Cent. Eur. J. Geosci. • 6(1) • 2014 • 42-55
DOI: 10.2478/s13533-012-0160-5

VERSITA

Central European Journal of **Geosciences**

Predictive modeling for preventive Archaeology: overview and case study

Research article

Maria Danese[1]* , Nicola Masini[1], Marilisa Biscione[1], Rosa Lasaponara[2]

1 IBAM, CNR, Contr.da S. Loja,
Tito Scalo (Pz), 85050, Italy; +39 0971 427309

2 IMAA, CNR, Contr.da S. Loja,
Tito Scalo (Pz), 85050, Italy

Received 5 August 2013; accepted 31 December 2013

Abstract: The use of GIS and Spatial Analysis for predictive models is an important topic in preventive archaeology. Both of these tools play an important role in the Support Decision System (SDS) for archaeological research and for providing information useful to reduce archaeological risk. Over the years, a number of predictive models in the GIS environment have been developed and proposed. The existing models substantially differ from each other in methodological approaches and parameters used for performing the analysis. Until now, only few works consider spatial autocorrelation, which can provide more effective results. This paper provides a brief review of the existing predictive models, and then proposes a new methodological approach, applied to the neolithic sites in the Apulian Tavoliere (Southern Italy), that combines traditional techniques with methods that allow us to include spatial autocorrelation analysis to take into account the spatial relationships among the diverse sites.

Keywords: Apulian Neolithic • Autocorrelation • Archaeology, Predictive Models • Spatial analysis

1. Introduction

Predictive models are useful tools for archaeological research. They can provide an important decision support system to obtain useful information for defining survey priority and facilitating new discoveries; saving time and money, especially in large areas. Moreover, predictive models can also contribute to the preservation of archaeological areas and features, and witnesses of the human past, providing information useful for reducing archaeological risk linked both to anthropic and natural risk factors. Archaeological risk can be defined as the product of the hazard, the vulnerability and exposure [2]. The hazard expresses the probability that an event causes damages, due to, for example, anthropic hazard such as urban sprawl [3], large-scale infrastructure [4] looting linked to the illicit trade of antiquities [5], or natural risk factors, such as landslides, fire, earthquakes, etc [6]. The exposure expresses the value of the element at risk, which is particularly difficult to evaluate in an objective way in the case of archaeological heritage. The vulnerability is connected to elements at risk herein represented by archaeological sites. According to most of Italian regional planning laws, archaeological risk maps are important instruments for public authorities to prevent and mitigate

*E-mail: (m.danese-n.masini-m.biscione)@ibam.cnr.it

risk factors, on the basis of a documentary knowledge in both urban and territorial planning. These maps have a probabilistic nature, depending on both hazard and vulnerability components, that is the site distribution on the territory. Unfortunately in the case of archaeological remains, the complete and real spatial distribution of archaeological sites can not be known and, in these cases, the use of predictive models can provide useful information for obtaining archaeological sensibility maps as results of a compromise between the reduction of costs and effectiveness of results. Predictive models have been introduced by different authors in the last thirty years [7], [8], [9] for landscape archaeological studies, which focus on the relationships between man and natural environment [10]. According to [11], landscape is the backdrop that affects limits or determines cultural forms and it is considered to be composed of several phenomena. Each phenomenon is an indicator of cultural transformations in the investigated region [12] and is considered in predictive models as an including or excluding factor for the presence of archaeological sites of a particular historical period or civilization. The work flow of predictive models starts from the study of known archaeological sites including their environmental and social characteristics. Known sites are studied to understand the factors which influenced their position in space, and consequently, to find settlement rules inside the model adopted to make prediction [13]. In this process, new technologies, such as GIS and remote sensing can be profitably used. Remote sensing and GIS have been becoming widely used in recent years thanks to data availability in conjunction with the development of user friendly softwares and tools. GIS environments allow us to find and overlay the model parameters, whereas remote sensing provides a useful data source for facilitating the detection of archaeological features and sites. The first main aspect of predictive modelling is the research and the use of including and excluding factors. In literature this is pursued in two main ways, deductive, inductive. With deductive approaches, location rules derive from a theoretical approach and from the consequent theoretical knowledge of archaeological sites. These elements provide settlement patterns and land use criteria adopted in the past. By using an inductive approach, rules inserted in the predictive model are extracted on the basis of a dataset, composed by an official database of well-known sites or from data picked up from archaeological surveys and remote sensing images [14], [15]. However, the choice of parameters is, linked to the historical period investigated and consequently in literature there is a great variation of factors used, even if the most important are always present. These are: (i) land use, (ii) elevation and (iii) the proximity to the water bodies [16]. Other parameters, such as environmental and social elements, are less frequently employed [17], [18], [19], as: 1) the sun exposure (that is the average of solar radiation hours received by a site in one day or in a season or during the year); 2) viewshed analysis and viewshed indexes, able to take in count the presence of vegetation and of sediments that increase the visible surface; 3) the distance from different elements such as the coast or reliefs; 4) euclidean distance or cost distance between sites. Another important aspect in predictive models is the choice of the method used to combine including and excluding factors. The most common methods in literature are statistics; in particular Markov's Chain [12], Dempster-Shafer's belief theory [20], logistic or linear regression, or multi-fractal analysis [21]. However, simpler methods such as map algebra functions are also used (See, just for some examples, [13], [16], [17]). One aspect that is very rarely treated is the presence of spatial autocorrelation [22] in the distribution of archaeological sites. In this paper we propose a predictive model that considers spatial autocorrelation analysis to include the distribution of archaeological sites. The paper is organized as follows: in the first section a review of existing predictive models is presented; in the second section spatial autocorrelation and spatial techniques that are currently used are described; in the third section the case study of Neolithic sites of the Tavoliere (in North-West of Apulia region, Southern Italy) is introduced and characterized in order to select parameters and to construct the model; and finally in the last section some final considerations and the validation of the model are presented.

