

The use of spatial statistics to analyze the periurban belt

Beniamino Murgante, Giuseppe Las Casas, Maria Danese

DAPIT, Università degli Studi della Basilicata,

10, Via dell'ateneo lucano, 85100 Potenza, Italia

E-mail address: beniamino.murgante@unibas.it

Abstract

The expression “periurban area”, despite its large use, does not have a clear and unambiguous definition. Different points of view generate more barriers in determining the exact edge of these zones. These various approaches are due to the complexity of the phenomenon to be analyzed and to the huge variety of territorial contexts in which it may reveal. The phenomenon is characterized by an urban growth with soil consumption generating loss of competitiveness for agricultural activities. Some experiences, in a spontaneous way, take into account only proximity to urban areas. It is obvious that contiguity condition alone is not sufficient to define such a complex phenomenon. The aim of this paper is to define more precise rules in order to describe the periurban phenomenon, using techniques of spatial statistic and point pattern analysis. This approach has been tested in the case of study of Potenza Municipality. This area, about 18,000 ha wide, is located about 900 m a.s.l. and is characterized by a relevant presence of forests, very steep slopes and a large number of areas with geological instability phenomena. Our interest in this area especially comes after the earthquake of 1980, when a huge migration of inhabitants began towards the countryside around the Town of Potenza. It was so intense that it deeply altered the urban morphology and modified the traditional physical functional relationships between the town and the rural territory.

INTRODUCTION

In most cases urban planners pay attention to urban or metropolitan areas without considering rural and periurban zones. Only in recent times the expression “periurban area” has been frequently used in planning documents. In planning literature, the periurban phenomenon has been defined in different ways; Thomas Le Jeannic (1997) describes the population displacement as the need to escape from the dense city in order to have more space and a better environment. Also, growth of periurban belt is due to high costs of flats, the need of individual dwellings and land rents (Guerois and Pumain 2001). Since 1980s it is less and less possible to distinguish town from country, denying the concept of two separate entities which was for many years one of the cornerstones of spatial planning (Hidding, *et al.* 2000; Van Den Berg and Wintjes 2000). Rural areas are more urbanized and periurban belt have increased the number of inhabitants with an uncontrolled growth. At the same time, urban areas have lost resident population gaining population in transit (Alberti, *et al.* 1994) because of the activities concentration in urban areas. All these situations produce a huge commuting phenomenon (Cavailhès, *et al.* 2004). The main feature of this trend is a low density of urbanization which spreads in all directions (Camagni, *et al.* 1998). Growth of these areas is strictly related to urban sprawl, generating negative repercussions to agricultural activities. A great amount of roads have been built to improve dwellings accessibility and car is the only means of transport (Camagni, *et al.* 2002).

This is an opposite tendency compared to the period after the Second World War, when urban planners used statistical methods to give a dimension of the migratory phenomenon towards towns. Urban sprawl is so complex to analyze, that classical statistics are not enough for a complete understanding of the phenomenon. Settlement location in zones surrounding urban areas takes into account environmental features, accessibility, agricultural losses of productivity. In order to achieve a more complete analysis it is important to analyze each phenomenon with its spatial location, so that it

is possible to consider the concentration of some events in some areas and their possible interactions. Geostatistics can be useful in order to study this problem with an innovative approach compared to the classic socio-economic techniques. This method allows an analysis which may determine the actual trend in one region. This technique has been applied in Potenza Municipality, where a migratory phenomenon began from urban to rural areas after a strong earthquake occurred in 1980.

All the informative layers have been combined with a land suitability procedure in order to define a periurban fringe with a certain precision.

SPATIAL STATISTICS TECHNIQUES

The main aim of spatial analysis is a better understanding of spatial phenomena aggregations and their spatial relationship. Spatial statistical analyses are techniques which use statistical methods in order to determine if data show the same behaviour of the statistical model. Data are treated as random variables. The *events* are spatial occurrences of the considered phenomenon, while *points* are each other arbitrary locations. Each event has a set of attributes describing the nature of the event. *Intensity* and *weight* are the most important attributes; the first is a measure identifying the event strength, the second is defined by the analyst who assigns a parameter in order to define if an event is more or less important according to some criteria. Spatial statistics techniques can be grouped in three main categories: *Point Pattern Analysis*, *Spatially Continuous Data Analysis* and *Area Data Analysis*.

The first group considers the distribution of point data in the space. They can follow three different criteria:

- random distribution: the position of each point is independent of the others points;
- regular distribution: points have an uniform spatial distribution;
- clustered distribution: points are concentrated in some building clusters.

