A spatial rough set for extracting the periurban fringe

Beniamino Murgante*, Giuseppe Las Casas*, Anna Sansone*,

* University of Basilicata, Viale dell'Ateneo Lucano 10, 85100, Potenza
beniamino.murgante@unibas.it

Abstract. To date the availability of spatial data is increasing together with techniques and methods adopted in geographical analysis. Despite this tendency, classifying in a sharp way every part of the city is more and more complicated. This is due to the growth of city complexity. Rough Set theory may be a useful method to employ in combining great amounts of data in order to build complex knowledge about territory. It represents a different mathematical approach to uncertainty by capturing the indiscernibility. Two different phenomena can be indiscernible in some contexts and classified in the same way when combining available information about them. Several experiences exist in the use of Rough Set theory in data mining, knowledge analysis and approximate pattern classification, but the spatial component lacks in all these research streams.
This paper aims to the use of Rough Set methods in geographical analyses. This approach has been applied in a case of study, comparing the results achieved by means of both Map Algebra technique and Spatial Rough set. The study case area, Potenza Province, is particularly suitable for the application of this theory, because it includes 100 municipalities with a different number of inhabitants and morphologic features.

1 Introduction

In a few years' time a transition has been occurred from the traditional town, characterized by static social contexts, to today's city more dynamic and very hard to control in small details. Until few decades ago the social structure of the city was characterized by a population with strong social ties, whose life was oriented by institutions, rules, authorities.

These inhabitants are now leaving the historic part of the cities more and more occupied by transition population (students, tourists, etc.). Often the centre of town is a big shopping centre with museums, libraries and other services but without residents. The new population neither has roots in those places nor the prospect of living there for the whole life. Older inhabitants have moved out of the urban area by creating a sort of dispersed spatial form (Indovina, 1990). This phenomenon occurs on the fringe of urban areas through progressive "coagulation" of buildings. Neighbourhoods without centre and with poor social relationships have been realized. These form of urban sprawl has been encouraged by the increase of number of infrastructures, the growth of income and demand for goods and services. Urban sprawl can be considered a long-term trend for successful economic-territorial systems (Camagni, et al. 2002), characterized by soil consumption generating loss of competitiveness for agricultural activities (Murgante, et al. 2008).

The term periurban area has been recently coined in order to represent this sort of transition city and it has been frequently used in planning documents. One might ask: “what does periurban area mean, exactly?”. This term is not fully understood from planners, because planning systems do not give a clear and unambiguous definition. Urban planners have two different approaches: the first one 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 real estate economic value, due to the transformation.

These different points of view generate uncertainty in defining the exact edge of these zones. Consequently Periurban fringe can be considered as an object with indeterminate boundaries (Burrough and Frank 1996).

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

For instance, proximity to urban areas, contiguity to road network, presence of utilities and urban services, population density higher than in rural areas can generate a set of inclusion rules and if some of these rules are satisfied, the area can be included in the periurban fringe. In the same way, exclusion rules consider archaeological sites, heritage areas, environmental preservation areas, steep slope terrains, landslide areas, erosion areas. If even one of this rules is satisfied, the area cannot be included in the periurban fringe.

The possibility of providing different degrees of land suitability has become more and more feasible in site selection by the use of geographical information systems combined with evaluative methods. In the literature about land classification, a relevant number of experiences combining GIS with multicriteria methods or fuzzy set approach exists (Malczewski 1999, Eastman 1999, Thill 1999, Leung 1988, Murgante and Las Casas 2004). Only few experiences exist in the use of Rough Set to compute spatial information. This approach has been tested in the case of study of Potenza Province. The phenomenon of urban sprawl is very common, despite the small number of inhabitants in the studied region.

2 An overview of rough set theory

Rough set theory (Pawlak, 1982) is based on the hypothesis that each element is associated to several information in the Universe of discourse. Some objects are indiscernible from others if they are classified in the same way according to their associated information. In other words, two different elements can be indiscernible in some circumstances while in other contexts they may belong to different classes. This methodology is based on the concepts of indiscernibility relation, upper and lower approximation and accuracy of the approximation.

In an *Information System* IS=(U,A), U is a set of defined objects U= $X_1, X_2, \ldots, X_n$ and A is a set of attributes A= $a_1, a_2, \ldots, a_n$. U and A are non empty sets. For each attribute $a \in A$, we associate a set of values V_a called domain of a. It is possible to define an attribute in order to classify all cases. A *Decision System* $DS=(U,A \cup d)$ is an information system in which a decision attribute, d, affects the classification.

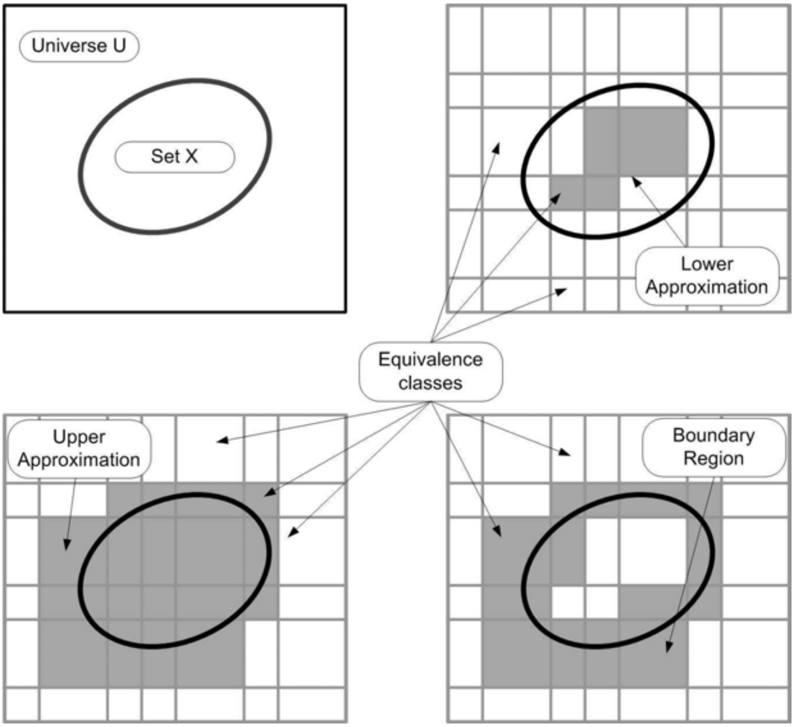

FIG. 1 – *Boundary, Upper and Lower Approximation of a set X.*

If we consider a set of attributes B⊂A, it is possible to define the following indiscernibility relation Ind(B): two elements (Xi, Xj)∈ Ind (B) 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) is called elementary set in B. Lower and upper approximation (figure 1) are defined as the elements contained with certainty in the set (equation 1) and as the objects which probably belong to the set (equation 2), respectively. The difference between upper and lower approximation defines the boundary of set X (equation 3).

$$LX = \{x_i \in U [x_i]_{nd(B)} \subset X\} \qquad \text{EQ. 1}$$

$$UX = \{x_i \in U [x_i]_{nd(B)} \cap X \neq 0\} \qquad \text{EQ. 2}$$

$$BX = UX - LX \qquad \text{EQ. 3}$$

If BX = 0 then the set X is Crisp. The accuracy (equation 4) is defined as the ratio of cardinality of lower and upper approximation:

$$\mu_B(X) = card(LX)/card(UX) \qquad \text{EQ. 4}$$

The result must be included between 0 and 1.

