GEOGRAPHIC INFORMATION SYSTEMS AND DECISION PROCESSES FOR URBAN PLANNING: A CASE STUDY OF ROUGH SET ANALYSIS ON THE RESIDENTIAL AREAS OF THE CITY OF CAGLIARI, ITALY

Sabrina LAI, Corrado ZOPPI

Dipartimento di Ingegneria del Territorio, Sezione di Urbanistica, Università di Cagliari, Piazza d'Armi 16, 09123 Cagliari, Italy, tel.: Italy–(0)70–6755216, telefax: Italy–(0)70–6755215, e-mail: sabrinalai@unica.it, zoppi@unica.it

ABSTRACT

In Italy, urban planning is based on the city Masterplan. This plan identifies the future urban organization and a system of zoning rules. Land-use policies are based on these rules. The zoning rules should synthesize environmental and spatial knowledge and policy decisions concerning the possible futures, with reference to the different urban functions.

In this essay, a procedure of analysis of the city Masterplan of Cagliari, the regional capital city of Sardinia (Italy), is discussed and applied. This procedure is referred to the residential areas. The procedure tries to explain the urban organization of the housing areas by using a system of variables based on the integration of different branches of knowledge concerning the urban environment.

The decisions on the urban futures that the zoning rules entail are critically analyzed in terms of consistency with this knowledge system.

The procedure consists of two phases. In the first phase, the urban environment is analyzed and described. This is done by defining and developing a geographic information system. This system utilizes a spatial analysis approach to figure out the integration of the residential areas into the urban fabric.

The second phase is inferential. Based on the geographic information system developed in the first phase, a Rough Set Analysis (RSA) method is applied. This technique allows to recognize the connection patterns between the urban knowledge system and the city planning decisions.

The patterns, the decision rules which come from the RSA implementation, are important starting points for further investigation on the development of decision models concerning urban planning.

1. INTRODUCTION[1]

This paper analyzes and discusses the relationship between the geography of the residential areas of the city of Cagliari (Italy) and a system of variables based on the integration of different branches of knowledge concerning the urban environment using a *Rough Set Analysis* (RSA) method. The city of Cagliari, located on the south-eastern part of Sardinia, is the regional capital city and the most important urban area of the island of Sardinia. The paper generates a set of rules, based on the RSA, concerning correlations between subsets of variables and areas identified as residential by the city Masterplan. The rules give the private and public agents a picture of how and why the geography of housing areas has developed, and provide sound basis for discussing, recognizing and addressing the mutual interests and expectations for planning policies for these areas.

The spatial configuration of the urban residential areas is considered as dependent on: the distribution of the resident population across the city, the zoning rules concerning the characteristics of the residential areas, the urban land-use structure, and the layout of the zones on which building is forbidden (areas around the city cemeteries; wetlands; historic and panoramic sites; coastal areas; natural reserves).

As such, this issue is a complex problem which entails several variables and no prior on how these variables are connected to each other. The reason why this paper uses RSA is that it is specifically designed to address this kind of problems.

Indeed, the most relevant methodological point of this paper is to demonstrate how the RSA procedure can be used to address an important problem of spatial analysis.

There are several types of RSA procedures and a multiplicity of fields of application. For a thorough discussion of each method and their virtues and deficiencies in the fields of: pharmacology, banking technique and practice, mechanics, linguistics, environmental science, materials science, graphic recognition and genetics, see Pawlak (2001); gastroenterologic diagnosis, see Carlin et al. (1998); radiology, see Gardner et al. (1996); aircraft design, see Peña et al. (1999); automobile headlights, see Lee e Vachtsevanos (2002).

The doctoral thesis of Aleksander Øhrn (1999) is an important reference point for the RSA. It contains a thorough and comprehensive discussion on the main theoretical issues, starting from the pioneer essays of Pawlak (1982; 1991), and on applications, particularly in the medical field. Moreover, Øhrn (1999; 2001) implements a interdisciplinary project named "Rosetta" for a software program designed to develop cooperation between theoreticians and practitioners who utilize rough sets and methods based on discernibility-based data analysis. This software program is freely available on the Internet and continuously updated through the voluntary contribution of the participants to the project.

RSA is not a favorite tool among urban and regional planners at least for two reasons. First, computer programs implementing RSA applied to urban and regional problems are not currently available. Second, a consolidated and structured literature does not exist on this issue as much as it does for other methodologies derived from econometrics, applied statistics and other techniques of knowledge discovery in databases (KDD). Moreover, while the RSA methodologies build knowledge without any prior, traditional evaluation methods of regional and urban planning, e.g. Cost-Benefit and Multicriteria Analysis, have some built-in procedure on how to evaluate and draw conclusions.

The RSA methodologies make it possible to build a step-by-step explanation of the inferential process. This is an important difference with respect to other KDD techniques commonly used in regional and urban planning, such as neural networks, fuzzy logic and cellular auto-

[1] This paper is part of the project of relevant national interest on "Sustainable Development and E-Governance in Urban Planning," funded by the Italian Ministry of Education, University and Research for the period 2006-2008.

mata. Since the inferential process is transparent, it is relatively easy for public officials, politicians, practitioners, scientists, citizens, entrepreneurs and others, to understand the reasons why decisions are taken. Thus, public discussion and negotiation could take place on a sound informational basis.

The RSA methodologies applied to regional and urban planning problems imply the use of spatial data, and inference on spatial phenomena. As a consequence, the results of the inferential process can be represented through geographic information systems. Moreover, the results from the RSA can be integrated into other evaluation methodologies, such as Multicriteria Analysis and Contingent Valuation.

Bruinsma's et al.'s (2002) paper is among the earliest and most important ones in this area. It discusses the location problem of industrial activities in some urban regions of Belgium, Germany, France and Denmark. Cerreta and Salzano (2004) develop a RSA-based cognitive approach to evaluate alternative planning policies for conservation of urban cultural heritage. They integrate RSA into Institutional Analysis and Community Impact Evaluation. Murgante and Sansone (2005) use RSA to identify the boundaries of the urban and suburban areas of the region of Basilicata (Southern Italy).

This paper is organized as follows. In the second section, the RSA methodology is discussed within the context of the case study. The third section discusses the geographic information system which describes the urban residential areas of Cagliari and the spatial variables connected to their configuration. The fourth section presents the analysis of the configuration of the urban residential areas of Cagliari based on the RSA procedure. Finally, the fifth section summarizes the findings and discusses the implications of the RSA methodology to spatial analysis as a way of dealing with complex spatial planning issues.

2. METHODOLOGY

RSA makes it possible to implement effective knowledge-building processes concerning problems which entail several variables and no prior on how these variables are connected to each other. We next describe the RSA methodology used in the paper[2].

Let OBJ be a set of objects and let A be a set of attributes of these objects. The RSA provides the *rough sets* (RS) of these objects. The RS are subsets of OBJ characterized by a indiscernibility relation, that is internal homogeneity and heterogeneity with respect to the other objects of OBJ.

Let OBJ = $\{O_1, O_2, \ldots, O_M\}$. A RS of OBJ is identified by its upper (UA) and lower approximation (LA). The UA is the union of the equivalence classes of the elements of OBJ which have a non-empty intersection with RS. The LA is the union of the equivalence classes of the elements of OBJ which are contained in RS. Moreover: the boundary region (BR) of a RS is the set of the elements of UA which are not contained in LA; the negative region of a RS is the set of elements of OBJ which are contained in equivalence classes whose elements are not contained in UA.

Let A = $\{A_1, A_2, \ldots, A_N\}$ be a set of attributes of the elements of OBJ; equivalence classes of OBJ can be obtained by defining an equivalence relation as follows.

Let $R(A_j)$ be an equivalence relation, where A_j is the j-th element of A. Given two elements of OBJ, o_1 and o_2, we say that o_1 is equivalent to o_2 if they have the same value of the attribute A_j. The symbolic notation is the following:

$$o_1 \; R(A_j) \; o_2. \qquad (1)$$

[2] The RSA methodology used in the paper is based on the discussion of Pawlak and Slowinski (1994). This discussion synthesizes two previous papers of Pawlak (1982; 1991). A couple of presentations given by Yang et al. (2002) in a course on data mining taught by Anita Wasilewska at Stony Brook State University of New York have also been very useful.

An equivalence class is a subset of elements of OBJ for which property (1) holds.
The attributes of A are either characteristics of the elements of OBJ (condition attributes) or a classification order of the elements of OBJ (decision attribute). For example, let OBJ be the set of the days of year 2003 and let A = {temperature (°C), pressure (bar), air humidity (x.xx %), rainfalls (at least 10 mm or less than 10 mm)}. The first three attributes of A are "condition attributes," the fourth is a "decision attribute," since it expresses a classification of the elements of OBJ based on a difference (with respect to the rainfall heaviness).
The RS can be identified by the so-called "reducts" of the elements of OBJ, defined through the attributes of A.
The "reduct relative to the k-th element of OBJ," RED_K, is the minimal combination of condition attributes of A which discerns the k-th element from the other elements of OBJ. The "decision-relative reduct relative to the k-th element of OBJ," $DRED_K$, is the minimal combination of condition attributes of A which discerns the k-th element from the other elements of OBJ, under the assumption that a RS is identified by all the elements of OBJ which have the same value of the decision attribute. In other words, discernibility between the elements of OBJ is based on the value of the decision attribute.
Once the $DRED_K$'s are identified, "decision rules" can be defined which connect the condition attributes and their values to the decision attribute and its values.
The symbolic notation for a decision rule is the following:

$$DRED_K(B) \Rightarrow A_D, \qquad (2)$$

or

$$DRED_K(B) \Rightarrow A_P \text{ OR } A_Q \text{ OR } A_R \text{ OR}\ldots \qquad (2')$$

This notation reads as follows:

- left-hand side: "the decision-relative reduct relative to the k-th element of OBJ, for the values of the attributes of vector B (these attributes are a subset of the condition attributes of A);"
- central (arrow): "implies;"
- right-hand side: (2) "a value of the decision attribute equal to A_D," or (2') "a value of the decision attribute equal to A_P, or a value of the decision attribute equal to A_Q, or a value of the decision attribute equal to A_R, or a value of the decision attribute equal to … ."