The second group takes into account the spatial location and the attributes associated to points, which represent discrete measures of a continuous phenomenon. The third group analyzes aggregated data which can vary continuously through space and can be represented as point locations. This analysis aims to identify the relationships among variables and *spatial autocorrelation*. If some clusters are found in some regions and a positive spatial autocorrelation is verified during the analysis, it can describe an attraction among points. The case of negative spatial autocorrelation happens when deep differences exist in their properties, despite the closeness among events. It is impossible to define clusters of the same property in some areas; a sort of repulsion occurs. Null autocorrelation arises when no effects are surveyed in locations and properties. Null autocorrelation can be defined as the case in which events have a random distribution over the study area (O'Sullivan and Unwin, 2002). Essentially, the autocorrelation concept is complementary to independence: events of a distribution can be independent if any kind of spatial relationship exists among them.

Spatial distribution can be affected by two factors:

- first order effect, when it depends on the number of events located in one region;
- second order effect, when it depends on the interaction among events.

If these two definitions seem more clear, it isn't as much clear as the recognition of these effects over the space.

Kernel density

Kernel density is one of the point pattern analysis techniques, where input data are point themes and outputs are grids. While simple density computes the number of events included in a cell grid considering intensity as an attribute, *kernel density* takes into account a mobile three-dimensional surface which visits each point. The output grid classifies the event S_i according to its distance from the point S_i , which is the centre of the ellipse generated from the intersection between the surface and the plane containing the events (Bailey and Gatrell, 1995). The *influence function* defines the influence of a point on its neighbourhood. The sum of the *influence functions* of each point can be calculated by means of the *density function*, defined by:

$$\text{Equation 1} \quad \lambda(L) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{L-L_i}{\tau}\right)$$

where:

- λ is the distribution intensity of points, measured in L;
- L_i is the event i ;
- K is the kernel function;
- τ is the bandwidth.

The first factor influencing density values is bandwidth: if τ is too big the value of λ is closer to simple density; if τ is too small the surface does not capture the phenomenon. The second factor influencing density values is cell size like in every grid analysis.

Straight Line Distance

The straight line distance is a function measuring the distance between each cell and the nearer source. The source can be in vector or grid format. In the case of grid format some cells will contain information about the source and some others will not, while in the case of a vector theme it will be necessary a previous transformation in grid before determining the distance.

The output of straight line distance is in grid format and the distance is measured between the barycentres of the cells. Also, in this case it is important to estimate some factors such as the maximum distance within which one has to assess measures and sizes of cells.

Moran index

Moran index (Moran, 1948) allows to transform a simple correlation into a spatial one. This index takes into account the number of events occurring in a certain zone and their intensity. It is a measure of the first order property and can be defined by the following equation:

$$\text{Equation 2} \quad I = \frac{N \sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j w_{ij}) \sum_i (X_i - \bar{X})^2}$$

where:

- N is the number of events;
- X_i and X_j are intensity values in the points i and j (with $i \neq j$), respectively;
- \bar{X} is the average of variables;
- $\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})$ is the covariance multiplied by an element of the weight matrix. If X_i and X_j are both upper or lower than the mean, this term will be positive, if the two terms are in opposite positions compared to the mean the product will be negative;
- w_{ij} is an element of the weight matrix which depends on the contiguity of events. This matrix is strictly connected to the adjacency matrix.

There are two methods to determine w_{ij} : the “Inverse Distance” and the “Fixed Distance Band”. In the first method, weights vary in inverse relation to the distance among events:

$$w_{ij} = d_{ij}^z$$

where z is a number smaller than 0.

The second method defines a critical distance beyond which two events will never be adjacent. If the areas to which i and j belong are contiguous, w_{ij} will be equal to 1, otherwise w_{ij} will be equal to 0. *Moran index* I can have values included between -1 and 1.

If the term is high, autocorrelation is positive, otherwise it is negative. Moran index vanishes in very rare cases, but usually the convergence is towards the theoretical mean value $E(I)$, where each value is independent from the others.

$$E(I) = -\frac{1}{N-1}$$

If $I < E(I)$, then the autocorrelation is negative, if $I > E(I)$ the autocorrelation is positive.

The significance of Moran index can be evaluated by means of a standardized variable $z(I)$ defined as:

$$z(I) = \frac{I - E(I)}{S_{E(I)}}$$

where $S_{E(I)}$ is the standard deviation from the theoretical mean value $E(I)$.

Local Indicators of Spatial Association: G function by Getis & Ord and Local Indicator of Spatial Association (LISA)

Both LISA and G function take into account disaggregated measures of autocorrelation, considering the similitude or the difference of some zones. These indexes measure the number of events with homogenous features included within a distance d , located for each distribution event. This distance represents the extension within which clusters are produced for particularly high or low intensity values.

