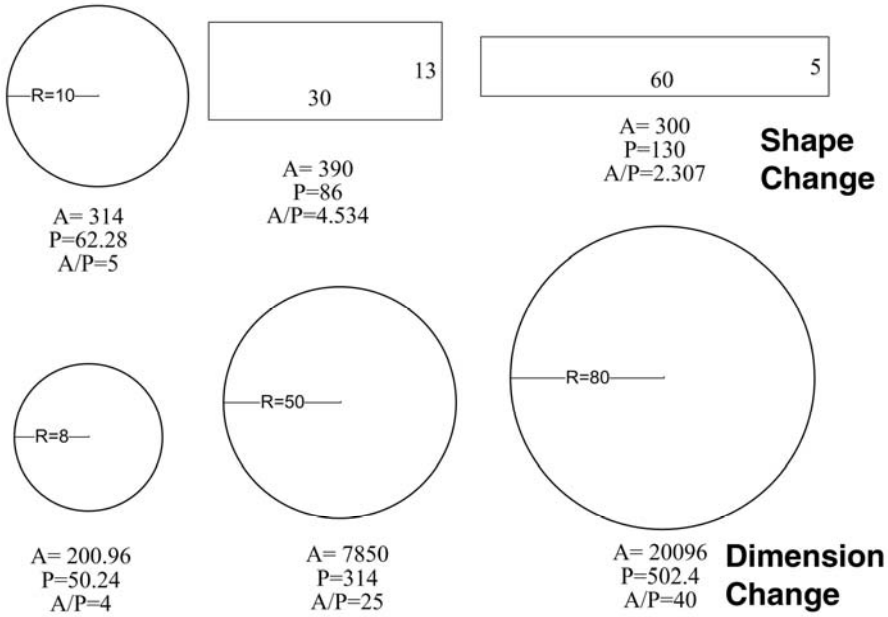

Fig. 8 Changes in area/perimeter ratio

4 Results and final discussion

Rough Set theory has been applied for the classification of different geographic layers. In the case study the entire provincial territory constitutes the universe [U], the inclusive rules and exclusive rules are the attributes [A] and the objects (the cells of the grid) have been classified following the rules previously defined. The indiscernibility relations have been calculated in [U] considering six attributes, obtaining the subset [X]. Therefore, all the cells which satisfy at the same time the exclusive and the inclusive rules are assumed as a subset [X], contained in [U]. The set [X] of all provincial territory and a detail of southern zone is reported in figure 9.

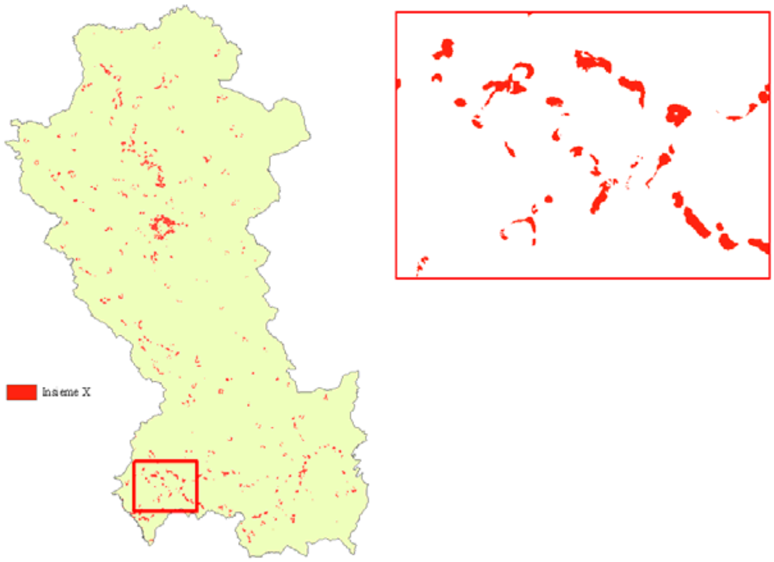

Fig. 9 Set [X] of all provincial territory with a detail of southern zone.

Lower and Upper approximation of [X] have been defined considering the two contiguity rules previously described. For both cases Lower and Upper approximation, Boundary and Accuracy have been calculated.

In the first case the lower approximation belongs to the set [X] and at the same time satisfies the first contiguity rule. Figure 10 shows the results of this procedure with a detail of the municipalities of Rivello and Nemoli.

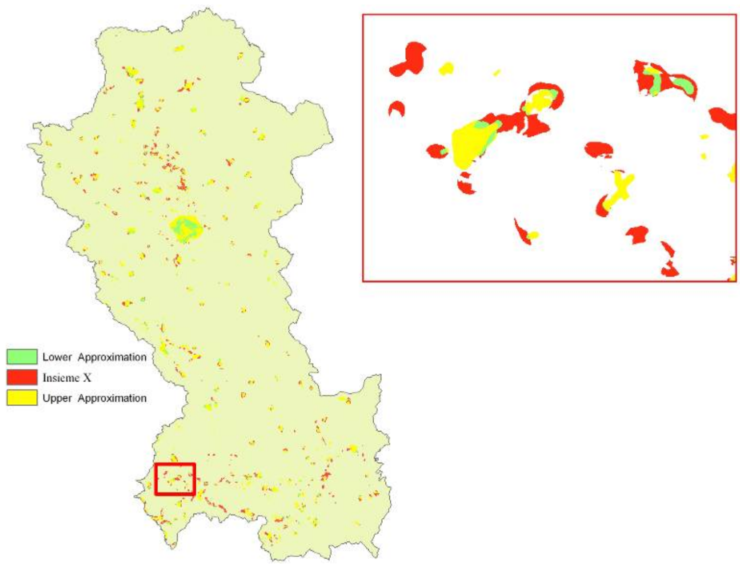

Fig. 10 Set [X]: Lower and Upper approximation of all provincial territory with a detail of the municipalities of Rivello and Nemoli.