3 A spatial extension of rough set theory

Geographical applications of this methodology are not much experimented. A first theoretical approach has been developed by Worboys and Duckham (2004) and Beaubouef et al. (2007). Spatial information can be classified following two criteria: nature of data and information reliabilities. A classification of uncertainty in spatial information is reported in figure 2.

Geographical information can be certain or not. In the case of well defined data, a degree of uncertainty can be eliminated using probability theory or, in case of several alternatives, it can be solved by adopting multicriteria methods. Poorly defined data have been classified in three groups.

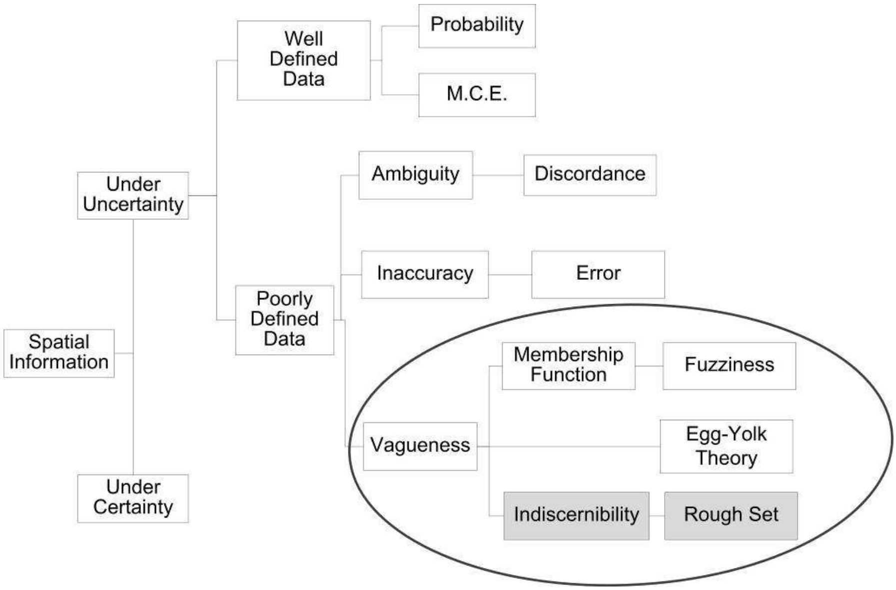

FIG. 2 – *A Classification of uncertainty in spatial information (adapted from Murgante and Las Casas, 2004).*

Some data are said to be ambiguous if they can have at least two particular interpretation. Ambiguity leads to a discordance in data classification due to a different perception of the phenomenon. Inaccuracy produces uncertainty in the case of low quality of data, due to a certain degree of error.

Great part of GIS functionality is based on Boolean operators which are founded on a two-valued logic. Vagueness (Erwig and Schneider, 1997) takes into account multi-valued logic and it is based on the concept of "boundary region" which includes all elements that cannot be classified as belonging to a set or its complement (Pawlak, 1998).

Three theoretical approaches to vagueness exist: the first one is based on fuzzy set theory (Zadeh, 1965) which accounts for partial membership of elements to a set; the second is Egg-Yolk Theory (Cohn and Gotts 1996, Hazarika and Cohn 2001) based on the concepts of "egg", i.e. the maximum extension of a region, and "yolk", i.e. the inner region boundary; the third approach is rough set theory.

It is possible to define the rough object in analogy with the experience of Cheng et al. (2001). Traditional objects with sharp boundaries can be defined as Crisp-Crisp Objects (CC-Objects). Rough objects are grouped in the following three classes (figure 3):

- Crisp-Rough Objects (CR-Objects) with well defined boundaries and uncertain content;
- Rough-Crisp Objects (RC-Objects) with precise content and undefined spatial edge;
- Rough-Rough Objects (RR-Objects) with uncertain both contents and boundaries.

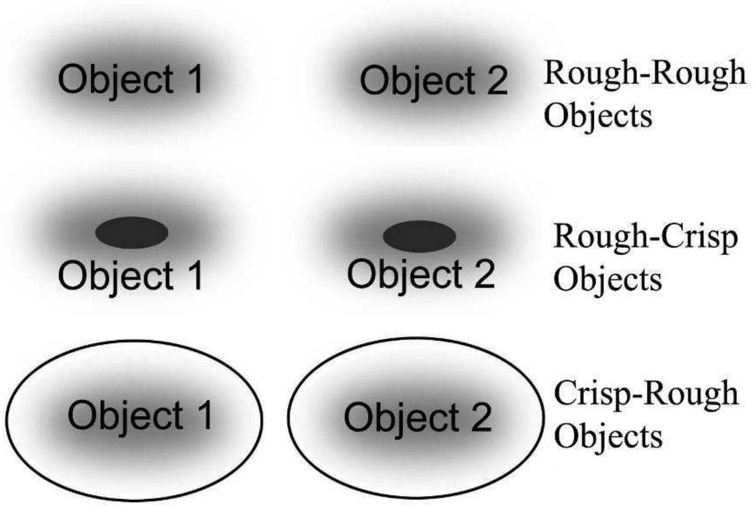

FIG. 3 – *Different classes of rough objects.*

Topological relationships are defined by the well known 9-intersection model (Egenhofer and Herring 1991). Later on Clementini and Di Felice (2001) developed an extension of this model, called "broad boundaries model". A broad boundary consists of inner and outer edges which increase the 8 cases of 9-intersection model to 44 cases. Also, a topological model has been developed (Wang et al., 2004) in rough set theory.

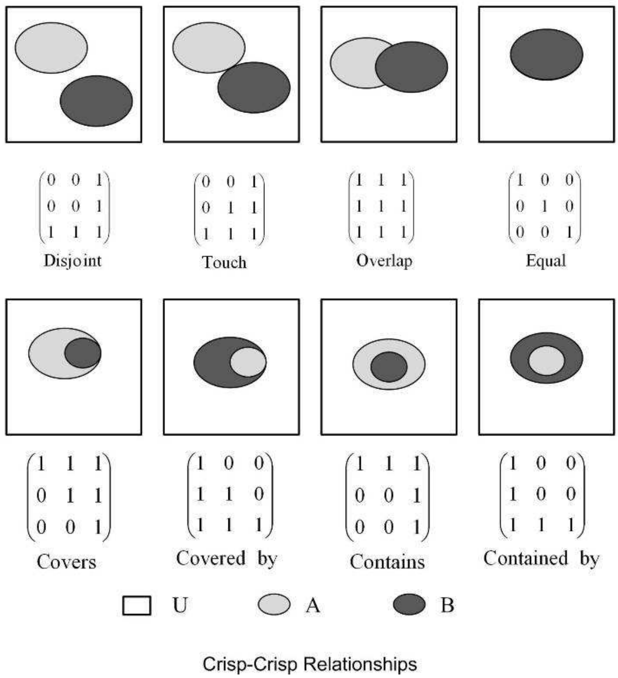

FIG. 4 – *Crisp-Crisp relationship (adapted from Wang et al., 2004).*

$$\begin{pmatrix} Pos(A) \cap Pos(B); & Pos(A) \cap Bnd(B); & Pos(A) \cap Neg(B) \\ Bnd(A) \cap Pos(B); & Bnd(A) \cap Bnd(B); & Bnd(A) \cap Neg(B) \\ Neg(A) \cap Pos(B); & Neg(A) \cap Bnd(B); & Neg(A) \cap Neg(B) \end{pmatrix} \quad \text{EQ. 5}$$

This model (equation 5 and figures 4, 5, 6) is based on the concepts of positive region (Lower approximation), boundary region and negative region (Universe without Upper approximation).