The Local Indicator of Spatial Association (Anselin, 1995) is defined as:

$$\text{Equation 3} \quad I_i = \frac{(X_i - \bar{X})}{S_X^2} \sum_{j=1}^N (w_{ij} (X_j - \bar{X}))$$

where symbols are the same used in Moran's I , except for S_X^2 which is the variance.

The function by Getis & Ord (1992) is represented by the following equation:

$$\text{Equation 4} \quad G_i(d) = \frac{\sum_{i=1}^n w_i(d) x_i - \bar{x}_i \sum_{i=1}^n w_i(d)}{S(i) \sqrt{\left[(N-1) \sum_{i=1}^n w_i(d) - \left(\sum_{i=1}^n w_i(d) \right)^2 \right] / N - 2}}$$

which is very similar to Moran index, except for $w_{ij}(d)$ which, in this case, represents a weight which varies according to distance.

THE STUDY CASE

These techniques have been applied in Potenza municipality, located in the southern Apennines area of Italy, with a very low residential density. In the last decades a migratory process began from the urban areas to the rural ones. The causes of this phenomenon are various (social, economic,

cultural etc.), so that it is not easy to define how the spatial distribution of events can vary. On this purpose, it can be useful to apply spatial statistic techniques to define criteria concerning suitability in building new settlements, considering particular tendencies that some specific areas already manifest.

A first factor considered in this study is density. Periurban area is characterized by a spread of settlements with extensive features, compared to the urban area which has a greater density. Lower density is the first condition distinguishing periurban areas from urban ones.

Rural sites have a strong connection with agricultural activities and the relationship with the urban area is weak. It is also necessary to establish a lower threshold which can distinguish periurban areas from rural ones.

In order to calculate density, all the polygons representing buildings have been converted in points which are the events to take into account in point pattern analysis. The ratio between the number of flats and the number of buildings has been calculated from census data; this value has been considered as the intensity of events. Figure 1 compares the density of scattered settlements between 1987 and 2004 and it shows the huge growth of urban sprawl.

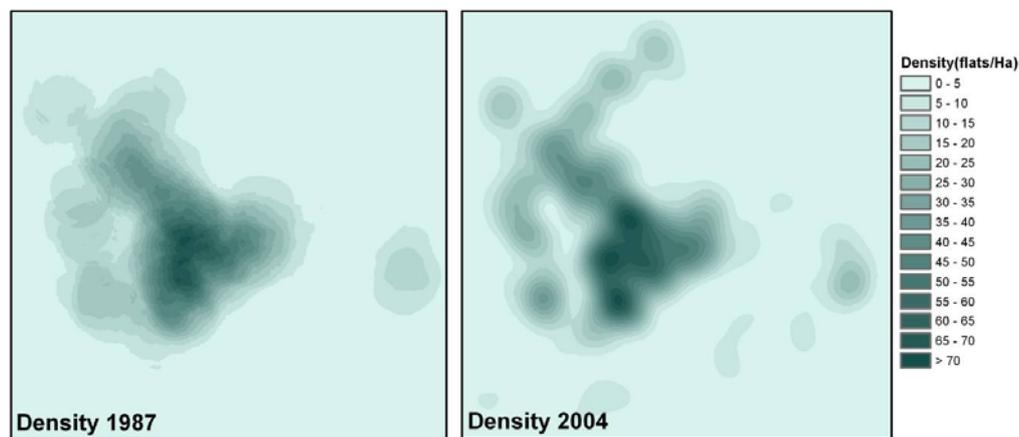


Figure 1: Density of scattered settlements in 1987 and 2004 (flats/hectare).

In the case of study, a value of bandwidth of 400 m and a cell size of the grid of 10 m have been used.

A first rough analysis of periurban fringe takes into account zones with a low density expansion including areas with values of kernel density included between 1 and 18 flats/ha (fig.2).

Orography and accessibility define the second factor, which consists of the distance from infrastructures because urban growth is more concentrated along the main line of road network.

In order to locate areas with a good accessibility, distance from infrastructures has been defined so that it represents the tendency. Straight Line Distance identified areas with a distance from the main infrastructures within 200 meters.

The third factor is the spatial autocorrelation which has been analyzed considering Moran Index, G function as developed by Getis & Ord and Local Indicator of Spatial Association (LISA).