A_D, A_P, A_Q, A_R are values of the decision attribute. In case (2) the decision rule is *exact*, in case (2') the decision rule is *approximate*.
The key to identify the $DRED_K$'s is determining the "decision-relative discernibility matrix," DRDM. The following is an example of how to determine a DRDM.
Let OBJ = {O_1, O_2, O_3, O_4, O_5, O_6} and let A = {A_1, A_2, A_3, A_4}, where: the elements of OBJ are six applicants for a position of assistant professor of urban planning at an Italian university; A_1 is a condition attribute relative to articles published in international journals with refereeing process (the values are: "A" for more than six essays; "M" if the articles are between six and three; "L" if the articles are less than three); A_2 is a condition attribute relative to the university courses with a number of credits equal to five or more taught by the applicant (the values are: "H" for more than ten courses; "I" if the courses are between ten and seven; "J" if the courses are between six and three; "K" if the courses are less than three); A_3 is a condition attribute relative to the knowledge of the English language (the values are: "S" for very good knowledge; "G" for intermediate level; "L" for pre-intermediate level); A_4 is the dichotomous decision attribute (the values are: "W" if the applicant takes the position; "F" if the applicant fails).
This information is summarized in Table 1. The matrix of Table 1 is named either "information matrix" or "object-attribute matrix."

	A_1	A_2	A_3	A_4
O_1	A	I	S	W
O_2	A	H	L	F
O_3	L	J	G	F
O_4	M	J	G	F
O_5	L	K	L	F
O_6	M	H	S	W

Table 1. An example of an information matrix

The DRDM derives from the information matrix. The DRDM is the symmetric matrix of the binary comparisons of the elements of OBJ. The result of a binary comparison of two elements of OBJ, O_i and O_j, is expressed by the condition attributes whose values are different for O_i and O_j, given O_i and O_j have different values of the decision attribute. If O_i and O_j have the same value of the decision attribute, they are considered indiscernible and the result of their binary comparison is the empty set. If the result of the binary comparison of O_i and O_j is the empty set, the two elements belong to the same RS.
The DRDM is shown in Table 2. The $DRED_1$, for $A_4 = F$, can be calculated as follows (the symbolic notation refers to the classical set theory):
$DRED_1=\{\{A_2\}U\{A_3\}\}\cap\{\{A_1\}U\{A_2\}U\{A_3\}\}\cap\{\{A_1\}U\{A_2\}U\{A_3\}\}\cap\{\{A_1\}U\{A_2\}U\{A_3\}\}=\{A_2,A_3\}$.
This result has the following meaning: "The element O_1 of the RS $\{O_1, O_6\}$ has a reduct relative to the decision attribute A_4 equal to the set {I, S}. In other words, the element O_1 is discernible by the values of the condition attributes A_2 and A_3."
The definition of $DRED_1$ implies the following rule (see notation 2):

$$DRED_1(A_2=I, A_3=S) \Rightarrow A_D=W.$$

With the same procedure, the following rule can also be identified:

$$DRED_6(A_2=H, A_3=S) \Rightarrow A_D=W.$$

	O_1	O_2	O_3	O_4	O_5	O_6
O_1	-	A_2,A_3	A_1,A_2,A_3	A_1,A_2,A_3	A_1,A_2,A_3	-
O_2	A_2,A_3	-	-	-	-	A_1,A_3
O_3	A_1,A_2,A_3	-	-	-	-	A_1,A_2,A_3
O_4	A_1,A_2,A_3	-	-	-	-	A_2,A_3
O_5	A_1,A_2,A_3	-	-	-	-	A_1,A_2,A_3
O_6	-	A_1,A_3	A_1,A_2,A_3	A_2,A_3	A_1,A_2,A_3	-

Table 2. An example of a decision-relative discernibility matrix

The descriptive statistics of the decision rules are the following[3]:

- the left-hand side support (LHSS) of a decision rule "$DRED_K(B) \Rightarrow A_D$" is the number of elements of OBJ for which condition $DRED_K(B)$ holds;

[3] For a thorough discussion of these statistics and of their use for filtering the decision rules see Øhrn (1999), Sections 6.2 and 6.3.

- the right-hand side support (RHSS) of a decision rule is the number of elements of OBJ for which the decision rule holds; in case the rule is expressed by the symbolic notation (2'), the right-hand side support is a n-tuple of values;
- the right-hand side accuracy (RHSA) of a decision rule is equal to the right-hand side support divided by the left-hand side support;
- the left-hand side coverage (LHSC) of a decision rule is the fraction of elements of OBJ for which the rule holds;
- the right-hand side coverage (RHSC) of a decision rule is equal to: (i) the number of elements of OBJ for which the rule holds divided by the number of elements of OBJ for which the decision attribute has the value "A_D," in case of an exact decision rule; (ii) the number of elements of OBJ for which the rule holds divided by the number of elements of OBJ for which the decision attribute has the values "A_P OR A_Q OR A_R OR…," in case of an approximate decision rule; in case of an approximate decision rule, the right-hand side coverage is a n-tuple of values.

The RHSA is based on the LHSS and the RHSS. This statistic informs on how much a rule is reliable. In other words, it indicates if and how much we can be confident that the rule holds, provided that the reduct $DRED_K(B)$ holds.

The LHSC is based on the RHSS. It makes it possible to put in evidence and compare the different shares of the phenomenon represented by the decision attribute that each rule contributes to explain.

The RHSC is based on the RHSS. It informs on how much a particular state of the phenomenon, represented by a value of the decision attribute, is explained by a decision rule.

In the RSA methodology, the attributes define the characteristics of a set of objects. These characteristics refer to states, behaviors, preferences, expectations, needs. The identification of the relations between the attributes makes it possible to recognize what makes the objects different from each other with reference to the decision attribute, that is which characteristics are decisive in decision-making processes.

3. GEOGRAPHIC INFORMATION SYSTEM

In order to achieve the proposed aim, geographic information systems are, by far, the most powerful and user-friendly tool which allows us not only to depict the present characteristics of a particular area, but also to analyze and integrate past knowledge and to develop new knowledge.

The chapter is divided into two parts. In the first one, the construction of the geographic information system is described; the second part presents the outcomes, by means of a comprehensive table and some detailed pictures.

3.1. Methodology

3.1.1. Data

Our analysis aims at ascertaining whether the zoning rules of the city Masterplan of Cagliari are somehow linked with a system of environmental, demographic and legal variables. In order to look into such a connection, we need to select properly the type of information to be utilized.

Choosing which data are to be used is a crucial issue, since the choice may significantly affect the outcomes. On the other hand, it should be pointed out that a mere introduction of all the available data, without incorporating any prior knowledge which takes into account "likely relationships [...] and [...] patterns already known" (Fayyad et al., 1996), implies making a real effort in pinpointing reasonable patterns, because of the number of the derived rules (Curry, 2003).

On this basis, the following variables are selected:

- resident population in 2001 (ISTAT, 2001);
- land use, classified according to the European project "Corine Land Cover" (fourth level; Regional Administration of Sardinia, 2003; Cilloccu and Cumer, 2002);
- areas where building is forbidden, due to environmental or binding reasons (City of Cagliari, 2004a; City of Cagliari, 2004b; Regional Administration of Sardinia, 2003); these are:

- parks and protected natural areas (according to the Regional Law n. 31/1989);
- Sites of Community Interest (proposed according to the Council Directive n. 92/43/EEC on the conservation of natural habitats and of wild fauna and flora);
- protection zones (according to the Council Directive n. 79/409/EEC on the conservation of wild birds);
- buffer zones around coastal line (300 meters) and around rivers, canals and streams (150 meters) (according to the National Law enacted by Decree n. 42/2004, article 142);
- landscape-protection areas (according to the National Law enacted by Decree n. 42/2004, articles 136 and 157);
- buildings and areas protected because of cultural or historic reasons (according to the National Law enacted by Decree n. 42/2004, articles 10 and 128);
- the area of the San Michele Landscape Plan (City of Cagliari, 2004a);
- buffer zones around the cemeteries (City of Cagliari, 2004a).

3.1.2. Residential zones

Next, we define the area to be examined and the boundaries of the housing zones. According to the city Masterplan of Cagliari, the latter are classed into four categories:

- historic center zone ("A" zone);
- residential completion zone ("B" zone);
- residential expansion zone ("C" zone);
- enterprise zone ("EZ" zone).

We choose to limit our analysis to B, C, and EZ zones. The historic center zone is a single, dense and central area in the urban fabric, which dates from the Middle Ages and hosts buildings important for cultural, artistic and historic reasons. Specific rules apply to this area, in order to avoid an increase in built volume, preserve the facades and control the building uses. The peculiarity of the historic center zone is that it is not a residential zone. Rather, it is a mixed-use zone, which entails public services, commercial and residential uses. For this reason, we do not consider this zone as part of the spatial configuration of the residential areas, even though we consider it as an important spatial reference to explain this configuration.

The "B" zones are built-up areas which consist mainly of dense residential blocks. A partially-built area is generally considered to belong to a "B" zone when its area is smaller than 5000 square meters and more than a 30 percent of the volume has already been built.

As a general rule, in a B zone building is limited, on a single building lot, to 3 cubic meters per square meter. However, there are several different types of B zones, which may be classed into three groups. The most important one consists of nine subtypes; for each of them, the Table 3 synthesizes the main overall limits on: ratio of maximum volume to area of the lot (MVL); minimum distance between house and side of the street (mDHS); minimum distance between house and boundaries of the property not coinciding with streets (mDHB); minimum distance between two houses (mDHH); maximum height (MZ); maximum built area (MBA); minimum building lot area (mBLA).

Type	MVL (cubic meters per square meter)	mDHS (meters)	mDHB (meters)	mDHH (meters)	MZ (meters)	MBA (square meters)	mBLA (square meters)
B1	5	0	5	10	22		
B2	5	0	5	10	16		
B3	5	0	5	10	13		
B4	3	0	5	10	22		
B5	5	5	6,5	10	17	360	600
B6	3	4	6	10	10,5	360	400
B7	1	5	5	10	7,5	360	600
B8	3	0	5		10,5		
B9	0,375		8		7,5		

Table 3. Building limits on some different B zones

The other two groups include: (i) areas to which special rules apply, for instance in order to maintain architectural or urban features, align the facades, regenerate districts in crisis, and, (ii) areas designed to host a combination of urban facilities (60 percent) and private buildings (40 percent).

The "C" zones are still-not-built or partially-built parts of the city (when less than a 30 percent of the volume is already built). These zones are bound to be residential areas. Restrictions on built volume are far stricter than those imposed for the B zones and equal to 1,5 cubic meters per square meter per building lot. Furthermore, in order to obtain a building permission, a plan must be approved by the local municipality. This plan must indicate the spatial distribution of the building lots, as well as a portion of the area (depending on the estimate of the number of the future residents; this is estimated on the basis of the amount of the housing volume, therefore on the ratio of maximum volume to the area of the lot). This has to be handed over to the municipality, in order to build public services and infrastructure.