In the same way as for rough object formalization, it is possible to distinguish three kinds of rough relationships:

- Crisp-Crisp (CC), rough relationships between crisp entities (fig. 4);
- Rough-Crisp (RC), rough relationships between rough entities and crisp entities (fig. 5);
- Rough-Rough (RR), rough relationships, between rough entities (fig. 6).

The Crisp-Crisp relationships considering objects with sharp boundaries have the same formalization of the 9-intersection model.

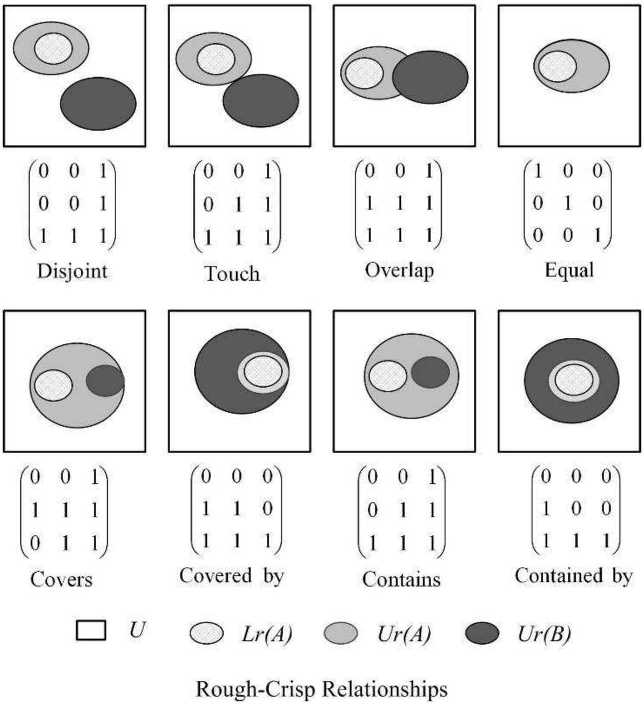

FIG. 5 – *Rough-Crisp relationship (adapted from Wang et al., 2004).*

In Rough-Crisp relationships, A is a rough set with lower Lr(A) and upper Ur(B) approximation, U is the universe and B Ur(B) is a crisp set where Lr coincides with Ur.

In Rough-Rough relationships both A and B are rough sets with lower Lr(A) Lr(B) and upper Ur(A) Ur(B) approximation and U is the universe.

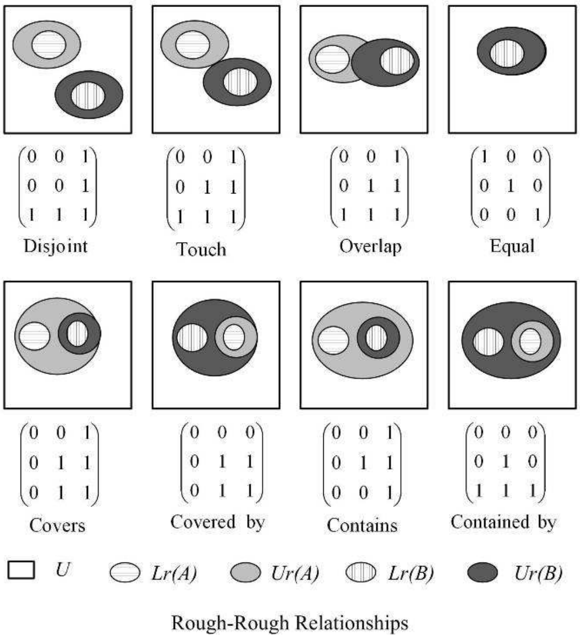

FIG. 6 – *Rough- Rough relationship (adapted from Wang et al., 2004).*

The method proposed by Wang et al. (2004) can be considered as a special case of Clementini and Di Felice's study. In the case of RC and RR relationships the number of possible matrices is not greater than eight as in the 9-intersection model. If spatial data are within the lower approximation they surely belong to the set and cannot be included in other kinds of approximations. If geographical entities are located outside the upper approximation they are not in the set. It is possible to overlap (figure 7) only upper approximations with other upper approximations (Ahlqvist et al., 2000). A lower approximation can be overlapped only with the equivalent upper approximation (Ahlqvist et al., 1998).

As a direct consequence:

$$Lr(A) \cap Lr(B) = 0\,;\; Ur(A) \cap Lr(B) = 0; \qquad \forall_A \forall_{B \neq A}$$

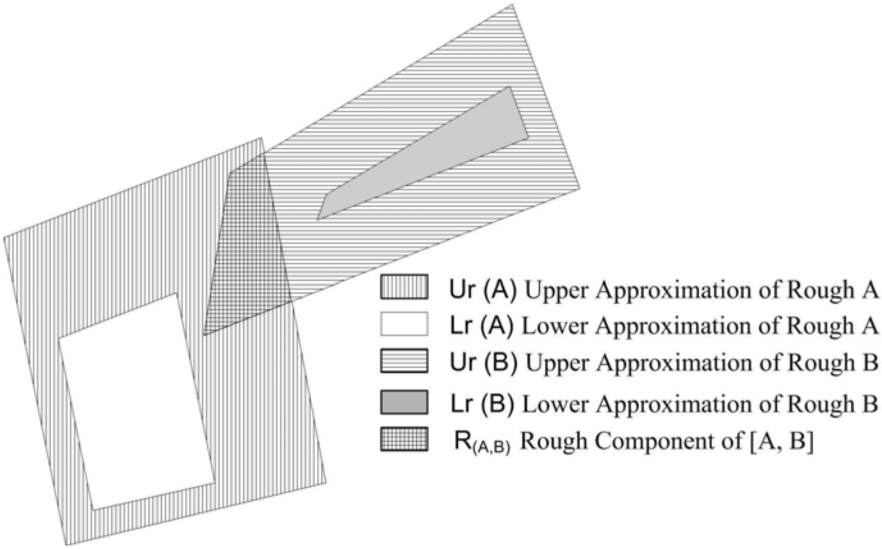

FIG. 7 – *Spatial Rough classification.*

4 Point Pattern Analysis Techniques

Usually, a geographical phenomenon can be analyzed in terms of first and second order properties. First order properties describe geographical events in terms of density or intensity, while second order properties analyze the relationships among spatial phenomena in terms of distance.