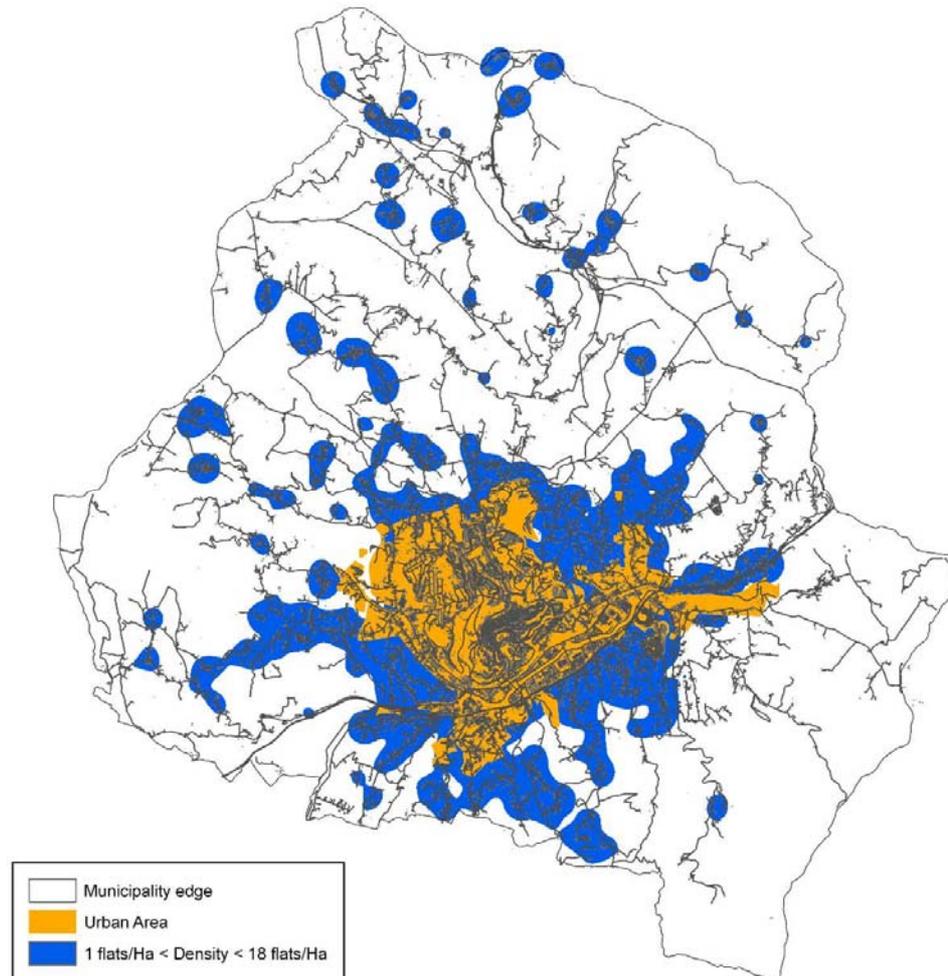


Figure 2: Areas which have Kernel density included between 1 and 18 flats/hectare.

In this case intensity of events is obtained as the ratio between number of inhabitants and number of buildings in each census zone. Moran index is able to specify if an event is clustered, scattered or with a random distribution. It has been calculated by means of the inverse distance method considering data in two different periods, 1987 and 2004, to evaluate the variation of scattered rate of settlements. The following values have been achieved:

- Moran Index at 1987: $I_{1987} = 0.0698$;
- Moran Index at 2004: $I_{2004} = 0.0722$.

These two indexes show a low autocorrelation in both cases, and the second one is higher than the first one. These data can be interpreted as growth of settlements concentrated in some particular zones. The next step of our study was to calculate the contiguity belt considered as the area where the phenomenon grows homogenously and where it will intensify in the future. Moran index depends from the distances among points; it is possible to calculate a distance value which produces an index I

with the maximum level of correlation among events, by maximizing the deviation z . This value has been calculated in 1600 m and it has been used as an input parameter in LISA and then in Getis & Ord function determining zones where events are autocorrelated.

A LISA positive value indicates a positive autocorrelation; obviously a negative autocorrelation corresponds to a negative value .

For the periurban fringe it is important to pay attention to the medium-low level of intensity, so the classes in table 1 have been considered.

Class	Autocorrelation	LISA
no correlation	Negative autocorrelation	-106,9 ÷ 0
1 low	Positive autocorrelation among lower bounds	0 ÷ 14
2 low	Positive autocorrelation among low bounds	14 ÷ 28
3 low	Positive autocorrelation among medium-low bounds	28 ÷ 54
1 high	Positive autocorrelation among high bounds	54 ÷ 84.7

Table 1

In Getis & Ord function, highest and lowest values of G mean highest and lowest values of phenomenon intensity.

The classes in table 2 have been considered.