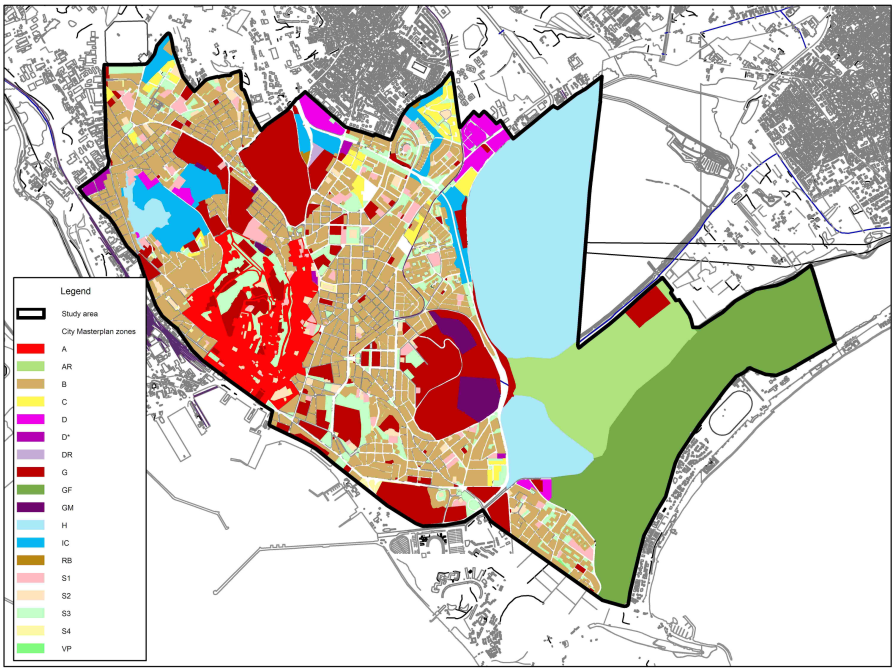

Figure 1. The zones of the Masterplan of Cagliari within the study area

The "EZ" zones are still-not-built or partially-built parts of the city, where an integration of different functions (residential buildings, public facilities and recreational areas) is required. For each EZ zone, the city Masterplan sets specific rules on the combination of functions. For instance, in an EZ zone important for environmental reasons a maximum of 35 percent of the area is available for housing areas, and a 0 percent for public facilities; a 65 percent has to be reserved for recreational areas. A stronger residential EZ is characterized by a 93 percent - 7 percent - 0 percent. An EZ zone located in spoiled city outskirts is characterized by a 70 percent - 30 percent - 0 percent.
In order to analyze the relationships between the different types of housing areas, an aggregation of residential blocks is required. In doing so, we assume that two homogeneous areas can be aggregated provided that they are parted from each other only by secondary roads or by areas specifically designed for local recreational, cultural, social or sports activities.

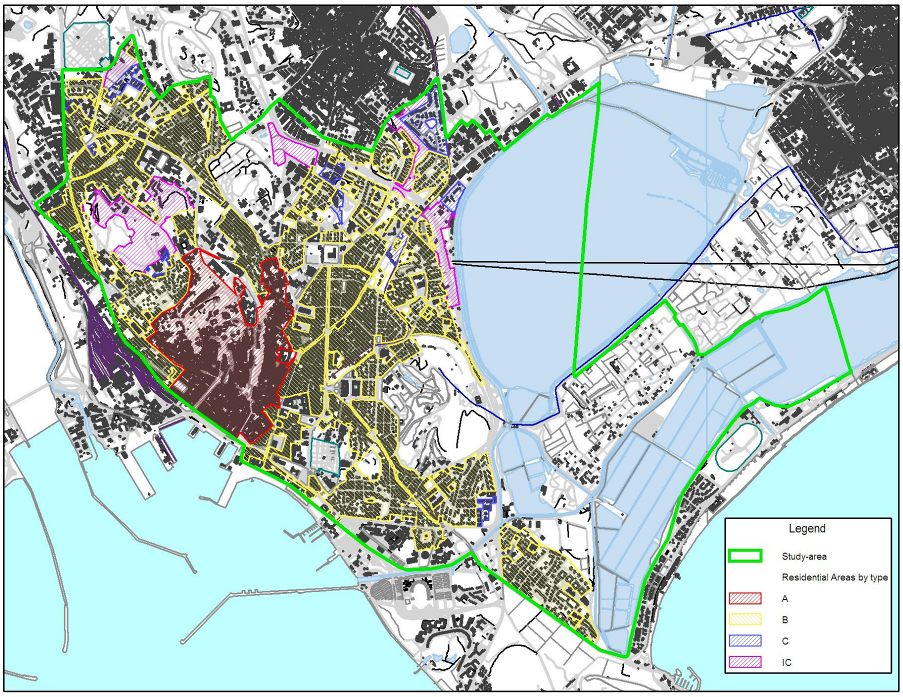

Figure 2. Location of the residential areas, aggregated and classed according to their type

3.1.3. Spatial analysis approach
Once the geometry of the housing areas is constructed, the following operations are implemented, in order to link the type of a residential area with the relevant information.
As a general procedure, we choose to analyze a buffer zone around each housing area, where distance value equals 150 meters. This choice is based on the assumption that the residential characterization of an area is influenced not only by its own morphologic, environmental, demographic and legal qualities, but also by those of its surrounding areas.
The most suitable spatial analysis methodology to deal with this issue is one of the earliest and simplest ones, the overlay mapping technique for vector datasets. Given two different datasets, this technique makes it possible to obtain a third dataset, which incorporates all the information provided by the input. The new dataset contains new levels of site characteristics.[4]

[4] This paper does not aim at discussing how the method works and which problems might occur; for a discussion covering both geometry and topology in overlay mapping, see Chrisman, 2002.

By intersecting the residential areas and the zones in which building is forbidden, it is immediately possible to determine whether a part of the area is affected by this prohibition, and which is the measure of the overlap zone (this is not true, as we explain later, for buildings and areas characterized by a cultural importance).
The same methodology, consisting of a combination of buffer creation and overlay mapping, is also used to fill in the fields related to land cover, resident population, and buildings/areas characterized by their cultural/historic importance; however, some additional hypotheses and procedures are required, either because of data geometry or because of overlaying outcomes.
First, the study area consists of 33 different types of fourth-level land uses. Such a detailed definition is assumed to be relevant only for land uses which detail type 1.1, "urban fabric," which is the most common land use in urban environment. A combination of the fourth and third level has been introduced. This simplification makes the number of land-cover types decrease to 24. Moreover, each residential zone is characterized by a combination of land uses (minimum 2, maximum 11) so peculiar that the association housing area - combination of land uses would have been unique, and the finding of a pattern based upon land cover would have been impossible, had we used the fourth and third-level land uses. Hence, a further simplification is made, by assuming that each housing area can be described as a combination of land uses which represents at least a 60 percent of the buffer zone around a single residential area. As a result, the number of combinations of land uses in our study area plunges to five.
Second, when calculating the resident population for each zone, a problem occurs, because of the boundaries of the Census Tracts, which do not coincide exactly with those of residential zones, despite the fact that they do represent the same real object (usually, a side of a street). When using data obtained from different sources, with different levels of accuracy, or produced in different periods of time, this is a fairly common problem, which affects the outcome of the spatial intersection between two datasets by producing a number of small polygons. [5] Because of the presence of slivers, and in order to avoid an overestimation, we assume that the resident population is uniformly distributed in each Census Tract.
Finally, another hypothesis, concerning zones protected by the law due to their historic importance, is assumed. In our database, the geometry type is a mix of points (24 out of 42) and polygons (18). The presence of the points makes it impossible to know the amount of the protected area contained in each buffer zone. We assume that each unknown area associated with protected buildings equals the smallest known one. This assumption implies that the bigger the area, the stronger the importance.

3.2. Results of the spatial analysis

This paragraph presents the outcomes of the spatial analysis procedure.
The information table associated with geographic features has the following fields (see Table 4):

- "Area Code," which takes three values (B, C, EZ);
- "Area," which contains the value of the area of a residential zone;
- "Area_A" (or "Area_B," or "Area_C," or "Area_EZ"), which is the area of the A (or B, or C, or EZ) zone which overlaps the buffer around a residential zone;
- "Population," which contains the number of residents in a buffer;
- "Land-use Code," which is the combination of the prevailing land uses in a buffer;
- "Nobuild Area," the area affected by building prohibition inside a buffer.

[5] The so-called *slivers*; see, for instance, Longley et al., 2001.