2. Spatial analysis

In this paper two main methods have been used: map algebra and point pattern analysis. These are both techniques of spatial analysis. They permit investigation into the spatial distribution of phenomena, aggregation shapes and existing relationships between the phenomenon and multiple factors.

2.1. Map algebra

Map Algebra (MA) is a high level spatial modeling language, including base elements (operators), complex elements (functions) and formal language (instructions), together with elements needed to program and develop complex models [23]. Main MA functions are [24] local, focal and zonal functions. In local functions there are one or more input raster and the output raster comes from operations done cell-by-cell; in focal functions there is only one input raster and each pixel of the output raster is the result of operations calculated in a defined neighborhood of that input raster; finally in zonal functions the output raster comes from operations done on the values of one raster inside areas cover by another zonal raster. Surface analyses, useful to calculate morphological elements of a landscape, come from the combination of these groups of functions.

2.2. Point pattern analysis

In Point Pattern Analysis (PPA) the dataset under consideration is represented and analyzed as a point pattern. One of the aims of PPA is to find spatial autocorrelation, expressed by the first principle of geography: "Everything is related to everything else, but nearest things are more related than distant things" [25]. Studying the autocorrelation in one variable is important because it causes a variation of its spatial distribution; in fact under the hypothesis of complete spatial randomness, that is events completely independent, this distribution would be different. A variable is considered autocorrelated according to similarity and, at the same time, to distances between events. These properties can be affected by two factors: first order effects are related to region properties and are expressed by the spatial variation in the expected value (mean) of events; second order effects are related to local interactions between events and are expressed by the spatial variation of covariance. Due to first and second order effects, the spatial autocorrelation of a distribution can assume three different types of configuration [26] (fig. 1): a) positive spatial autocorrelation or attraction between events, when points are concentrated in some building clusters (clustered distribution); b) negative spatial autocorrelation or repulsion, when events are near in the space but they have different quantitative properties, so it is impossible to find homogeneous areas (uniform or regular distribution); and c) null autocorrelation (random distribution) when it is not possible to find any spatial effects, for what concerns both spatial location and properties of each event. Consequently, when null autocorrelation is verified, events exhibit a random spatial distribution; so in this case the complete randomness hypothesis is verified.

Usually, in predictive models spatial autocorrelation is not considered. To consider it we used the following two families of PPA: Nearest Neighbor method and Kernel Density Estimation.

2.2.1. Nearest neighbor method

Nearest Neighbor is the most common distance-based method. It provides information about the interaction among events at the local scale (second order property). Nearest-Neighbor Index (NNI) is defined by the following equation (1):

$$NNI = NNOD/NNED \tag{1}$$

The numerator of equation (1) represents the average distance between N events (Nearest Neighbor Observed Distance, NNOD) considering the minimum distance of each event from the nearest one, and it can be represented by equation (2):

$$NNOD = \frac{\sum_{i=1}^{n} d_{min}(s_1, s_j)}{n}, \tag{2}$$

where d_{min} (S_i,S_j) is the distance between each point and its nearest neighbour, and n is the number of points in the distribution. Finally, the denominator of equation (1) represents the nearest neighbour expected distance (NNED), based on a completely random distribution. It is expressed by the following formula (3):

$$NNED = 0.5\sqrt{\frac{A}{n}}. \tag{3}$$

where A is the area of the spatial domain.
NNI ¡ 1 means that NNOD¡NNED, so the point pattern shows a positive spatial autocorrelation, because events are closer to each other than expected.
NNI ¿ 1 means that NNOD¿NNED, so events are more scattered than expected.