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

4.1 Kernel Density Estimation

Kernel density is a technique of point pattern analysis, which considers each point as an event and the associated attributes as the intensity of the phenomenon.

Bailey and Gatrell (1995) developed a set of spatial analysis techniques applied in the field of epidemics spread. These techniques are based on Waldo Tobler's (1970) first law of geography: "Everything is related to everything else, but near things are more related than distant things". Compared to classical statistical approaches, it is important to locate data, considering the events as spatial occurrences of the phenomenon studied. Each Li event is located in the space in an unambiguous way by its coordinates (xi, yi).

An Li event (equation 6) is a function of its position and attributes characterizing it and quantifying its intensity:

$$L_i = (x_i, y_i, A_1, A_2, \ldots, A_n) \qquad \text{EQ. 6}$$

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

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

$$\hat{\lambda}(L) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{L - L_i}{\tau}\right) \qquad \text{EQ. 7}$$

where, $\lambda(L)$ is the intensity of point distribution, quantified at point L; L_i is i^{th} event, k() represents kernel function and τ is the bandwidth. τ can be defined as the radius of a circle generated from the intersection between the surface and the plan containing the study region R. Two major factors influence the results: dimension of grid and bandwidth (Batty *et al.*, 2003). Bandwidth produces a three-dimensional surface, more or less corresponding to the phenomenon, allowing to analyze its distribution at different scales.

Bandwidth choice remarkably influences the surface of estimated density. If bandwidth is high, kernel density is closer to the values of simple density. With a narrow bandwidth, the surface will capture local events, with density close to zero for the elements of the grid located far from each event. The right bandwidth can be determined by estimating the phenomenon and, if it is important, by highlighting peaks of distribution or smooth spatial variations.

4.2 Nearest neighbour distance

Distance analysis among events generally represents an alternative to measures based on density, but in several cases it could be an input datum for KDE. Nearest neighbour distance is the most common distance-based method and it provides information about the interaction among events at the local scale (second order property). Nearest neighbour distance considers nearest neighbour event-event distance, randomly selected. The distance between events can be calculated using Pythagoras theorem:

$$d(L_i, L_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \qquad \text{EQ. 8}$$

If dmin (Li) is the nearest neighbour distance for an Li event, it is possible to consider the mean nearest neighbour distance defined by Clark and Evans (1954) as:

$$\bar{d}_{min} = \frac{\sum_{i=1}^{n} d_{min}(L_i)}{n} \qquad \text{EQ. 9}$$

5 The case study

Potenza Province is located in Southern Italy, it has a low population density (400,000 inhabitants, over 650,000 hectares).

The chief town is Potenza with 70,000 inhabitants, while the other municipalities can be classified in three groups. Twelve towns count more or less 12,000 inhabitants, twenty municipalities have a population of about 5,000 inhabitants, the population of the remaining 67 municipalities varies from 700 to 2,000 inhabitants.

Generally urban sprawl is more common in metropolitan areas, whereas it is not frequent in regions with low population density. This phenomenon in small municipalities is generated by the abandonment of old town centres, while in bigger towns it is produced by high costs of flats. An accurate analysis of the phenomenon has been carried out using all the potentialities of Geographical Information Systems. All the polygons which represent buildings have been converted into points in order to use spatial statistic techniques. The number of flats for each building indicates the intensity of the event in order to achieve a continuous surface of spatial densities. In the application a bandwidth value of 400 m has been used with a grid cell dimension of 10 m.

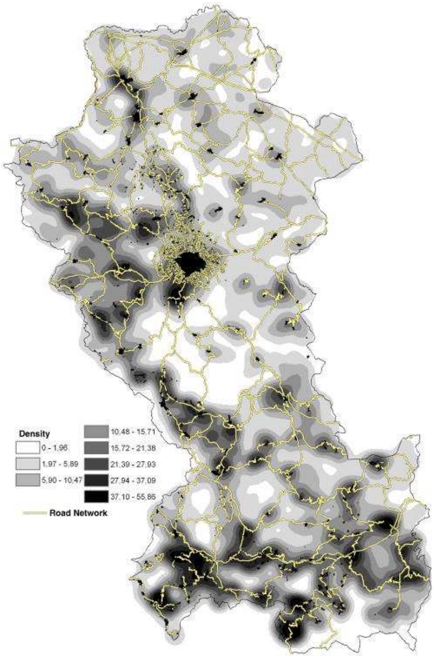

FIG.8 – *Density of scattered settlements compared to road network.*

The phenomenon intensification has been evaluated calculating kernel density from cartographies at 2004 and 1987. This comparison has highlighted more precisely zones with the greatest growth of urban sprawl phenomenon. After the localization of areas where the phenomenon is more considerable, it is important to understand the factors which could lead to its increase.

Kernel density can give further interpretations of the settlement dispersion phenomenon (figure 8). Growth has been developed as a crown surrounding the urban area or along the road network and it is indiscriminately located on landslides and steep slopes.

The greatest increase of urban sprawl has occurred in zones mostly situated on mountains, while urban growth is considered as a threat in areas with intensive agricultural activities.

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

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

The second class generated by Kernel density is exactly one of the inclusion rules adopted in the application. The phenomenon of urban growth has been observed in various municipalities with different sizes and this study indicates that new buildings are completely within a distance of 200 m from the road network. Proximity to road network has been calculated by means of straight line distance, assigning a distance value to each cell. Areas where distance between buildings is less or equal to 100 m. have been obtained by means of nearest neighbour distance. Distances from road and between buildings complete the set of inclusion rules. The following exclusion rules have been considered: area included within a distance of 150 m from rivers and streams, slope higher than 35%, Nature 2000 sites, zones at hydrogeological risk.

These inclusion and exclusion rules have been combined with contiguity rules.

Depth of contiguity zone, for each centre, has been located using a shape index for the boundary of the urban area. 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 index can assume values greater than one. The more the value is greater than one, the more the shape of the settlement will be long, jagged 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 to an index greater than 2.4. In table 1 the 100 municipalities of Potenza Province are grouped in three classes according to compactness rate. This table highlights the low level of compactness of urban areas of Potenza Province.

	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.*

Two criteria have been considered for contiguity: the first one 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.

All these rules with the same thresholds has been adopted in both techniques Map Algebra and Spatial Rough set.

5.1 Land classification with map algebra

Three different periurban areas have been identified by combining the previous rules with map algebra:

- the first edge has been obtained considering all previous (inclusion-exclusion) rules and first contiguity, which takes into account the ratio between area and perimeter of the urban region;
- the second boundary has been achieved considering all previous rules and the second contiguity, which considers area and perimeter of the circle inscribing the urban region;
- the third zone does not take into account any contiguity.