Area Code	Area (square meters)	Area_A (square meters)	Area_B (square meters)	Area_C (square meters)	Area_EZ (square meters)	Population	Land-use code	Land-use area (square meters)	Nobuild area (square meters)
B	8335	0	0	6969	68474	416	1112	95728	62886
B	3621	0	42252	0	327	917	1111-1112	71809	0
B	500416	0	217307	0	0	11223	1111-1112	1140922	368277
B	323029	0	149847	22815	77396	6247	1111-1112	691978	242041
B	42665	0	57228	4560	28842	1778	1112	210392	0
B	437279	0	199216	25579	0	12343	1111-1112	897643	125383
B	94822	0	36893	50062	55147	1298	1112-121	323598	0
B	43030	11253	2270	0	64840	1486	1111	173140	18891
B	329394	0	314867	3791	0	12724	1111	778955	1529
B	114846	0	132787	0	0	4407	1111	354449	28306
B	141418	0	132544	20227	0	6414	1111-1112	363691	196
B	5263	0	42211	0	10	748	1111-1112	95280	0
B	135148	0	212029	0	0	9011	1111	378611	196
B	150587	58828	126518	0	0	8759	1111	462255	99692
B	173056	70283	145212	0	0	8072	1111	540456	79247
B	194992	0	179038	746	99071	10176	1111-1112	517814	0
B	90866	0	184100	7264	0	7512	1111-1112	388518	196
B	20764	0	14105	0	116487	1676	121	119702	0
B	84640	0	75437	27082	0	4629	1112	275598	109930
B	135231	0	117145	58361	13160	7740	1112	447403	33591
B	79250	0	90838	15096	0	4570	1112-121	313198	196
B	17774	0	65182	4842	61102	1065	1112	150178	0
B	264844	105253	63718	11131	71239	3752	1111	521259	161360
B	116678	103875	68029	0	0	3765	1111	341613	145337
B	53874	0	39669	19762	18705	2487	1112	228265	57600
B	92625	0	64884	0	39985	3462	1112-121	300594	2473
B	310353	0	102195	0	2551	10212	1112-121	600360	167458
B	286462	85996	40762	0	0	8651	1111-1112	780699	290070
B	49697	0	106633	25542	5744	2567	1112	267460	0
B	422135	0	0	534	0	5778	1112	652233	191100
B	149764	77650	38048	0	0	3640	1111-1112	511887	532302
B	110194	0	104158	30415	60633	4301	1112	386309	0
B	77776	0	38412	43142	21877	2581	1111-1112	244837	3329
B	5088	0	44267	20944	0	812	1112	85782	0
B	102496	0	96174	40003	5838	4219	1112	398655	0
B	1037	0	36609	1416	0	1174	1111	62854	0
C	30404	0	17086	0	30582	170	121	51501	68202
C	130690	0	90688	2422	78244	2007	1111-1112	383821	196
C	25579	0	67775	0	0	1730	1112	143891	32564
C	41569	0	81494	11454	47899	2653	1112	275583	44464
C	6210	2836	39866	3427	43389	433	1111	114959	2921
C	3971	0	74479	4724	13317	492	1111	110148	0
C	950	0	24617	3928	39171	192	1111	56942	0
C	7832	0	43383	0	0	1370	1112-121	142460	449
C	7264	0	80635	0	0	2121	1111-1112	133826	196
C	28123	0	70175	7993	7467	1115	1111-1112	182316	0
C	20944	0	111828	1045	0	1547	1112	152760	0
C	21706	0	73758	9159	0	1501	1111-1112	179780	0
C	1870	0	53530	1225	10851	631	1112	99820	0
EZ	121149	0	143275	22916	0	1926	1112	264728	146174
EZ	4060	0	58251	0	9806	1014	1112	112776	0
EZ	75452	0	224127	111684	4060	4122	1111-1112	533163	196
EZ	348331	4766	230140	10970	2166	5904	1111	657172	5063
EZ	82385	0	16364	12110	0	263	1112-121	190585	155856
EZ	2166	0	55901	0	12016	1172	1111-1112	64032	0
EZ	56152	0	43822	37590	49947	1611	1112	230114	123878

Table 4. Descriptive attributes of the database

The following paragraphs describe the database, and try to connect it both to type and to spatial location of the housing areas.

3.2.1. <u>Area of the A zone contained in the buffer around a residential zone</u>

Only in 9 cases out of 56 does the attribute Area_A take a value different from zero, which means that only nine residential areas are closer to the historic center less than 150 meters.

Seven are B zones, one is a C zone and one is an EZ zone; additionally, the latter are characterized by the smallest overlap values (the C zone because of its small area, and the EZ zone due to its distance from the A zone). Since the A zone corresponds to the central area of the city, most of the inner part of the urban fabric is classified as B zone. Thus, the inner part of the city is almost completely built, with only few areas available for new housing.

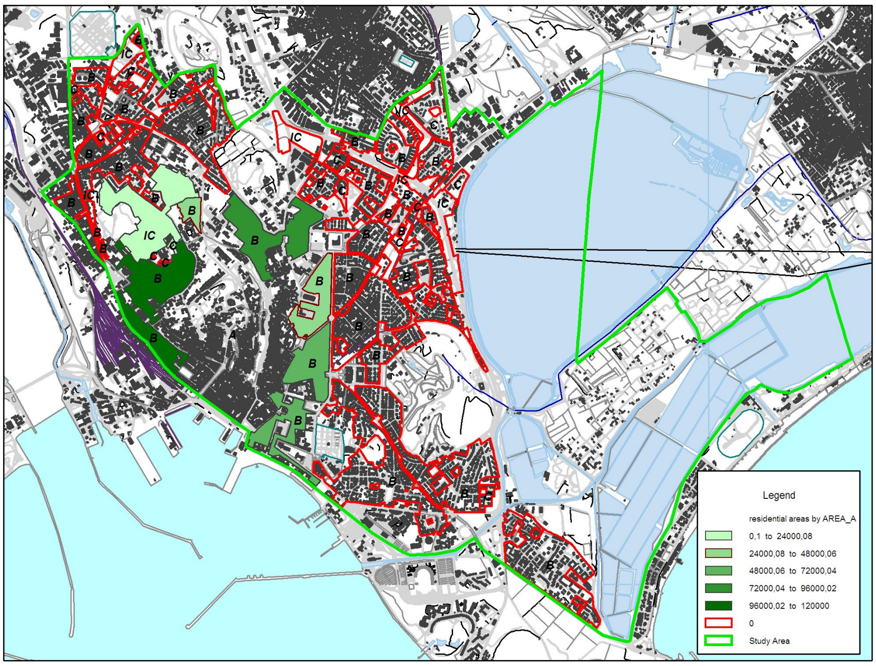

Figure 3. Residential zones classified according to the attribute “Area_A”

<u>3.2.2. Area of the B zone contained in the buffer around a residential zone</u>

In 54 cases out of 56 the attribute Area_B takes value different from zero.

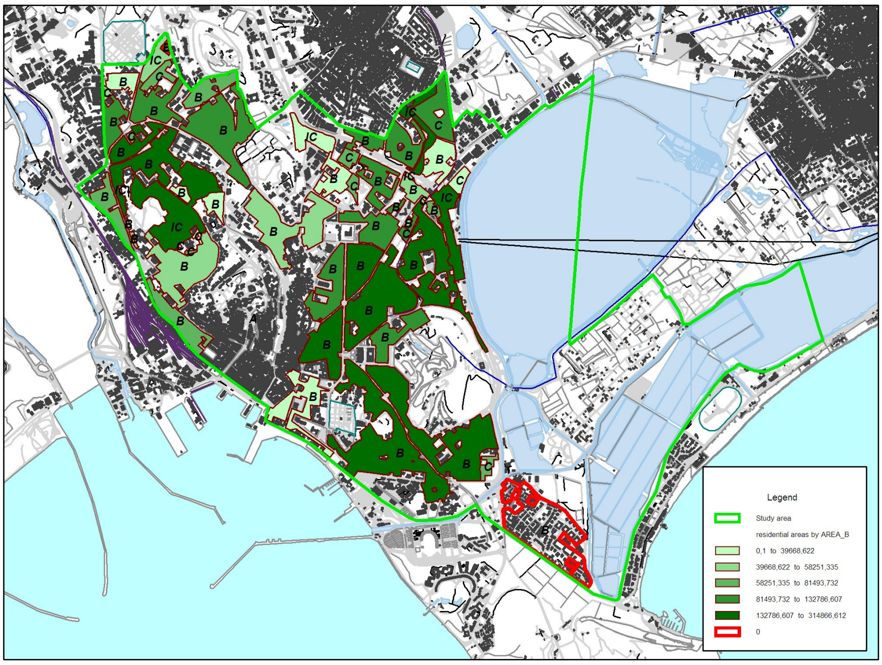

Figure 4. Residential zones classified according to the attribute “Area_B”

This means that in our study area almost every residential zone, regardless of its own type, is very close to at least one B zone. There are only two areas (both of them are B zones) which do not overlap any buffer around a completion zone. The first one is completely parted from other residential zones, since it is surrounded by an area reserved for military uses, a wetland and a natural park, whereas the second one is surrounded by an EZ zone and a recreational urban area. The highest values of this attribute are always connected with B zones and EZ zones. This seems to happen both because of their dimension and because of their position.

<u>3.2.3. Area of the C zone contained in the buffer around a residential zone</u>

In 36 cases out of 56 the attribute Area_C takes non-zero values. Twenty-two are B zones, 9 are C zones and 5 are EZ zones. Since the C zones are mostly located in the outer part of the city, areas with highest values of this attribute (B and EZ zones) shape a sort of buffer around the central districts.

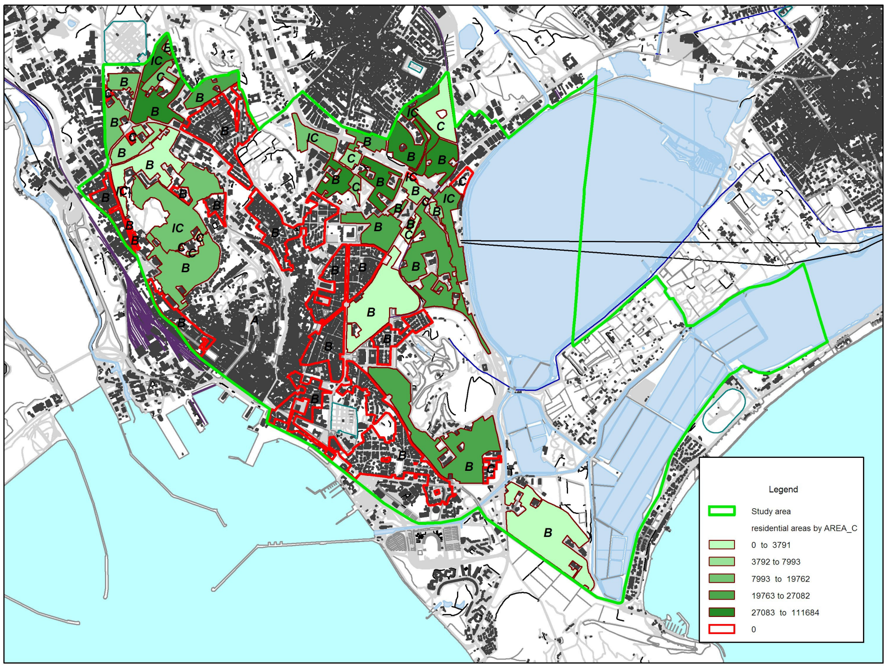

Figure 5. Residential zones classified according to the attribute "Area_C"

<u>3.2.4. Area of the EZ zone contained in the buffer around a residential zone</u>

In 32 cases out of 56 the attribute Area_EZ takes non-zero values. Nineteen are B zones, 8 are C zones and 5 are EZ zones. With reference to their spatial distribution, these areas form two main clusters in the northern outer part of the city. The reason must be sought both in the small number of enterprise zones in our study area and in their proximity to each other.