2.2.2. Kernel density estimation

While classic density only computes the number of events included in a cell grid, kernel density estimation (KDE) is a moving three-dimensional function, weighting events within their sphere of influence according to their distance from the point at which intensity is being estimated [27]. The method is commonly used in a more general statistical context to obtain smooth estimates of univariate (or multivariate) probability densities from an observed sample of observations [28]. In each spatial point KDE is defined as:

$$\lambda(L) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{L - L_i}{\tau}\right), \tag{4}$$

where λ is the distribution intensity of points, Li is the event i, k is the kernel function and ? is the bandwidth. The main factor influencing density values is the 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, as in every grid analysis. An extension of KDE is KDEN, where the point estimation is evaluated

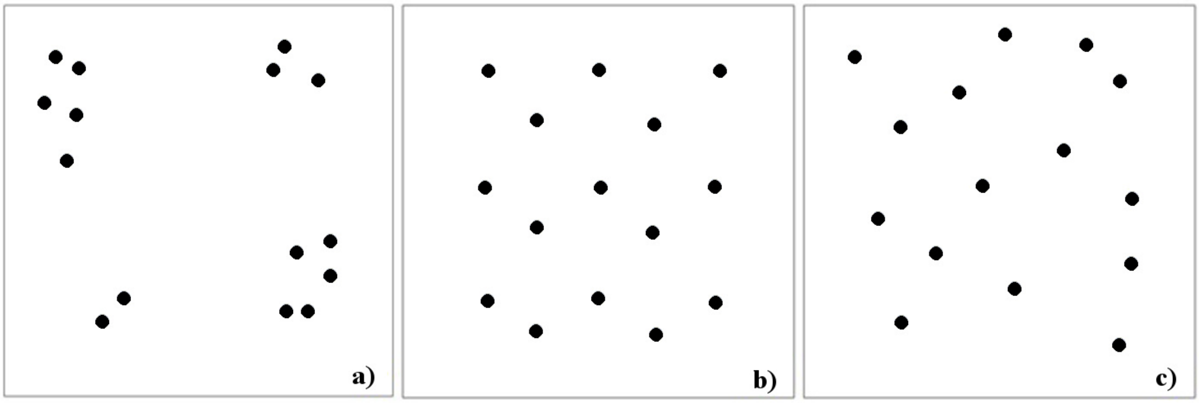

Figure 1. a) positive autocorrelation, b) negative autocorrelation, c) random distribution.

along a network. KDEN could be calculated in a very different way and in different application fields (see for example [29], [30], [31], [32], [33], [34], [35], [36]), but it has not yet been applied in the archaeological field. In this paper the method proposed in SANET software has been used [37], [38], [39].

3. Neolithic sites of the tavoliere (Apulia)

3.1. The case study and the working flow

The above explained methods have been applied to Apulian Tavoliere, an extended lowland plain covering an area of 4300 km2 in Apulia region (Southern Italy). It is limited to the east by the Adriatic Sea, to the northwest by the Gargano Massif, and to the south by the Murge plateau. The investigated area is involved in an intensive mechanized agricultural activity and characterized by a long human presence as attested by a rich archaeological heritage. [40]. On the basis of the aerial photographs taken by Bradford during his 'Apulian Expedition' [41] this area has interested archaeological research for 50 years. In the 80s there were significantly important publications in conference proceedings about Daunia's Prehistory-Protohistory-History [42], [43], [44], [45] and also monographs on investigations on some specific sites [46]. The area is furthermore the ideal terrain for applications and palaeo-environmental research in relation to the settlement choices for both the exceptional documentation on the distribution of prehistoric settlements and the sparse population; from the reconstruction of the ancient landscape observing the geology, hydrogeology and sedimentology [47] up to the palaeo-economy

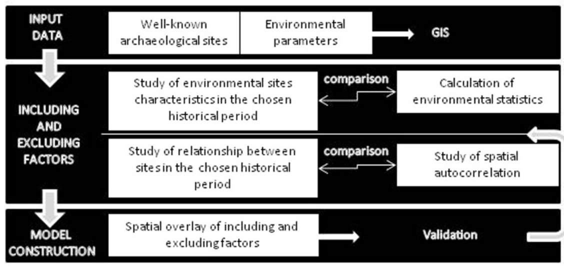

Figure 2. The working flow chart for the predictive model of the Apulian Neolithic sites

research as regards settlement choices and conformation of the territory [48], [49], [50],[51]. Other studies on demographic estimates, social organization and settlement [52], [53] have been conducted starting from the classes of size of sites [40], [54]. The working flow pursued the following steps (fig. 2): 1) Creation of a GIS with the dataset of already known archaeological sites and with the environmental parameters characterizing the study region; 2) Research of including and excluding factors for the Neolithic Apulia, through the calculation of spatial statistics and the use of PPA methods and the comparison of results obtained with the study of environmental and social characteristics of sites in the chosen historical period; and 3) Overlay of including and excluding factors with the formalization and the following validation of the model. The result is a map of survey priorities that contains areas that describe the probability of finding an element of archaeological interest.
For the creation of the GIS, archaeological features (fig. 3) that were used for this case study come from scientific works, such as: - proceedings of national conferences on Daunia's Prehistory-Protohistory-History [55], [43], [44],