Class	Autocorrelation	Intensity value X_i (inhabitants/buildings)	G'
no correlation	Negative autocorrelation	$X_i \leq 18$ $X_i \geq 18$	-1.3 ÷ - 2 -6.3 ÷ 1
1 low	Positive autocorrelation among lower bounds	$X_i \leq 18$	-1.3 ÷ - 2
2 low	Positive autocorrelation among low bounds	$X_i \leq 18$	-2 ÷ - 4
3 low	Positive autocorrelation among medium-low bounds	$X_i \leq 18$	-4 ÷ - 6.3
1 high	Positive autocorrelation among high bounds	$X_i \geq 18$	1 ÷ 11.9

Table 2

Next figure (Fig. 3) compares results, showing the similitude of areas with positive autocorrelation achieved with both indicators. In our study case Getis & Ord function fits the phenomenon better because it is more sensible to intensity changes allowing more accurate classification.

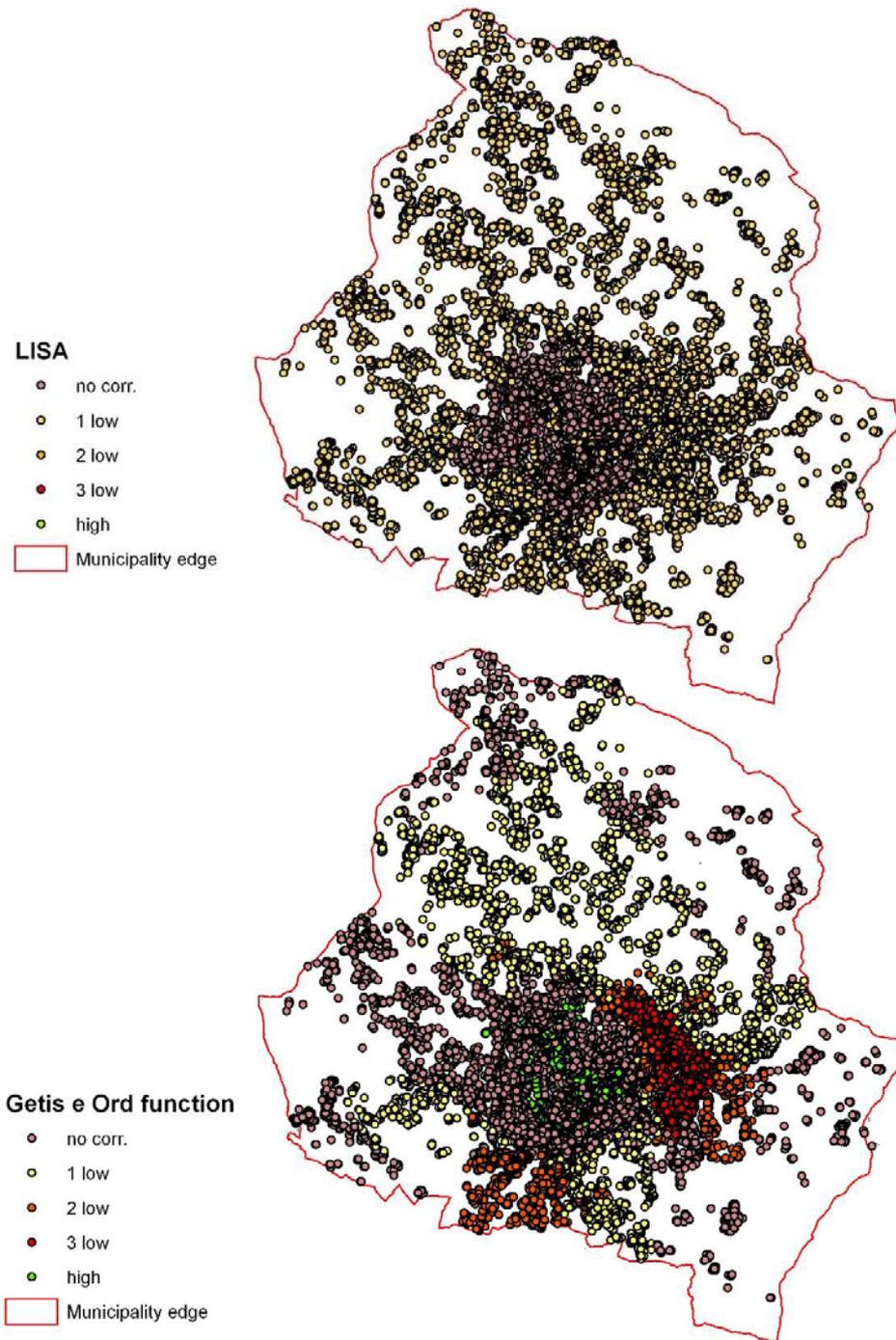


Figure 3: Clusters location.