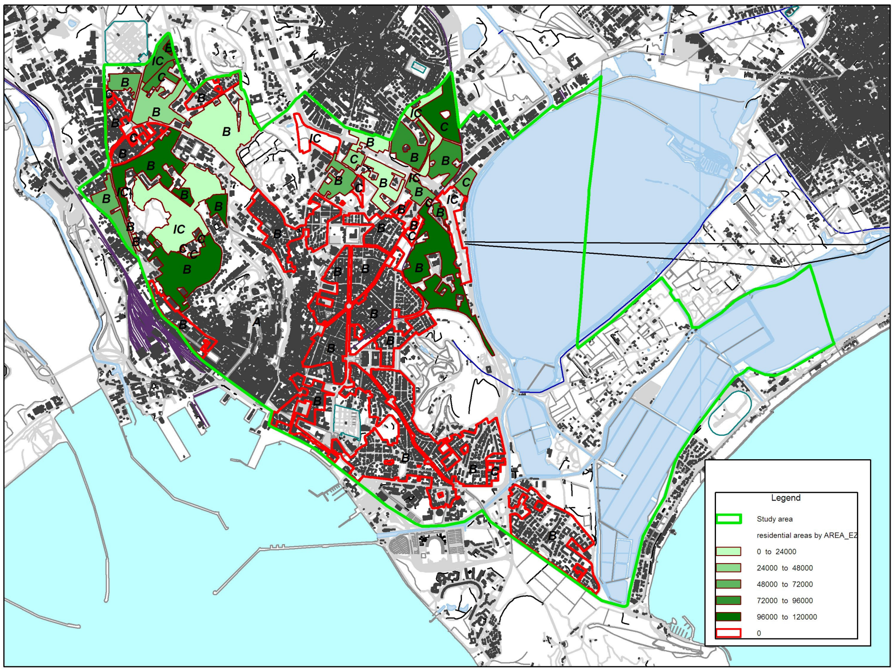

Figure 6. Residential zones classified according to attribute “Area_EZ”

<u>3.2.5. Land use</u>

Out of 56 residential areas, 13 are classified as “1111” (dense-residential-settlement land use), 18 as “1112” (scattered-residential-settlement land use), 17 as “1111-1112” (a mix of dense and scattered-residential-settlement land use), 6 as “1112-121” (a mix of scattered-residential-settlement land use, and industrial, commercial and services land uses) and 2 as “121” (industrial, commercial and services land uses).

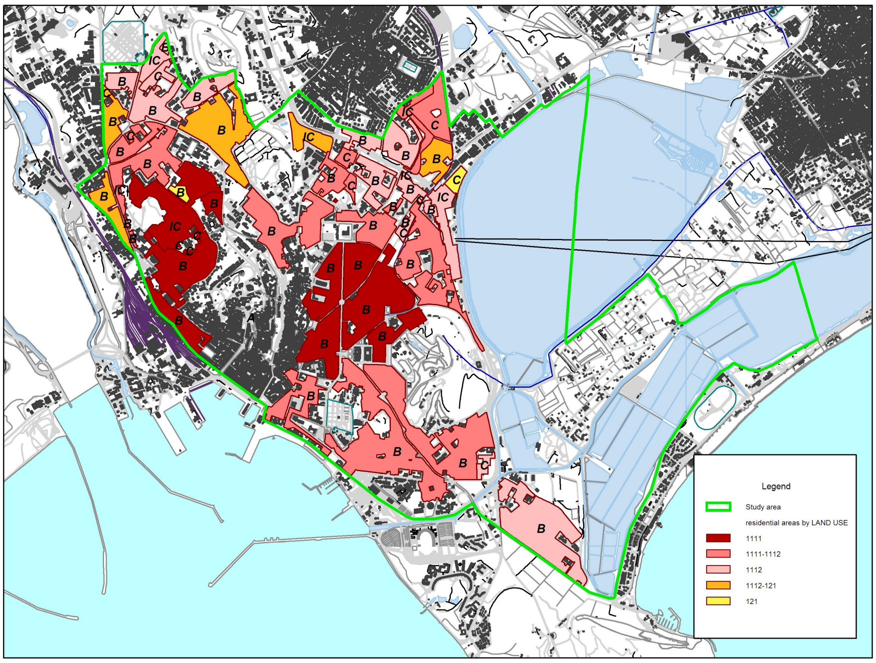

Figure 7. Residential zones classified according to the attribute “LAND_USE”

Figure 7 shows a gradual transition in prevailing land uses, when moving from the center to the outer parts of the city, from 1111 to 1111-1112, then to 1112-121. The absence of commercial and services uses in the city center might be surprising. However, this happens because of the procedure used in constructing residential areas. It should be reminded that areas reserved for local recreational, cultural, social or sports activities are included in residential zones, whereas areas for services and trading with a greater level of importance are excluded.

3.2.6. Population

As Figure 8 shows, areas with the highest values of population, which are always B zones, form a continuous shape that surrounds the city center. Outer zones are characterized by lower values. However, the same happens with some inner areas. Three main reasons justify this spatial distribution.

First, the dimension of each residential zone (and, as a consequence, the dimension of its buffer, with respect to which resident population is calculated). Second, the differences in the amount of built volume permitted in each type of residential zone (as explained in 3.1.2.). Third, the different types of buildings, particularly with reference to the number of stories (this is connected to the second point above).

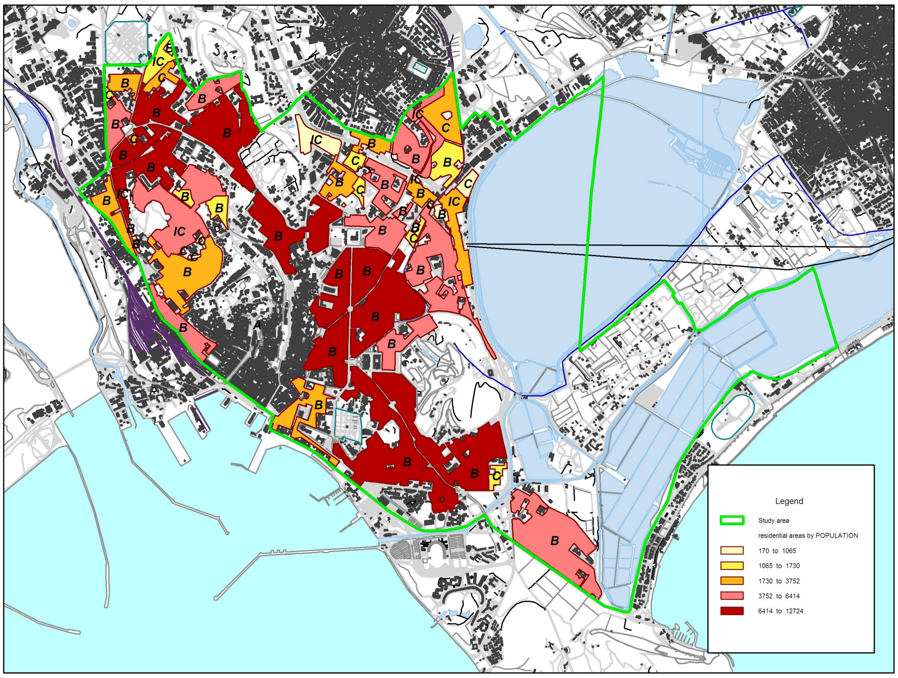

Figure 8. Residential zones classified according to attribute "Resident population"

3.2.7. Building prohibition

Spatial distribution of areas where building is not permitted is influenced by many factors.

On the one side, the inner areas (all of which are B zones) are affected primarily by their proximity to the historical center and by the presence of monuments and important buildings/areas. On the other side, the eastern and the western residential areas are affected by their closeness to environmentally-relevant sites (parks, wetlands, coastline). The northern area is influenced by a landscape plan and by the presence of the city cemetery. Thus, different key factors play a role in prohibition on new housing, and they are scattered throughout the study area.

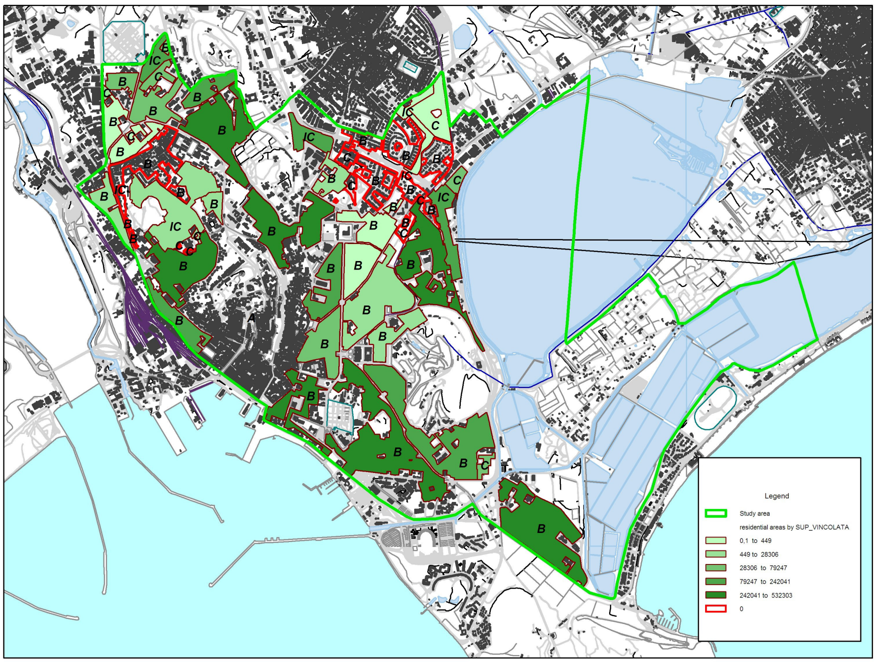

Figure 9. Residential zones classified according to attribute “Nobuild area”

4. ANALYSIS OF THE SPATIAL CONFIGURATION OF THE RESIDENTIAL AREAS OF CAGLIARI

The analysis of the spatial configuration of the residential areas of Cagliari is presented in two parts. The first part contains the definition of the sets of the objects, OBJ, and of the attributes, A (see Section 2 for the meaning of these notations), and a description of the implementation of the RSA procedure based on the methodology discussed in Section 2. The second part presents the results from the implementation of the RSA procedure.

4.1. Objects, attributes and decision rules

The elements of OBJ are fifty-six residential zones of the city of Cagliari (see Section 3). Each zone is identified as "residential" by the zoning rules of the city Masterplan. It contains at least one block. Thus, each area is bounded by city streets.