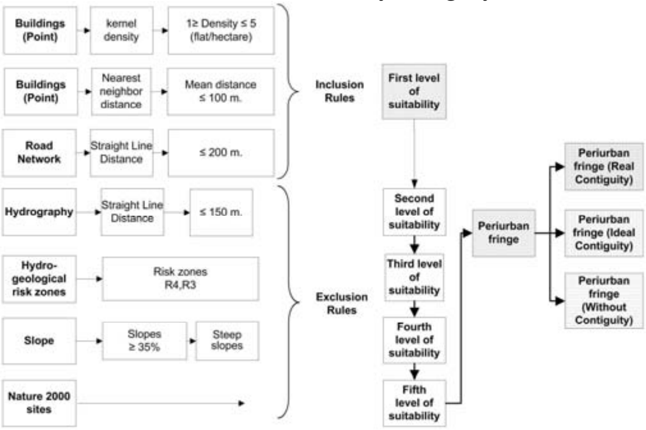

FIG.9 – *Scheme of the procedure for the location of Periurban fringe.*

As expected, results are rather different according to type of periurban area. In most municipalities, the smallest area is achieved considering the first of cases mentioned above. The biggest region is yield without considering any contiguity rule.

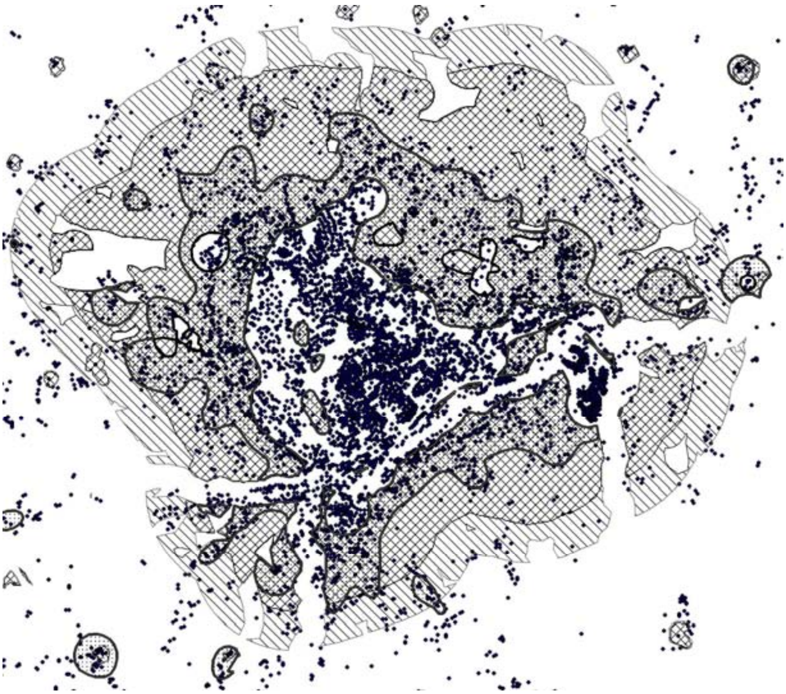

FIG.10 – *Comparison among the three periurban fringes in Potenza municipality. The inner area (dotted hatch with bold boundary) does not consider the contiguity. The intermediate zone (squared hatch) adopts the first contiguity rule, the outer region (hatch with lines) uses the second contiguity rule.*

This trend is completely reversed in the case of Potenza municipality (figure 10). The periurban fringe obtained without taking into account any contiguity rule is the smallest one, because the kernel function captures a low density of buildings in these zones. This result implies that areas close to the urban region are represented by settlements with at most 2 flats for buildings. In confirmation of this hypothesis the relationship between the two periurban fringes and the two contiguity rules hold the same sequence.

However, the greatest part of cases follows the order "first contiguity rule - second contiguity rule -without any contiguity rule".

Considering the morphology of settlement system of Potenza Province, constituted in most of cases by urban areas sited on the spire of Apennine and scattered settlements located near the road network along the valleys, the contiguity rule does not entirely capture the phenomenon leading to a wrong interpretation.

Considering Avigliano municipality (figure 11), the use of contiguity rules implies the exclusion of north-western and south-eastern settlements. From figure 11 it is easy to understand that the transition from the first to the second contiguity rule does not cause a large increase of extension. In this figure a huge increase of periurban fringe is yield without any contiguity rule, which better represents the actual situation.

In all cases, the more reliable region is the third one which is achieved without any contiguity rule, because it identifies the areas in which new transformations are more likely to occur.

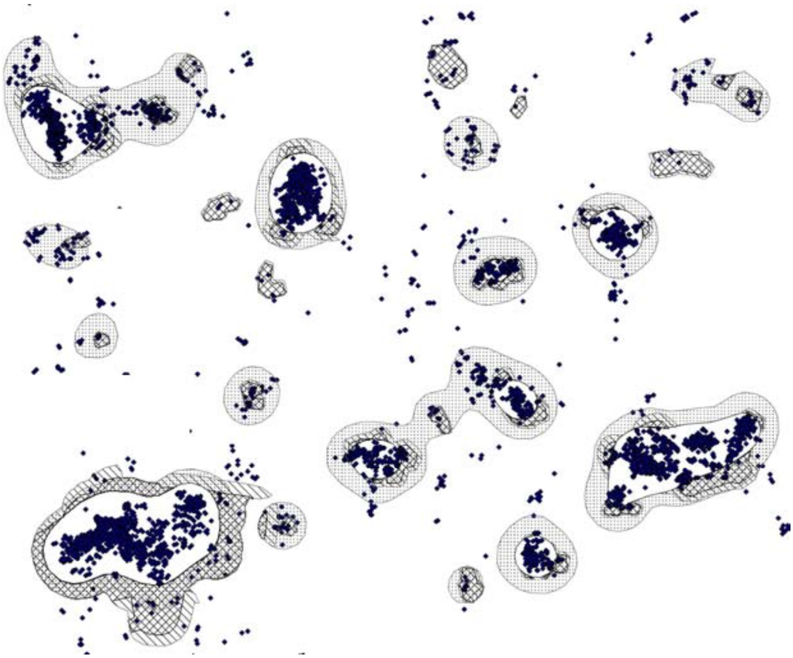

FIG.11 – *Avigliano, middle size municipality, the squared hatch adopts the first contiguity, the hatch with lines uses the second contiguity rule and the dotted hatch does not consider the contiguity.*

5.2 Land classification with Spatial Rough

Rough Set theory has been applied for the classification of different geographic layers. In the case of study, the whole provincial territory represents the universe [U]; inclusion and exclusion rules are the attributes [A] (table 2) and objects (cells of the grid) have been classified following previously defined rules. Indiscernibility relations have been calculated in [U] considering attributes, obtaining the subset [X]. Therefore, all the cells which satisfy at the same time exclusion and inclusion rules are assumed as a subset [X], contained in [U]. The decision system in this case is composed by the set X, i.e. the set of attributes where the decision variable is the nearest neighbour distance.