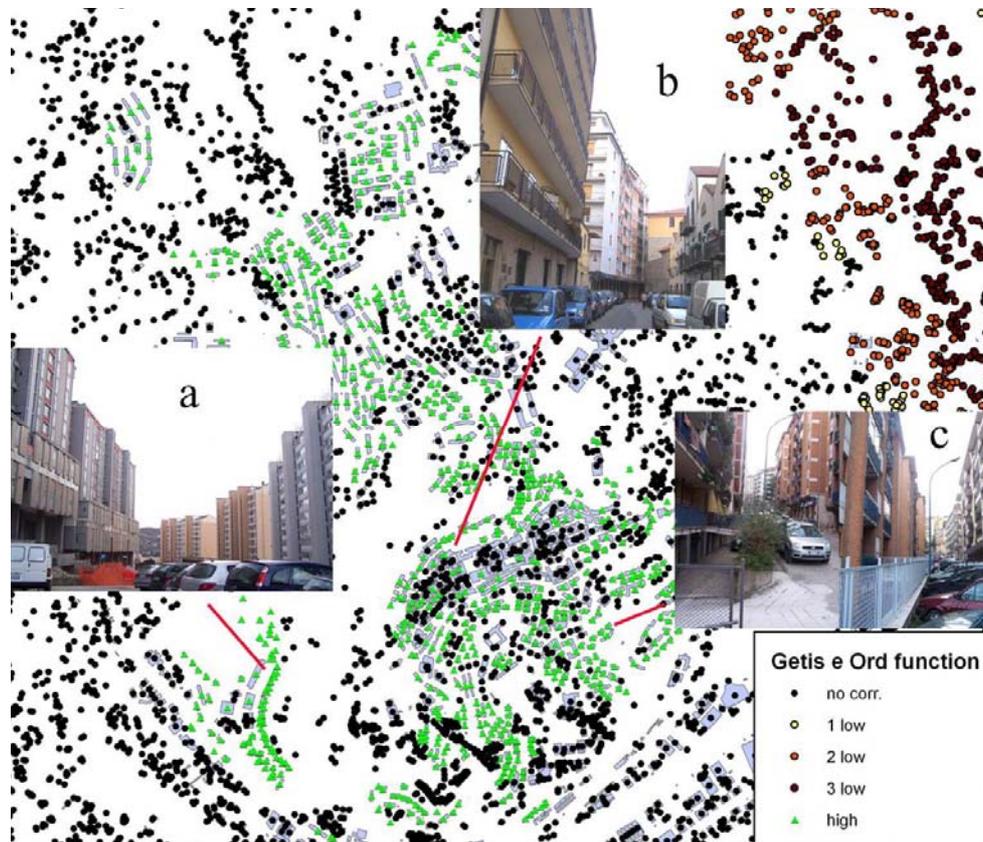


Figure 4: Detailed map with pictures which show autocorrelation difference.

Figure 4 shows in a more detailed map how deep are analysis produced with Getis & Ord function. Opposite values can occur in contiguous areas.

Picture a shows how the highest values of autocorrelation correspond to the highest buildings of the town. Picture b highlights an abrupt transition, in a few metres, from no correlation to high correlation passing from ancient low buildings to high concrete buildings. Picture c shows how very elevated values of autocorrelation correspond to high buildings separate by narrow streets.

RESULTS AND FINAL DISCUSSION

Autocorrelation phenomena included in medium medium-low values have been interpolated thus generating polygons which represent the contiguity belt. These polygons represent the second level of suitability. It is composed by the inclusion rules considered in land suitability procedures reduced considering the global and local measures of autocorrelation. It is obvious that kernel density (Fig. 1 and 2) is a rough measure which needs a deeper analysis. Moran index and Getis & Ord function give a further interpretation of phenomena considering contiguity not in all directions but only in some zones.

The exclusion rules in figure 5 have been considered in the present study: areas included within a distance of 150 m from rivers, streams and springs, slopes higher than 35%, Nature 2000 sites, hydro-

geological risk zones, areas higher than 1200 m a.s.l., landslides, areas close to railways and road networks.