The attributes are the following (all the attributes refer to the area which belongs to a buffer of 150 meters around one of the fifty-six residential zones identified by the city Masterplan, with the exception of the last attribute, which is the decision attribute):

- area of the historic center zone (name of the attribute: AREA_A) - this attribute takes two values: "YES" if a part of the historic center zone belongs to the buffer of the residential zone; "NO" otherwise;
- area of the residential completion zone (name of the attribute: AREA_B) - this attribute takes two values: "HIGH" if the area is greater than the fiftieth percentile (about seven hectares); "LOW" otherwise;
- area of the residential expansion zone (name of the attribute: AREA_C) - this attribute takes two values: "HIGH" if the area is greater than the fiftieth percentile (about 1,1 hectares); "LOW" otherwise;
- area of the enterprise zone (name of the attribute: AREA_EZ) - this attribute takes two values: "HIGH" if the area is greater than the fiftieth percentile (about three hectares); "LOW" otherwise;
- area of the prevailing land use (name of the attribute: L_USE) - this attribute takes four values: "DRS" if dense-residential-settlement land use prevails; "SRS" if scattered-residential-settlement land use prevails; "DRS-SRS" if a mix of dense and scattered-residential-settlement land uses prevails; "ICS" if industrial-commercial-service land use prevails;
- resident population (name of the attribute: POP) - this attribute takes two values: "HIGH" if the residents are more than the fiftieth percentile (about 2.500 people); "LOW" otherwise;
- area where building is forbidden, because of environmental protection, cemeteries, archaeological resources, landscape protection and so on (name of the attribute: NOBUILD) - this attribute takes two values: "HIGH" if the area is greater than the fiftieth percentile (about five hectares); "LOW" otherwise;
- class of the residential zone established by the zoning rules of the city Masterplan (name of the attribute: AREA_COD) - this attribute takes four values: "B-HIGH" if the zone is classified as "residential completion zone" and its area is greater than the fiftieth percentile (about eight hectares); "B-LOW" if the zone is classified as "residential completion zone" and its area is smaller than the fiftieth percentile (about eight hectares); "C" if the zone is classified as "residential expansion zone;" "EZ" if the zone is classified as "enterprise zone."

The analytical description of the values of the attributes is shown in Table 5.

The procedure described in Section 2 is used to define DRDM, $DRED_K$'s and decision rules within the context of the case study. This procedure is implemented through the program

"Rosetta," developed by Øhrn (1999; 2001). In this case study all the attributes are alphabetical, so there is not a problem of variable discretization. Moreover, there are not missing values in the information matrix, so there is not a problem of completion of the information matrix. Information is based on alphabetical values and it is complete.

Attributes-Values	Occurrences	Frequency (%)	Attributes-Values	Occurrences	Frequency (%)
AREA_A			**L_USE**		
YES	9	16,07	DRS	13	23,21
NO	47	83,93	SRS	18	32,14
			DRS-SRS	17	30,36
AREA_B			ICS	8	14,29
HIGH	28	50,00			
LOW	28	50,00	**POP**		
			HIGH	28	50,00
AREA_C			LOW	28	50,00
HIGH	18	32,14			
LOW	38	67,86	**NOBUILD**		
			HIGH	18	32,14
AREA_EZ			LOW	38	67,86
HIGH	16	28,57			
LOW	40	71,43	**AREA_COD**		
			B-HIGH	23	41,08
			B-LOW	13	23,21
			C	13	23,21
			EZ	7	12,50

Table 5. Descriptive statistics of the attributes

The information matrix is 56x8, since the elements of OBJ, the residential zones of the city of Cagliari, are fifty-six, and the condition and decision attributes are eight.

4.2. Results

The decision rules are generated using the Johnson algorithm of the Rosetta program. This algorithm is the most efficient in case of alphabetical attributes (Øhrn, 2001, pp. 24–26).
Table 6 shows the rules and their descriptive statistics. The rules are twenty-nine. Each rule refers to one or more values of the decision attribute, according to whether it is exact or approximate. The decision attribute can take four values: "B-HIGH," "B-LOW," "C," and "EZ." The rules which refer to: "B-HIGH" are thirteen (eleven exact rules – two approximate rules); "B-LOW" are seven (three exact rules – four approximate rules); "C" are eight (five exact rules – three approximate rules); "EZ" are seven (three exact rules – four approximate rules).
Table 6 shows the decision rules and their descriptive statistics. The RHSA of each exact rule equals 1. The sum of the RHSA of each approximate rule also equals 1. This indicates that all the rules are highly reliable.

DECISION RULES	LHSS	RHSS	RHSA	LHSC	RHSC
1. POP(HIGH) AND NOBUILD(HIGH) => AREA_COD(B-HIGH)	12	12	1,00	0,21	0,52
2. AREA_A(NO) AND AREA_C(LOW) AND POP(HIGH) => AREA_COD(B-HIGH)	9	9	1,00	0,16	0,39
3. AREA_A(YES) AND NOBUILD(HIGH) => AREA_COD(B-HIGH)	6	6	1,00	0,11	0,26
4. AREA_B(LOW) AND AREA_C(LOW) AND POP(HIGH) => AREA_COD(B-HIGH)	5	5	1,00	0,09	0,22
5. L_USE(DRS-SRS) AND NOBUILD(HIGH) => AREA_COD(B-HIGH)	5	5	1,00	0,09	0,22
6. AREA_EZ(LOW) AND L_USE(SRS) AND POP(HIGH) AND NOBUILD(LOW) => AREA_COD(B-HIGH) OR AREA_COD(B-LOW)	3	2, 1	0,67, 0,33	0,05	0,09, 0,08
7. AREA_A(YES) AND L_USE(DRS-SRS) => AREA_COD(B-HIGH)	2	2	1,00	0,04	0,09
8. AREA_B(HIGH) AND AREA_C(HIGH) AND L_USE(DRS-SRS) AND NOBUILD(LOW) => AREA_COD(B-HIGH) OR AREA_COD(EZ)	2	1, 1	0,50, 0,50	0,04	0,04, 0,14
9. AREA_B(HIGH) AND AREA_EZ(HIGH) AND L_USE(SRS) => AREA_COD(B-HIGH) OR AREA_COD(C)	2	1, 1	0,50, 0,50	0,04	0,04, 0,08
10. AREA_C(HIGH) AND AREA_EZ(HIGH) AND L_USE(ICS) => AREA_COD(B-HIGH)	1	1	1,00	0,02	0,04
11. AREA_B(LOW) AND L_USE(DRS-SRS) AND POP(LOW) => AREA_COD(B-LOW) OR AREA_COD(EZ)	3	2, 1	0,67, 0,33	0,05	0,15, 0,14
12. AREA_B(LOW) AND AREA_C(HIGH) AND AREA_EZ(LOW) AND L_USE(SRS) => AREA_COD(B-LOW)	2	2	1,00	0,04	0,15
13. AREA_B(LOW) AND AREA_C(HIGH) AND AREA_EZ(LOW) AND NOBUILD(LOW) => AREA_COD(B-LOW)	2	2	1,00	0,04	0,15
14. AREA_C(LOW) AND AREA_EZ(HIGH) AND L_USE(SRS) => AREA_COD(B-LOW)	2	2	1,00	0,04	0,15
15. AREA_B(LOW) AND AREA_C(LOW) AND AREA_EZ(LOW) AND L_USE(SRS) AND POP(LOW) => AREA_COD(B-LOW) OR AREA_COD(C) OR AREA_COD(EZ)	4	1, 2, 1	0,25, 0,50, 0,25	0,07	0,08, 0,15, 0,14
16. AREA_A(NO) AND AREA_B(LOW) AND AREA_EZ(LOW) AND L_USE(DRS) => AREA_COD(B-LOW)	1	1	1,00	0,02	0,08
17. AREA_B(HIGH) AND AREA_C(HIGH) AND L_USE(ICS) => AREA_COD(B-LOW)	1	1	1,00	0,02	0,08
18. AREA_B(LOW) AND AREA_C(HIGH) AND L_USE(DRS-SRS) => AREA_COD(B-LOW)	1	1	1,00	0,02	0,08
19. AREA_C(LOW) AND AREA_EZ(HIGH) AND L_USE(ICS) AND POP(LOW) AND NOBUILD(LOW) => AREA_COD(B-LOW)	1	1	1,00	0,02	0,08
20. AREA_A(YES) AND POP(LOW) => AREA_COD(B-LOW) OR AREA_COD(C)	2	1, 1	1,00	0,04	0,08, 0,08
21. AREA_B(HIGH) AND AREA_C(LOW) AND POP(LOW) => AREA_COD(C)	6	6	1,00	0,11	0,46
22. AREA_A(NO) AND AREA_EZ(HIGH) AND L_USE(DRS) => AREA_COD(C)	1	1	1,00	0,02	0,08
23. AREA_B(HIGH) AND AREA_C(LOW) AND L_USE(SRS) => AREA_COD(C)	1	1	1,00	0,02	0,08
24. AREA_B(LOW) AND AREA_C(LOW) AND AREA_EZ(LOW) AND L_USE(ICS) => AREA_COD(C)	1	1	1,00	0,02	0,08
25. AREA_EZ(HIGH) AND L_USE(ICS) AND NOBUILD(HIGH) => AREA_COD(C)	1	1	1,00	0,02	0,08
26. AREA_A(YES) AND AREA_B(HIGH) AND NOBUILD(LOW) => AREA_COD(EZ)	1	1	1,00	0,02	0,14
27. AREA_B(HIGH) AND AREA_C(HIGH) AND POP(LOW) => AREA_COD(EZ)	1	1	1,00	0,02	0,14
28. AREA_B(LOW) AND AREA_C(HIGH) AND AREA_EZ(HIGH) AND L_USE(SRS) => AREA_COD(EZ)	1	1	1,00	0,02	0,14
29. AREA_C(HIGH) AND L_USE(ICS) AND NOBUILD(HIGH) => AREA_COD(EZ)	1	1	1,00	0,02	0,14

Table 6. Rules and descriptive statistics

The LHSC of the rules ranges from 21 percent to 2 percent. This indicates that the different shares of the phenomenon represented by the decision attribute that each rule contributes to explain are generally small, even though significant differences exist. Table 6 shows the decision rules ordered by decision attribute and by LHSC. The first rule of Table 6 refers to the value "B-HIGH" of the decision attribute. This rule has the highest LHSC (21 percent), followed by the second rule concerning the value "B-HIGH" (16 percent). Four other rules have a comparatively high LHSC (11 percent): (i) the third rule concerning the value "B-HIGH;" (ii) the first rule concerning the value "C" (the twenty-first rule of Table 6). The LHSC of the other rules ranges between 9 percent and 2 percent. Each of them does not explain more than five occurrences.