Attributes and decision variable	Values	Inclusion rules
B= Nature 2000 sites	0 Nature 2000 sites 1 different from Nature 2000 sites	1 different from Nature 2000 sites
Fi= Hydrography	1 if the distance is less than 150 m. 2 if the distance is included between 150 m. and 800 m. 3 if the distance is bigger than 800 m.	2 if the distance is included between 150 m. and 800 m. 3 if the distance is bigger than 800 m.
Fr = Hydro-geological risk zones	0 classes R3 ed R4 1 different from classes R3 ed R4	1 different from classes R3 ed R4
P= Slope	1 if the slope is less than 23,6% 2 if the slope is included between 23,6% and 35% 3 if the slope is bigger than 35%	1 if the slope is less than 23,6% 2 if the slope is included between 23,6% and 35%
D= Density	1 if the density is less than 1 flats per hectare 2 if the density is included between 1 and 5 flats per hectare 3 if the density is bigger than 5 flats per hectare	2 if the density is included between 1 and 5 flats per hectare
V= Road Network	1 if the distance is less than 200 m. 2 if the distance is included between 200 and 700m. 3 if the distance is less than 700 m.	1 if the distance is less than 200 m.
C1= Real Contiguity	0, 1	1
C2= Ideal Contiguity	0, 1	1
Nei= Nearest neighbour distance (Decision Variable)	1 if the minimum distance is less than 100m 2 if the minimum distance is included between 100 and 200m 3 if the minimum distance is bigger than 200m	1 if the minimum distance is less than 100m 2 if the minimum distance is included between 100 and 200m

TAB. 2 – *Attributes and decision variables for rough classification.*

Set X has been obtained according to the decision variables (NEI) in order to achieve its lower and upper approximation. Three classifications have been done. Two ones take into account the first and the second contiguity rules, respectively; the third one does not consider any contiguity rule. Lower and Upper approximation, and accuracy for NEI equal to 1 have been computed for each case.

In the case of Real Contiguity, Lower approximation belongs to the set X and at the same time it is included in first contiguity belt with nei equal to 1. Upper approximation follows the same rules of the Lower approximation with the exception of Nei which can be equal to 1 or to 2 (table 3).

B	Fi	Fr	D	P	V	C_1	Nei 1	Nei 2	Total	Lower	Upper	Accuracy
1	2	1	2	1	1	1	103914		103914	138056	435856	31,67%
1	3	1	2	1	1	1	297750	50	297800			
1	2	1	2	2	1	1	8946		8946			
1	3	1	2	2	1	1	25196		25196			

TAB. 3 – *Decision Table, Lower and Upper approximation and Accuracy with Real Contiguity.*

As previously defined, Accuracy is the cardinality ratio between Lower and Upper approximation. In the case of Real Contiguity, Accuracy is equal to 31.67% (table 3).

B	Fi	Fr	D	P	V	C2	Nei 1	Nei 2	Total	Lower	Upper	Accuracy
1	2	1	2	1	1	1	121089	32	121121	44214	523862	8,44%
1	3	1	2	1	1	1	358388	139	358527			
1	2	1	2	2	1	1	12127		12127			
1	3	1	2	2	1	1	32087		32087			

TAB. 4 – *Decision Table, Lower and Upper approximation and Accuracy with Ideal Contiguity.*

The same procedure has been followed for the case of Ideal Contiguity. In this case Accuracy is equal to 8,44% (table 4).

The last classification has been applied without considering any contiguity rule. In this case Accuracy is equal to 7.84% (table 5).

B	Fi	Fr	D	P	V	Nei 1	Nei 2	Total	Lower	Upper	Accuracy
1	2	1	2	1	1	231461	102	231563	72857	928856	7,84%
1	3	1	2	1	1	624127	309	624436			
1	2	1	2	2	1	20390		20390			
1	3	1	2	2	1	52467		52467			

TAB. 5 – *Decision Table, Lower and Upper approximation and Accuracy without Contiguity.*

Comparing results of the three cases, Real Contiguity rule produces the best results (tables 3, 4, 5) with a better interpretation of the phenomenon (figures 12, 13, 14, 15, 16, 17).

6 Results and final discussion

The spatial extension of Rough Set theory is particularly suitable in defining the periurban belt because it can be included among Rough-Rough Objects (RR-Objects), following the previous rough objects classification, since a clear definition as well as unmistakable rules in defining the edge lacks. The periurban phenomenon is a case with a high level of uncertainty, with ill defined data for the whole area and not much clear rules.

Sorites paradox perfectly describes the periurban phenomenon:

- two or three buildings do not make a town;

- a million buildings do make a big town;
- if n buildings do not make a town neither do (n+1) buildings;
- if n buildings make a town, so do (n-1) buildings.

The first property combined with the third one implies that a million buildings do not make a town, in contradiction with the second property. In the same way, a combination of the second and fourth properties shows that two or three buildings do make a town, in contradiction with the first property (Fisher, 2000).

Comparison among the results achieved using Map Algebra technique and Rough Set method is the most interesting issue (figures 12, 13, 14, 15, 16, 17).

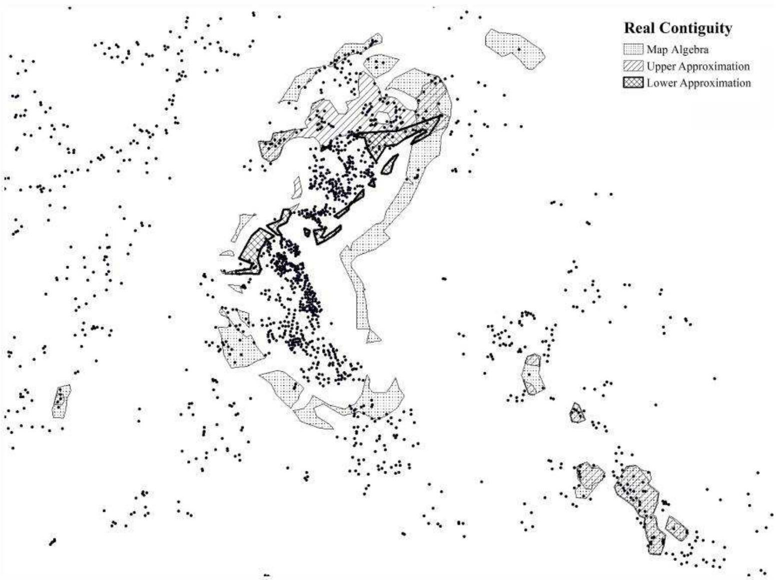

FIG.12 – *Comparison among the periurban fringe obtained with Map Algebra and Rough Set considering real contiguity for Lauria municipality.*

The periurban fringe achieved by means of spatial rough set is the smallest one in all municipalities and with every kind of contiguity. This tendency may be due to several factors:

- how certain combinations of attributes are interrelated;
- map algebra adopts Boolean operators, based on the true-false logic, producing either too narrow or too wide areas;
- decision variable influences results in a certain way, for this reason it is more important to adopt a decision variable which is a sort of key factor for the phenomenon to represent.