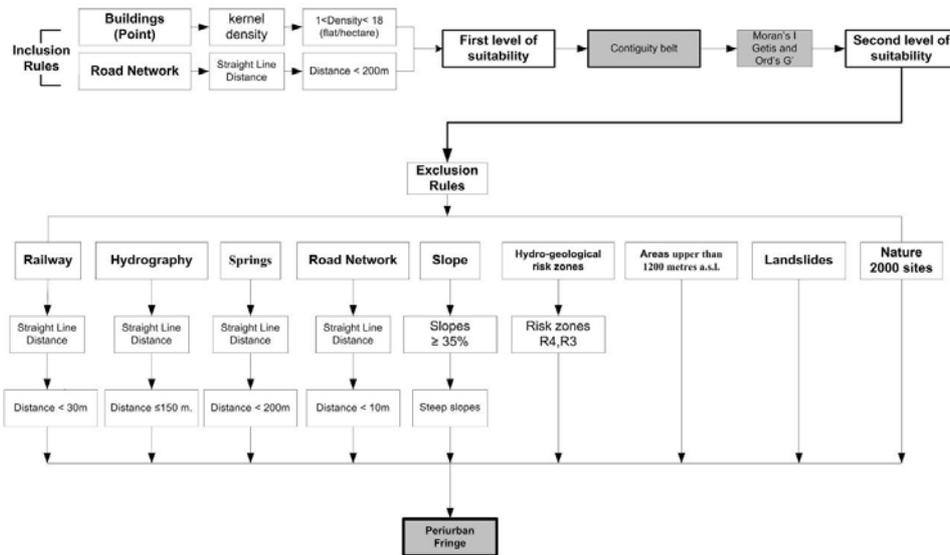


Figure 5: Scheme of the land suitability procedure for the location of Peri-urban fringe.

Figure 5 shows the flow chart of the land suitability procedure for the location of Peri-urban fringe. All these rules have been combined using map algebra techniques. Figure 6 quantifies the reduction of suitable areas achieved after the procedure.

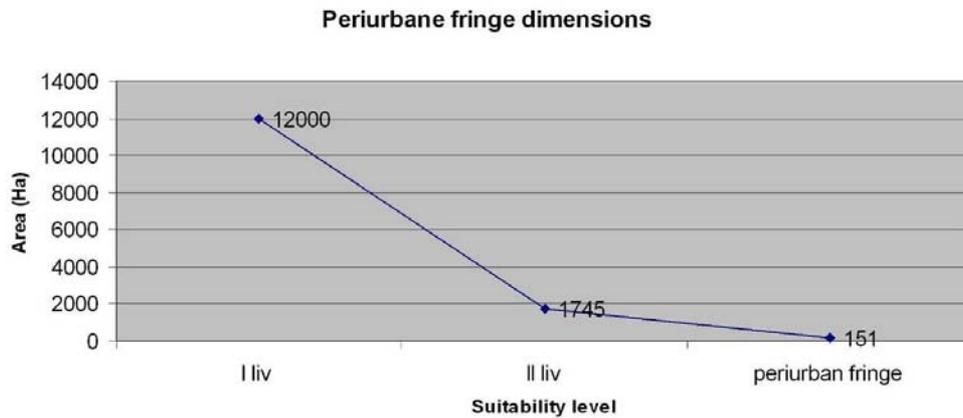


Figure 6: Size of suitable areas.

The results are illustrated as geographic components in figure 7. Location of the contiguity belt is determined by the highway, which determines two gates for the town. In these areas urban sprawl is more intensified, particularly in the eastern part where the other road, which connect the industrial

areas of Potenza with FIAT factory, amplifies the phenomenon. Steep slope obstructs urban growth in other zones. Periurban fringe (magenta coloured in figure 7) considers contiguity belt after the exclusion rules and represents areas suitable for the location of new settlements or for intensifying the existing ones.