The RHSC of the rules ranges from 52 percent to 8 percent. A 52 percent of the occurrences of the value of the decision attribute "B-HIGH" is explained by the first decision rule of Table 6; a 39 percent is explained by the second rule; a 26 percent is explained by the third rule. A 15 percent of the occurrences of the value of the decision attribute "B-LOW" is explained by the first, second and third rule concerning this value (the twelfth, thirteenth and fourteenth rule of Table 6). A 46 percent of the occurrences of the value of the decision attribute "C" is explained by the first rule concerning this value (the twenty-first rule of Table 6). Each of the seven rules concerning the decision attribute "EZ" explains one occurrence (a 14 percent of the occurrences).

Attributes-Values	Explained occurrences	Frequency (%)	Decision Attribute = "B-HIGH"	Frequency (%)	Decision Attribute = "B-LOW"	Frequency (%)	Decision Attribute = "C"	Frequency (%)	Decision Attribute = "EZ"	Frequency (%)
AREA_A	**22**	**27,50**	**17**	**38,64**	**2**	**13,33**	**2**	**14,29**	**1**	**14,29**
YES	11	13,75	8	18,18	1	6,67	1	7,14	1	14,29
NO	11	13,75	9	20,45	1	6,67	1	7,14	0	0,00
AREA_B	**32**	**40,00**	**7**	**15,91**	**8**	**53,33**	**11**	**78,57**	**6**	**85,71**
HIGH	14	17,50	2	4,55	1	6,67	8	57,14	3	42,86
LOW	18	22,50	5	11,36	7	46,67	3	21,43	3	42,86
AREA_C	**30**	**37,50**	**7**	**15,91**	**8**	**53,33**	**10**	**71,43**	**5**	**71,43**
HIGH	12	15,00	2	4,55	6	40,00	0	0,00	4	57,14
LOW	18	22,50	5	11,36	2	13,33	10	71,43	1	14,29
AREA_EZ	**22**	**27,50**	**4**	**9,09**	**10**	**66,67**	**6**	**42,86**	**2**	**28,57**
HIGH	9	11,25	2	4,55	3	20,00	3	21,43	1	14,29
LOW	13	16,25	2	4,55	7	46,67	3	21,43	1	14,29
L_USE	**34**	**42,50**	**10**	**22,73**	**12**	**80,00**	**7**	**50,00**	**5**	**71,43**
DRS	2	2,50	0	0,00	1	6,67	1	7,14	0	0,00
SRS	15	18,75	3	6,82	6	40,00	4	28,57	2	28,57
DRS-SRS	11	13,75	6	13,64	3	20,00	0	0,00	2	28,57
ICS	6	7,50	1	2,27	2	13,33	2	14,29	1	14,29
POP	**45**	**56,25**	**27**	**61,36**	**6**	**40,00**	**9**	**64,29**	**3**	**42,86**
HIGH	28	35,00	27	61,36	1	6,67	0	0,00	0	0,00
LOW	17	21,25	0	0,00	5	33,33	9	64,29	3	42,86
NOBUILD	**34**	**42,50**	**26**	**59,09**	**4**	**26,67**	**1**	**7,14**	**3**	**42,86**
HIGH	25	31,25	23	52,27	0	0,00	1	7,14	1	14,29
LOW	9	11,25	3	6,82	4	26,67	0	0,00	2	28,57

Table 7. Number and frequency of the explained occurrences of the decision attribute for each condition attribute with reference to the decision rules

Thus, the most part of occurrences are explained by a few rules with respect to a total of twenty-nine. The role each condition attribute plays in generating the decision rules is fundamental to understand the phenomenon represented by the values of the decision attribute.

The decision attribute can take four values: (i) "B-HIGH," which indicates a residential completion zone with an area greater than eight hectares; (ii) "B-LOW," which indicates a residential completion zone with an area smaller than eight hectares; (iii) "C," which indicates a residential expansion zone; (iv) "EZ," which indicates an enterprise zone.

The decision rules connect the condition attributes with each other in order to identify the value of the decision attribute, that is how the zoning rules of the Masterplan of Cagliari characterize the use of a city residential area.

The relations between condition and decision attributes are presented in four parts. Each part refers to one of the four values of the decision attribute. Table 7 shows the statistics concerning these relations.

Residential completion zones with an area greater than eight hectares (B-HIGH)

The most important condition attribute to identify relatively vast residential completion zones is POP. High values of POP identify a 61,36 percent of the occurrences of B-HIGH. Moreover, low values of POP are never connected with B-HIGH.

The attribute NOBUILD is also important (59,09 percent of the occurrences of B-HIGH). This attribute is fairly more important if its value is HIGH (52,27 percent versus 6,82 percent if its value is LOW).

The attribute AREA_A comes third (38,64 percent). This attribute is slightly more important if its value is NO (20,45 percent versus 18,18 percent if its value is YES).

The attribute L_USE identifies a 22,73 percent of the occurrences of B-HIGH. The most important types of land uses are DRS-SRS (a mix of dense and scattered-residential-settlement land uses, 13,64 percent) and SRS (scattered-residential-settlement land use, 6,82 percent). Less important is ICS (industrial-commercial-services land use, 2,27 percent). Moreover, DRS (dense-residential-settlement land use) is never connected with B-HIGH.

Each of the attributes AREA_B and AREA_C identifies a 15,91 percent of the occurrences of B-HIGH. They are more important if their value is LOW (11,36 percent versus 4,55 percent if its value is HIGH).

The least important condition attribute is AREA_EZ (9,09 percent). Its importance is the same whether its value is HIGH or LOW (4,55 percent).

Thus, relatively vast residential completion zones are mostly identified, in a buffer of 150 meters around the completion zone, by: the presence of a relatively high resident population (more than 2.500 residents); a low presence of residential expansion zones (less than 1,1 hectares), of residential completion zones (less than seven hectares), and of enterprise zones (less than three hectares); a non-dense-residential-settlement land use; the presence of a relatively vast zone where building is forbidden (more than five hectares). The presence and the absence of areas of the historic center zone do not play a definite role with respect to residential completion zones with an area greater than eight hectares.

The first five decision rules of Table 6 are very effective in summarizing the importance of the attributes in identifying the value B-HIGH of the condition attribute.

Residential completion zones with an area smaller than eight hectares (B-LOW)

The most important condition attribute to identify relatively small residential completion zones is L_USE (80,00 percent of the occurrences of B-LOW). The most important types of land use are SRS (40,00 percent) and DRS-SRS (20,00 percent). Less important are ICS (13,33 percent) and DRS (6,67 percent).

The attribute AREA_EZ is also important (66,67%). This attribute is fairly more important if its value is LOW (46,67 percent versus 20,00 percent if its value is HIGH).
The attributes AREA_B and AREA_C identify a 53,33 percent of the occurrences of B-LOW. AREA_B is fairly more important if its value is LOW (46,67 percent versus 6,67 percent if its value is HIGH). On the contrary, AREA_C is more important if its value is HIGH (40,00 percent versus 13,33 percent if its value is LOW).
The attribute POP comes fifth (40%). This attribute is fairly more important if its value is LOW (33,33 percent versus 6,67 percent if its value is HIGH).
Low values of the attribute NOBUILD identify a 26,67 percent of the occurrences of B-LOW. Moreover, high values of NOBUILD are never connected with B-LOW.
The least important condition attribute is AREA_A (13,33 percent). Its importance is the same whether its value is YES or NO (6,67 percent).
Thus, relatively small residential completion zones are mostly identified, in a buffer of 150 meters around the completion zone, by: a high presence of residential expansion zones (more than 1,1 hectares) and a low presence of enterprise zones (less than three hectares); a non-dense-residential-settlement land use; a relatively low resident population (less than 2.500 residents); a low presence of residential completion zones (less than seven hectares); the presence of a relatively small zone where building is forbidden (less than five hectares). The presence and the absence of areas of the historic center zone do not play a definite role with respect to residential completion zones with an area smaller than eight hectares.
The first four decision rules concerning the decision attribute B-LOW (rules 11-14 of Table 6) are very effective in summarizing the importance of the attributes in identifying the value B-LOW of the condition attribute. In the fourth rule, the value of AREA_EZ is HIGH, which should not be amazing, since the value LOW of AREA_EZ identifies a significant fraction of the occurrences of B-LOW, even though smaller than that identified by the value HIGH.

<u>Residential expansion zones (C)</u>

The most important condition attributes to identify residential expansion zones are AREA_B and AREA_C. AREA_B identifies a 78,57 percent of the occurrences of C. This attribute is fairly more important if its value is HIGH (57,14 percent versus 21,43 percent if its value is LOW). Low values of AREA_C identify a 71,43 percent of the occurrences of C. Moreover, high values of AREA_C are never connected with B-HIGH.
The attribute POP is also important. Low values of POP identify a 64,29 percent of the occurrences of C. Moreover, high values of POP are never connected with C.
The attribute L_USE comes fourth. It identifies a 50,00 percent of the occurrences of C. The most important types of land use are SRS (28,57 percent) and ICS (14,29 percent). Less important is DRS (7,14 percent). Moreover, DRS-SRS is never connected with C.
The attribute AREA_EZ identifies a 42,86 percent of the occurrences of C. Its importance is the same whether its value is HIGH or LOW (21,43 percent).
Attributes AREA_A and NOBUILD are the least important. The attribute AREA_A identifies a 14,29 percent of the occurrences of C. Its importance is the same whether its value is YES or NO (7,14 percent). High values of the attribute NOBUILD identify a 7,14 percent of the occurrences of C. Low values of NOBUILD are never connected with C.
Thus, residential expansion zones are mostly identified, in a buffer of 150 meters around the expansion zone, by: a low presence of residential expansion zones (more than 1,1 hectares) and a high presence of residential completion zones (more than seven hectares); a scattered-residential-settlement land use with industrial-commercial-service land use; a relatively low resident population (less than 2.500 residents); the presence of enterprise zones (whether their area is large or small, that is greater or less than three hectares). The presence and the absence of areas of the historic center zone do not play a definite role with respect to residential ex-

pansion zones. The condition attribute NOBUILD (areas where building is forbidden) is very slightly connected with the value C of the decision attribute, since in only one rule a (high) value of NOBUILD is associated with a C value of the decision attribute.
The rule 21 of Table 6 is very effective in summarizing the importance of the attributes in identifying the value C of the condition attribute.