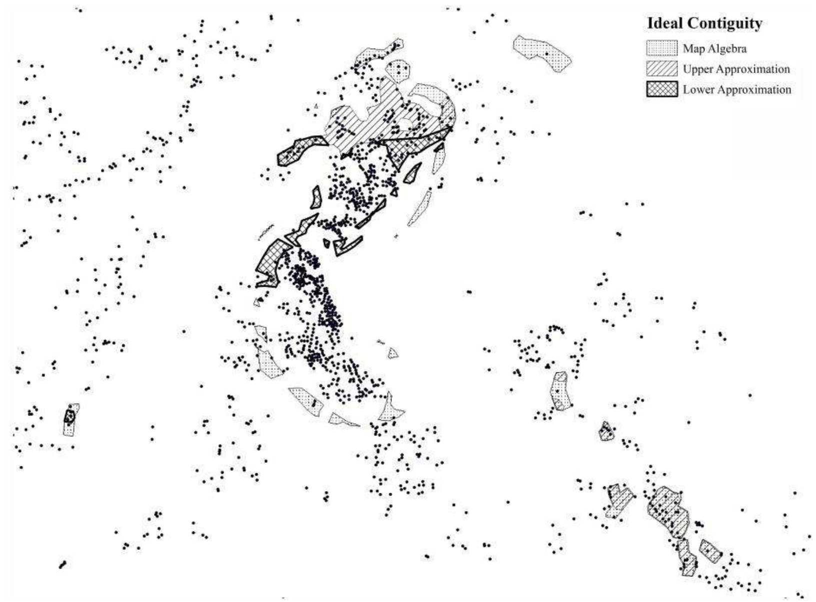

FIG.13 – *Comparison among the periurban fringe obtained with Map Algebra and Rough Set considering ideal contiguity for Lauria municipality.*

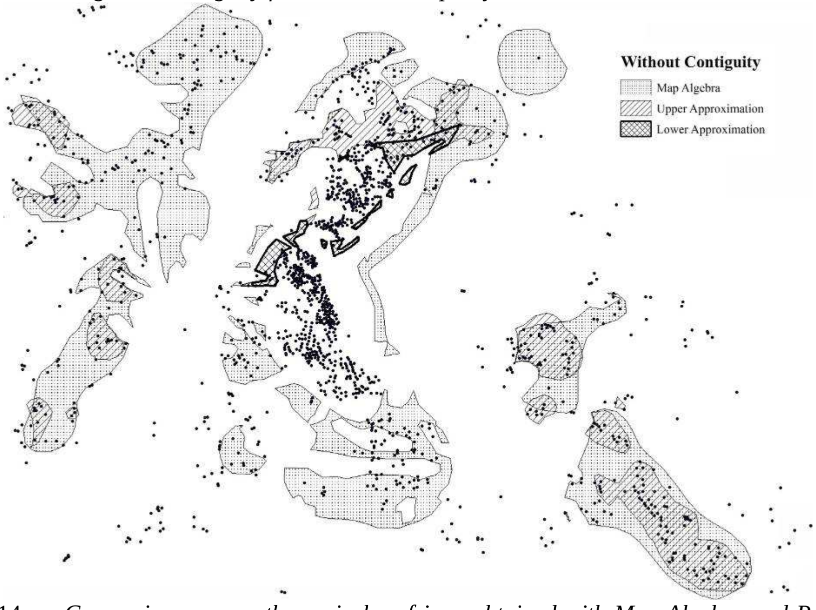

FIG.14 – *Comparison among the periurban fringe obtained with Map Algebra and Rough Set without considering any contiguity rule for Lauria municipality.*

FIG.15 – *Comparison among the periurban fringe obtained with Map and Rough Set considering real contiguity for Picerno municipality.*

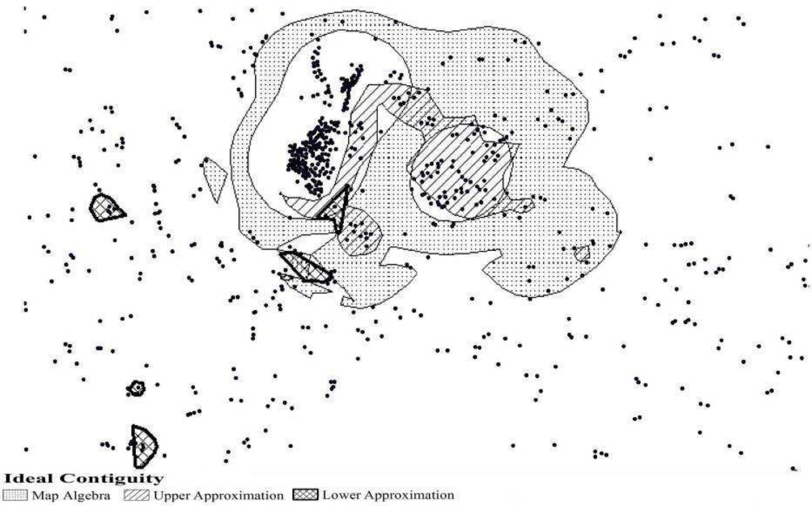

FIG.16 – *Comparison among the periurban fringe obtained with Map Algebra and Rough Set considering ideal contiguity for Picerno municipality.*

FIG.17 – Comparison among the periurban fringe obtained with Map Algebra and Rough Set without considering any contiguity rule for Picerno municipality

The results showed in figures 12 to 17 are confirmed in Table 6, which summarizes results achieved with each classification method, taking into account cell number included in periurban fringe.

	Map Algebra	Rough set		
		Lower	Upper	Accuracy
Without Contiguity	1721768	72857	928856	7,84%
Real Contiguity	766864	138056	435856	31,67%
Ideal Contiguity	1208171	44214	523862	8,44%

TAB. 6 – Comparison among the cell number included in periurban fringe for each classification method.

Despite widely held that Rough Set evaluations are used to achieve more soft constraints, our results show that the classification with Map Algebra produces a periurban fringe with a larger extension, while rough classifications generate a narrower fringe. Among the rough classifications, real contiguity produces an acceptable accuracy and it is more close to reality. The achieved area represents in an indiscernible way the region where the new development areas can be located.

7 Conclusion

Rough set theory reduces cognitive complexity (Gorsevski and Jankowski, 2008) using the indiscernibility relation, where objects characterized by identical attribute values occur in the information system. The universe can be partitioned into blocks of indiscernible objects, called elementary sets. The combination of these blocks defines subsets of the universe called lower and upper approximations. Knowledge about a real or abstract world can be built combining elementary sets, following certain rules (Greco et al., 2001).

Approximations of sets are basic operations in rough set theory and are used as main tools to deal with vague and uncertain data (Pawlak, 1997). In most cases data are achieved adopting spatial statistic techniques, where a certain level of uncertainty is due to several factors. For instance, Kernel Density, adopted in the present application, is a function of the choice of some key parameters, such as grid resolution, kernel function and, above all, bandwidth.