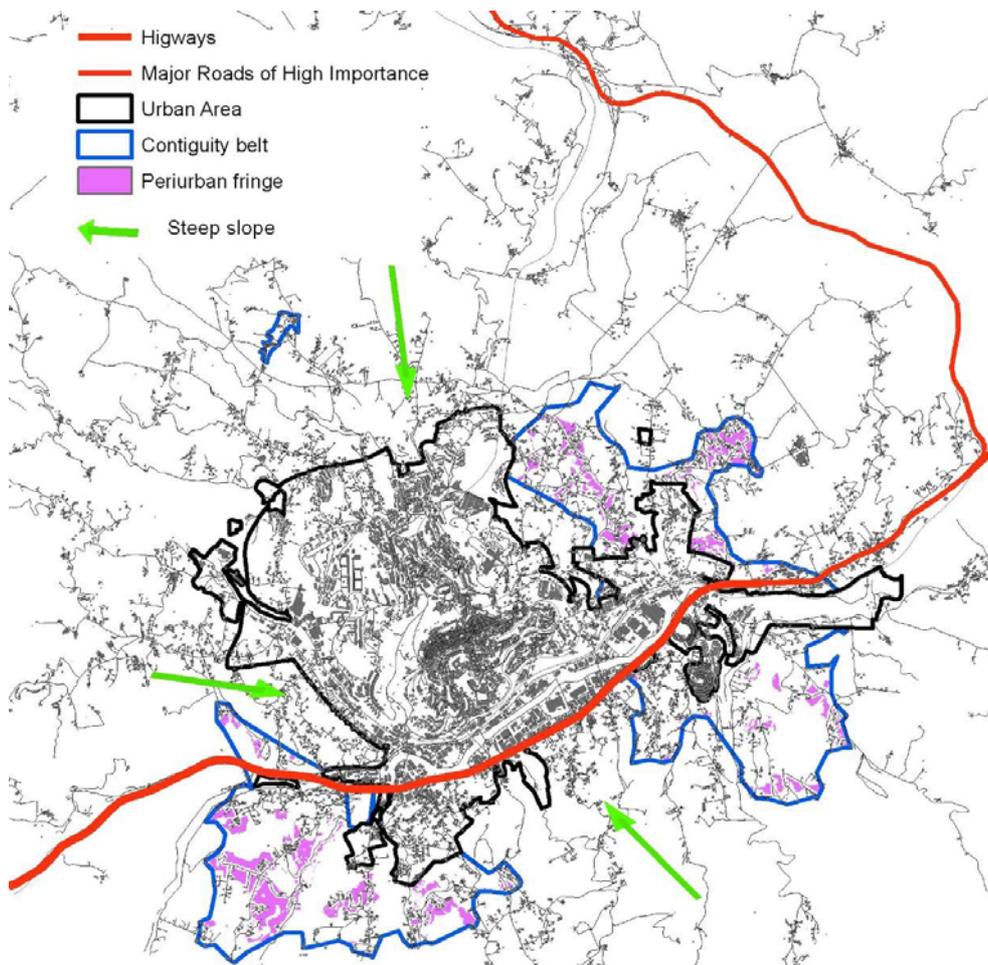


Figure 7: Periurban fringe after the land suitability procedure (magenta).

After the theorization by Waldo Tobler (1970), the first law of geography is reported here: “Everything is related to everything else, but near things are more related than distant things”. More experience exists of the use of spatial statistics in geographical analysis; for instance Kernel density has been applied for the location of epidemics (Gatrell *et al.* 1995) and studies on spreading of city services (Borruso and Schoier, 2004), while these techniques have not been used enough in the field of territorial planning. In this paper several kinds of spatial statistic functions have been applied for a deeper knowledge of territory and to give urban planners a better support for planning choices.

BIBLIOGRAPHICAL REFERENCES

- Alberti M., Solera G., Tssetsi V., (1994), *La città sostenibile*, F. Angeli,
- Anselin, L., (1995), "Local Indicators of Spatial Association – LISA." *Geographical Analysis* 27:93-115.
- Bailey T. C., Gatrell A. C. (1995), *Interactive spatial data analysis*, Prentice Hall
- Borruso G., Schoier G. (2004), "Density Analysis on Large Geographical Databases. Search for an Index of Centrality of Services at Urban Scale", in Gavrilova M. L., Gervasi O., Kumar V., Laganà A., Mun Y. and Tan K. J. (eds.), *Lecture Note in Computer Scienze Springer-Verlag, Berlin*,
- Camagni, R., Capello R.; Nijkamp P., (1998), "Towards sustainable city policy: an economy-environment technology nexus" *Ecological Economics* Volume: 24, Issue: 1, Elsevier
- Camagni, R., Gibelli M. C., Rigamonti P. (2002), "Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion" *Ecological Economics* Volume 40, Issue: 2, Elsevier
- Cavailhès J., Peeters D., Sekeris E., Thisse J.,(2004), "The periurban city: why to live between the suburbs and the countryside", *Regional Science and Urban Economics*, Volume 34, Issue 6, Elsevier
- Gatrell A. C. , Bailey T. C., Diggle P. J., Rowlingson B. S., (1995) "Spatial point pattern analysis and its application in geographical epidemiology" *Transaction of institute of British Geographer*
- Getis, A. and Ord, J. K. 1992 The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, 189-206.
- Guerois M., Pumain D., (2001), *Urban sprawl in France (1950-2000)*, Franco Angeli
- Hidding M., Needham B., Wisserhof, J., (2000), "Discourses of town and country" *Landscape and Urban Planning* Volume: 48, Issue: 3-4, Elsevier
- Moran, P. (1948), "The interpretation of statistical maps", *Journal of the Royal Statistical Society*, n.10
- O'Sullivan D., Unwin D., (2002), "Geographic Information Analysis" John Wiley & Sons
- Van Den Berg L., Wintjes A. (2000), New 'rural lifestyle estates' in The Netherlands, *Landscape and Urban Planning*, Volume: 48, Issue: 3-4 Elsevier