Enterprise zones (EZ)
The most important condition attribute to identify enterprise zones is AREA_B. It identifies a 85,71 percent of the occurrences of EZ. Its importance is the same whether its value is YES or NO (42,86 percent).
The attributes AREA_C and L_USE are also important. Each of them identifies a 71,43 percent of the occurrences of EZ. The attribute AREA_C is more important if its value is HIGH (57,14 percent versus 14,29 percent if its value is LOW). The most important types of land uses are SRS (28,57 percent) and DRS-SRS (28,57 percent). Less important is ICS (14,29 percent). Moreover, DRS is never connected with EZ.
Attributes POP and NOBUILD come fourth. Each of them identifies a 42,86 percent of the occurrences of EZ. High values of POP identify a 61,36 percent of the occurrences of EZ. Moreover, low values of POP are never connected with B-HIGH. The attribute NOBUILD is more important if its value is LOW (28,57 percent versus 14,29 percent if its value is HIGH).
The attribute AREA_EZ identifies a 28,57 percent of the occurrences of EZ. Its importance is the same whether its value is HIGH or LOW (14,29 percent).
The least important condition attribute is AREA_A. The YES value of this attribute identifies a 14,29 percent of the occurrences of EZ. Low values of AREA_A are never connected with EZ.
Thus, enterprise zones are mostly identified, in a buffer of 150 meters around the enterprise zone, by: the presence of residential completion zones (whether their area is large or small, that is greater or less than seven hectares); a high presence of residential expansion zones (more than 1,1 hectares); the presence of enterprise zones (whether their area is large or small, that is greater or less than three hectares); a non-dense-residential-settlement land use; a relatively low resident population (less than 2.500 residents); the presence of a relatively small zone where building is forbidden (less than five hectares). The condition attribute AREA_A (areas of the historic center zone) is very slightly connected with the value EZ of the decision attribute, since in only one rule a (YES) value of AREA_A is associated with a EZ value of the decision attribute.
There is no dominant rule among the seven concerning the decision attribute EZ.

5. DISCUSSION AND CONCLUSIONS

The analysis of the rules concerning the spatial configuration of the residential areas puts in evidence that large residential completion zones "B-HIGH" are mostly surrounded by highly-populated areas and areas where building is forbidden. Moreover, these zones are characterized by the proximity with non-dense-residential land uses, small residential completion and expansion zones.
There are important differences between large and small residential completion zones "B-LOW." The B-LOW's are mostly characterized by the proximity to residential expansion zones, enterprise zones and other B-LOW's, low-populated areas and areas with a low presence of building prohibitions. Above all, these zones are characterized by the proximity with non-dense-residential land uses.
Thus, the B-LOW's are boundary zones for the city residential completion sectors, whereas the B-HIGH's are located in the most internal areas of the completion sectors. The B-LOW's are generally characterized by surroundings with expansion and enterprise zones. These zones

are part of the most recent urban expansion. They are relatively less populated and less interested by building prohibitions.
The residential expansion zones are characterized by their proximity with residential completion zones (more than seven hectares), other residential expansion zones (less than 1,1 hectares), and enterprise zones. Moreover, these zones are characterized by the proximity with scattered-residential-settlement land uses. The residential expansion zones, characterized by scattered-residential land uses and a low resident population, work as a sort of a buffer between the completion zones, located in the inner-city highly-populated areas, and the enterprise zones.
The enterprise zones are characterized by their proximity with large residential expansion (more than 1,1 hectares) and completion zones. Thus, even though the expansion zones work as a buffer between completion and enterprise zones, the spatial configuration of the enterprise zones is not uniquely connected to the proximity to expansion zones. Rather, enterprise zones are spread over the urban fabric. These mixed-use (residential and public service) areas are used by the zoning rules of the city Masterplan without a preference for their surrounding zones, which might be whichever type of residential area.
As a result, the spatial configuration of the residential areas of Cagliari is generally characterized by a progressive increase in resident population, in density of the residential land uses, and in areas where building is forbidden, from the outer to the inner city. The completion zones are mostly located in the inner city, the expansion zones in the outer city. The enterprise zones are scattered and used to address the problem of the underendowment of public services for the residents, whether it arises in the inner or in the outer areas.
This paper has employed RSA to analyze the spatial configuration of the residential areas of the city of Cagliari. In doing so, it demonstrates how a spatial analysis approach based on a GIS can be utilized to figure out the geography of the residential zones within the urban fabric, thereby improving upon the objectivity and accuracy of RSA. Moreover, the application of this method allows for an integration of the results of the GIS and RSA approaches, which can be used by the city planners in the development of the policy-making processes concerning city residential areas. In this respect, the paper makes an important methodological contribution.
By applying the method developed in this paper, planners can better ensure that the policies they advocate are based on a thorough analysis of the spatial configuration of housing areas.
The zoning rules for urban planning of the Italian cities are quite similar to each other, since they are based on the same national law enacted by decree (n. 1444/1968). For this reason, the results obtained by the GIS-based RSA could be an important reference point to compare the characteristics of the spatial configuration of the residential areas of the Italian cities. In other words, an important feature of the methodology developed in this paper is that it is easily exportable, and, as a consequence, it allows for comparisons of different spatial configuration and policies. Moreover, the methodology can be applied to analyze the spatial configuration of the residential areas of other medium-sized European and North-American cities, once a thorough comparison of the zoning rules of their Masterplans is developed.
The optimal choice of the attributes to be included in the RSA includes as many variables as necessary to describe the housing market satisfactorily. Of course, this choice is heavily influenced by available information. In Italy, information quantity and quality are not as high as in the US, where empirical studies regarding the housing market have been largely developed.
The analysis implemented here is based on a set of variables representing the best choice given the information available, rather than the optimal choice. These variables should be considered a subset of the optimal variable choice.

Regarding this point, it must be stated that there are a number of variables that should have been included in the RSA and were not included since no information is available. One is the household income, which could be very important in determining the income effect on the spatial configuration of the residential areas. Moreover, data on capacity of the system of public infrastructure and services would be very helpful.

REFERENCES

Bruinsma, F., P. Nijkamp, and R. Vreeker. 2002. A comparative industrial profile analysis of urban regions in Western Europe: an application of rough set classification. *Tijdschrift voor Economische en Sociale Geografie* 93(4): 454-463.

Carlin, U., J. Komorowski, and A. Øhrn. 1998). Rough set analysis of patients with suspected acute appendicitis. In *Proceedings of the Seventh Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU '98).* Paris, France: EDK Editions Medicales et Scientifiques. 1528-1533.

Cerreta, M., and I. Salzano. 2004. Dalle valutazioni ex-post alle valutazioni ex-ante: un approccio metodologico per la conservazione integrata del patrimonio culturale [From ex-post to ex-ante assessment: a methodological approach for the integrated conservation of cultural heritage]. In *XXV Conferenza Italiana di Scienze Regionali [25th Conference of Regional Sciences]*, edited by Associazione Italiana di Scienze Regionali [The Italian Section of the Regional Science Association International]. Novara, Italy: Associazione Italiana di Scienze Regionali [The Italian Section of the Regional Science Association International]. Paper published in the compact disk of the Proceedings, Session E7: Valutazione delle politiche urbane [Urban policies assessment].

Chrisman, M. 2002. *Exploring Geographic Information Systems*. New York, United States: John Wiley & Sons.

Cilloccu, F. and A. Cumer. 2002. La nuova Carta dell'Uso del Suolo 1:25.000 della Regione Sardegna [Regional Administration of Sardinia's new Land Use Map scale 1:25.000]. *Documenti del Territorio*, 51(15):13-16.

City of Cagliari. 2004a. Masterplan of Cagliari. Available on the Internet at http:// comuneca.tiscali-business.it/ urbanistica/ Urbanistica/ Puc/ Indice.htm (last access: 12/04/2006).

City of Cagliari. 2004b. Plan for the historic center of Cagliari. Available on the Internet at http:// comuneca.tiscali - business.it/ urbanistica/ Urbanistica.php (last access: 12/04/2006).

Curry, B. 2003. Rough Sets: Current and Future Developments. *Expert Systems*, 20(5): 247-250.

Gardner, M., M. Lalmas, and I. Ruthven. 1996. Effect on noise reduction using rough set theory on the accuracy of K-NN classification of radiological reports. *Technical report* 1996-31. Glasgow, United Kingdom: Department of Computing Science, University of Glasgow.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth. 1996. Machine Learning and Data Mining. *AI Magazine*, 17(3): 37-54.

ISTAT (Italian Institute of Statistics). 2001. Censimento della popolazione e delle abitazioni [Population and Houses Census]. Rome, Italy: Istituto Poligrafico dello Stato.

Lee, S. and G. Vachtsevanos. 2002. An application of rough set theory to defect detection of automotive glass. *Mathematics and Computers in Simulation* 60: 225-231.

Longley, P.A., M.F. Goodchild, D.J. Maguire and D.W. Rhind. 2001. *Geographic Information. System and Science*. Chichester, United Kingdom: John Wiley & Sons.

Øhrn, A. 1999. *Discernibility and Rough Sets in Medicine: Tools and Applications*. Trondheim, Norway: Department of Computer and Information Science, Norwegian University of Science and Technology.

Øhrn, A. 2001. *Rosetta Technical Reference Manual*. Trondheim, Norway: Department of Computer and Information Science, Norwegian University of Science and Technology.

Murgante, B. and A. Sansone. 2005. L'uso della densità di Kernel e del rough set per la definizione degli ambiti periurbani [The use of the Kernel density and of the rough set analysis to define periurban areas]. Paper presented at the Seminar "Urbing – Giornate di lavoro sull'alta formazione sui temi della pianificazione, gestione e governo del territorio" [Urbing – Workshop on high education concerning regional and urban planning, management, governance]. Potenza, Italy. April 29-30.

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

Pawlak, Z. 1991. *Rough sets – Theoretical aspects of reasoning about data*. Boston, MA, United States, and London, United Kingdom: Kluwer Academic Publishers.

Pawlak, Z., and R. Slowinski. 1994. Decision analysis using rough sets. *International Transactions in Operational Research* 1(1): 107-114.

Pawlak, Z. 2001. Rough set theory. *KI* 15(3): 38-39.

Peña, J.M., S. Létourneau, and Fazel Famili. 1999. Application of rough sets algorithms to prediction of aircraft component failure. In *Advances in intelligent data analysis: Third international symposium, IDA-99, Amsterdam, The Netherlands, August 1999. Proceedings*, edited by D.J. Hand, J.N. Kok, and M.R. Berthold. Heidelberg, Germany: Springer-Verlag GmbH. 473-486.

Regional Administration of Sardinia. 2003. Sardinian Regional Geographic Database. Available on the Internet at http:// www.sitos.regione.sardegna.it/ entilocali/ pianificazione_territoriale/ html/ 008.htm (last access: 12/04/2006).

Yang, Z., T. Li, F. Jin, and S. Xu. 2002. Rough set in data mining. Lectures given during the course on "Data mining concepts and techniques" of Anita Wasilewska (course # CSE544) at the Stony Brook State University of New York, United States.