A deeper understanding of spatial and non-spatial relationships among dependent and independent geographical variables is a fundamental issue for a correct development of Rough Set approach to land classification. Precisely, using these variables plays an important role in land classification by spatial rough set. Choosing attribute combination and decision variable on one side affects results, whist on the other side produces simulations closer to reality. Rough Set approach allows a better interpretation of the periurban phenomenon. This method does not produce a sharp boundary which unambiguously defines the periurban edge, but it allows to locate cells with the same attribute values taken into account for the classification in an indiscernible way. This area, called boundary, exactly represents the transition zone from urban to rural part of the town, where urban sprawl phenomenon could be more concentrated.

References

Ahlqvist O., Keukelaar J., Oukbir K., (1998). *Using Rough Classification to Represent Uncertainty in Spatial Data*. SIRC98, Towards the Next Decade of Spatial Information Research. 10th Annual Colloquium of the: Spatial Information Research Centre, University of Otago, Dunedin, New Zealand.

Ahlqvist O., Keukelaar J., Oukbir K., (2000). *Rough classification and accuracy assessment*. International Journal of Geographical Information Science, vol. 14, no. 5, 475 – 496. London: Taylor and Francis.

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.

Beaubouef, T., Petry, F. and Ladner, R., (2007). *Spatial Data Methods And Vague Regions: A Rough Set Approach*. Applied Soft Computing Journal, vol. 7, pp. 425-440 Elsevier Science Publishers.

Burrough P. A., Frank A. U., (1996). *Geographic objects with indeterminate boundaries*, London: Taylor & Francis.

Camagni R., Gibelli M.C., Rigamonti P., (2002). *I costi collettivi della città dispersa*. Firenze: Alinea.

Cheng T., Molenaar M., Lin H., (2001). *Formalizing fuzzy object uncertain classification results*, International Journal of Geographical Science vol. 15, no 1, 27-42, London: Taylor and Francis.

Clark P.J., Evans F.C., (1954). *Distance to nearest neighbour as a measure of spatial relationships in populations*, Ecology, 35:445–453, 1954.

Clementini E., Di Felice P., (2001). *A spatial model for complex objects with a broad boundary supporting queries on uncertain data*. Data & Knowledge Engineering, 37(3):285-305.

Cohn A.G., Gotts N.M. (1996). *The 'Egg-Yolk' representation of regions with indeterminate boundaries"* in Burrough P. A., Frank A. U., (eds.). Geographic objects with indeterminate boundaries, London: Taylor & Francis.

Eastman J.R., (1999). *Multicriteria evaluation and GIS*, in Longley P.A., Goodchild M.F., Maguire D.J. Rhind D.W. Geographic information system and science. Jonh Wiley & sons.

Egenhofer M. J., Herring, J., (1991). *Categorizing binary topological relationships between regions, lines, and points in geographic databases*, Technical Report, Department of Surveying Engineering, University of Maine, Orono.

Erwig, M., Schneider M., (1997). *Vague Regions*. Proceedings of 5th International Symposium on Advances in Spatial Databases (SSD'97), LNCS 1262 pp.298-320

Fisher P., (2000). *Sorites paradox and vague geographies,* Fuzzy Sets and Systems 113, 7-18, Elsevier Science.

Hazarika S.M., Cohn A.G., (2001), *A Taxonomy for Spatial Vagueness, an alternative egg-yolk*, Proceedings of COSIT 2001, LNCS 2205 pp.92-107

Gatrell A. C. , Bailey T. C., Diggle P. J., Rowlingson B. S., (1996). *Spatial point pattern analysis and its application in geographical epidemiology*, Transaction of institute of British Geographer.

Gorsevski P. V., Jankowski P., (2008), *Discerning landslide susceptibility using rough sets*. Computers, Environment and Urban Systems 32, 53–65. Elsevier Science Publishers.

Greco S., Matarazzo B., Slowinski R., (2001). *Rough sets theory for multicriteria decision analysis.* European Journal of Operational Research, 129, 1-47. Elsevier Science Publishers.

Indovina F. (1990). *La città diffusa*. Venezia: Daest.

Leung Y.. (1988). *Spatial analysis and planning under imprecision*. Amsterdam, The Netherlands: Elsevier Science Publishers.

Malczewski J. (1999). *Gis and Multicriteria Decision Analysis*. John Willey & Sons Inc.

Murgante B., Las Casas G., Danese M., (2008). *The periurban city: geostatistical methods for its definition* in Rumor M., Coors V., Fendel E. M., Zlatanova S. (Eds), Urban and Regional Data Management. London: Taylor and Francis.

Murgante B., Las Casas G.. (2004). *G.I.S. and Fuzzy Sets for the Land Suitability Analysis,* Lecture Note in Computer Scienze. Vol. 3044. Berlin: Springer Verlag.

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

Pawlak Z., (1997). *Rough set approach to knowledge-based decision support*, European Journal of Operational Research, Volume 99, Number 1, pp. 48-57, Elsevier Science Publishers.

Pawlak Z., (1998). *Rough Sets theory and its applications to data analysis*, Cybernetics and Systems An International Journal, 29, 661- 688. London: Taylor and Francis.

Thill J.C., (1999). *Spatial Multicriteria Decision Making and analysis*, Ashgate, Aldershot.

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

Wang S., Yuan H., Chen G. , Li D., Shi W., (2004). *Rough Spatial Interpretation.* In Tsumoto S., Słowiński R., Komorowski J., Grzymala-Busse J. W. (eds), “Rough Sets and Current Trends in Computing”. Lecture Notes in Computer Science vol. 3066. Berlin: Springer Verlag.

Worboys, M.F. and Duckham, M., (2004). *GIS: A Computing Perspective*. CRC Press.

Zadeh, L. (1965). *Fuzzy sets*. Information Control. 8, 338-353.

Résumé

De nos jours la disponibilité des informations spatiales ne cesse d’augmenter avec l’amélioration des techniques et des méthodes adoptées dans les analyses géographiques. Malgré cette tendance, classifier précisément chaque partie d’une ville est une tâche de plus en plus difficile, en raison de la complexité croissante du tissu urbain. La théorie des ensembles approximatifs peut constituer une approche intéressante pour regrouper un grand volume de données dans le but d’élaborer une connaissance complexe d’une zone géographique. Elle constitue une approche nouvelle du principe d’incertitude, en capturant l’indiscernabilité.
Deux phénomènes différents peuvent être indiscernables dans des contextes particuliers, mais classifiés de la même manière quand on combine les informations les concernant. Il existe plusieurs expériences traitant de l’emploi de la théorie des ensembles approximatifs dans les phases de fouille de données, d’analyse des connaissances et de la classification approximatives des modèles. Néanmoins, tous ces domaines de recherche ne prennent pas en considération la composante spatiale.

Le but de cet article est l'emploi des méthodes fondées sur la théorie des ensembles approximatifs dans les analyses géographiques. Cette approche a été appliquée à un cas concret, en comparant les résultats obtenus en adoptant les techniques de l'Algèbre des Cartes avec ceux obtenus en adoptant la théorie des ensembles approximatifs. La zone d'étude correspond à la province de Potenza en Italie, et est particulièrement indiquée pour l'application de cette théorie, car elle comprend 100 municipalités avec des populations et des morphologies très différentes.