The difference between Upper and Lower approximation defines the boundary of set [X] represented in figure 11.

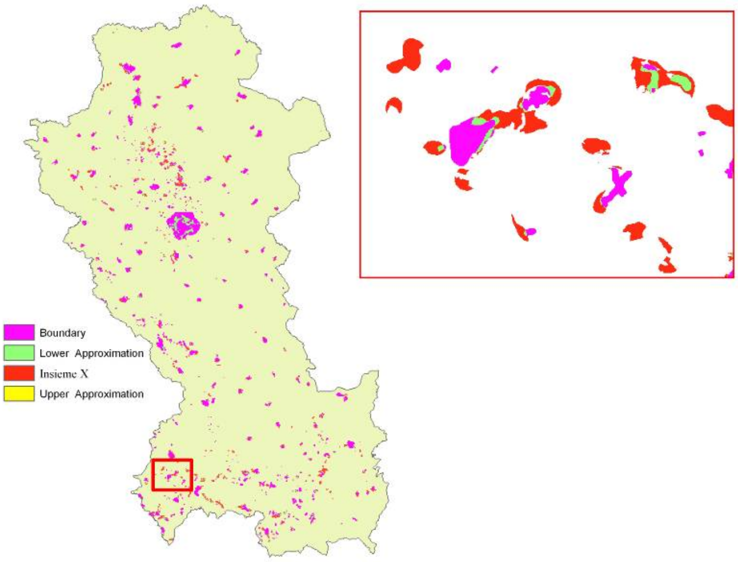

Fig. 11 Boundary of set [X] of all provincial territory with a detail of the municipalities of Rivello and Nemoli.

In the case under examination the accuracy is 49,38%: this result testifies that [X] is contained in [U], and the Lower and Upper approximation can represent the phenomenon.

The same procedure is followed for the second contiguity rule. Figure 12 shows Lower and Upper approximation and the Boundary of set [X].

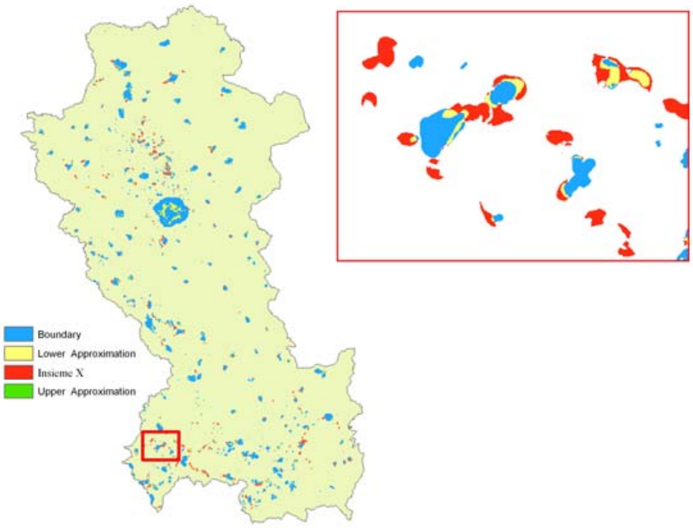

Fig. 12 Lower and Upper approximation, and the Boundary of set [X] with the second contiguity rule.

The accuracy in this case is 48,55%, leading to the consideration that the periurban phenomenon is better represented when considering the first contiguity rule.

5 References

Bailey T. C., Gatrell A. C. (1995), *Interactive spatial data analysis*, Prentice Hall.

Batty M., Besussi E., Maat K., Harts J. (2003), "Representing Multifunctional Cities: Density and Diversity in Space and Time" CASA Working Papers, ISSN: 1467-1298

Borruso G., Schoier G. (2004), "Density Analysis on Large Geographical Databases. Search for an Index of Centrality of Services at Urban Scale", in Gavrilova M. L., Gervasi O., Kumar V., Laganà A., Mun Y. and Tan K. J. (eds.), Lecture Note in Computer Scienze Springer-Verlag, Berlin, ISBN 3-540-22056-9

Gatrell A. C. , Bailey T. C., Diggle P. J., Rowlingson B. S., (1995) "Spatial point pattern analysis and its application in geographical epidemiology" Transaction of institute of British Geographer

Las Casas G., Murgante B. (2004) "Crisp and Fuzzy Sets for the Land Suitability Analysis" In Fendel E.M., Rumor M. (eds.), Information systems and the delivery of the promised societal benefits, Proceedings of the 24 Urban Data Management Symposium October 27-29, 2004, Chioggia, Venice, Italy

O'Sullivan D., Unwin D., (2002), "Geographic Information Analysis" John Wiley & Sons ISBN: 0471211761

Pawlak Z. (1982) Rough sets. International Journal of Information & Computer Sciences 11, 341-356.

Tobler, W. R. (1970). A Computer Model Simulating Urban Growth in the Detroit Region. Economic Geography, 46: 234-240.