[45] where results of traditional methods, such as field-walking, archaeological excavations and historical studies are reported; - recent works such as [14], where sites are extracted from both aerial and satellite images and further investigated by high-resolution magnetic survey, which provides a detailed identification of buried remains; - other scientific papers on the Neolithic in Tavoliere [40], [41], [42], [43], [56], [57]; - the important study of Bradford [41], [58], [59] based on aerial photographs taken during WW2. Owing to the extent of the area and the diachronicity at the basis of this work, the sources are clearly different and in particular the data are often heterogeneous, as we shall see later; in particular as regards studies of the paleo environment. However, the main aim of our prediction model is to provide a service to planners rather than to archaeologists; the latter mainly investigate the areas belonging to each site while the first need to know where they exist, generically, or possible sites. Moreover only by digging would it be possible to clarify the presence, nature and chronology of the archaeological remains and, additionally, the absence of data does not ensure the absence of the site. In the light of these considerations, we note the lack of homogeneity of data but for our aims the importantace is in recording their presence.

3.2. Characteristics of Neolithic sites

With the aim to find including and excluding factors the characteristics of Apulian Neolithic sites have been considered. They can be summarized as follows [4]:

1. Shape and dimension of sites: the shape is usually circular or semi-circular, marked by defensive structures (moat), draining elements or land demarcation [42], [60] around villages, spaces for animals and part of cultivated lands. Dimensions range from little villages (less than 1ha) to medium villages (1÷4ha) up to big villages (for example Passo di Corvo which is bigger than 28ha) [54].

2. Distances between sites: distances are often modular, but we should discern between sites contemporary or not [61].

3. Geological, morphological and paleo environmental characteristics: there is a preference for alluvial plain because it is advantageous for agriculture; especially where surface aquifers are located [54], [45], [62]. The agricultural vocation of the Tavoliere is well-known and palaeobotanical investigations highlight the importance and complexity of this economic activity even in the Neolithic. The climatic optimum of Atlantic phase - high temperatures and increased precipitation - means that the production system is closely linked to the environment [46]. The areas on the edge of the terraces down to the river, and less arid plains further inland, would provide better-drained soils that are easier for agriculture that requires a digging stick or hoe. For this kind of primitive agriculture the best lands were those of the terraces often with the presence of "crusta" covered by humus levels, which is essential for crops. The coincidence of a shallow aquifer with "crusta" surface was recently emphasized; in this case, this training is so rich in bases adapted to the cultivation of cereal crops. Terracing most high, bordered by rivers, was characterized instead by Mediterranean maquis [64]. In particular we can summarize a preference with the proximity to: i) mountain areas (piedmonts) for summer pasture , ii) rivers for water supply, fishing and harvesting, and iii) caves for lithic industries [61]. Moreover, sites are often located on low slope areas, close to marshlands [45], and not so far from higher sloped areas used for hunting and harvesting [54]. Finally neolithic sites seem to be completely absent in areas characterized by a geology with calcareous structures [64].

3.3. Research of including and excluding factors I: descriptive statistics of environmental parameters

Some descriptive statistics on environmental parameters have been calculated for the known sites. Results agree with elements explained in the previous paragraph. Data layers considered are hydrography, elevation, slope, aspect, lithology, land use and viewshed. The cell size used for raster used is 8m. Statistics have been calculated for distance to rivers, elevation, slope, aspect and viewshed. Statistics have not been calculated in each point that represents an archeological site because a site is not really punctual, but it occupies an area. Therefore, neighborhood statistics have been employed, by using a majority focal function and considering a circle area around points large 1ha, according to the values reported in paragraph 3.2 point 1. Results for the distance from rivers, elevation, slope and viewshed are shown in fig.4.

Additionally, in order to understand the preference of a river to a settlement, with regards to its distance, a KDEN has been calculated. The first interesting result is that the KDEN shows there is a density greater than 0 around high order rivers; that is rivers with a greater quantity of water. From these statistics, and considering the presence of outliers, the following including factors have been chosen (fig. 5):

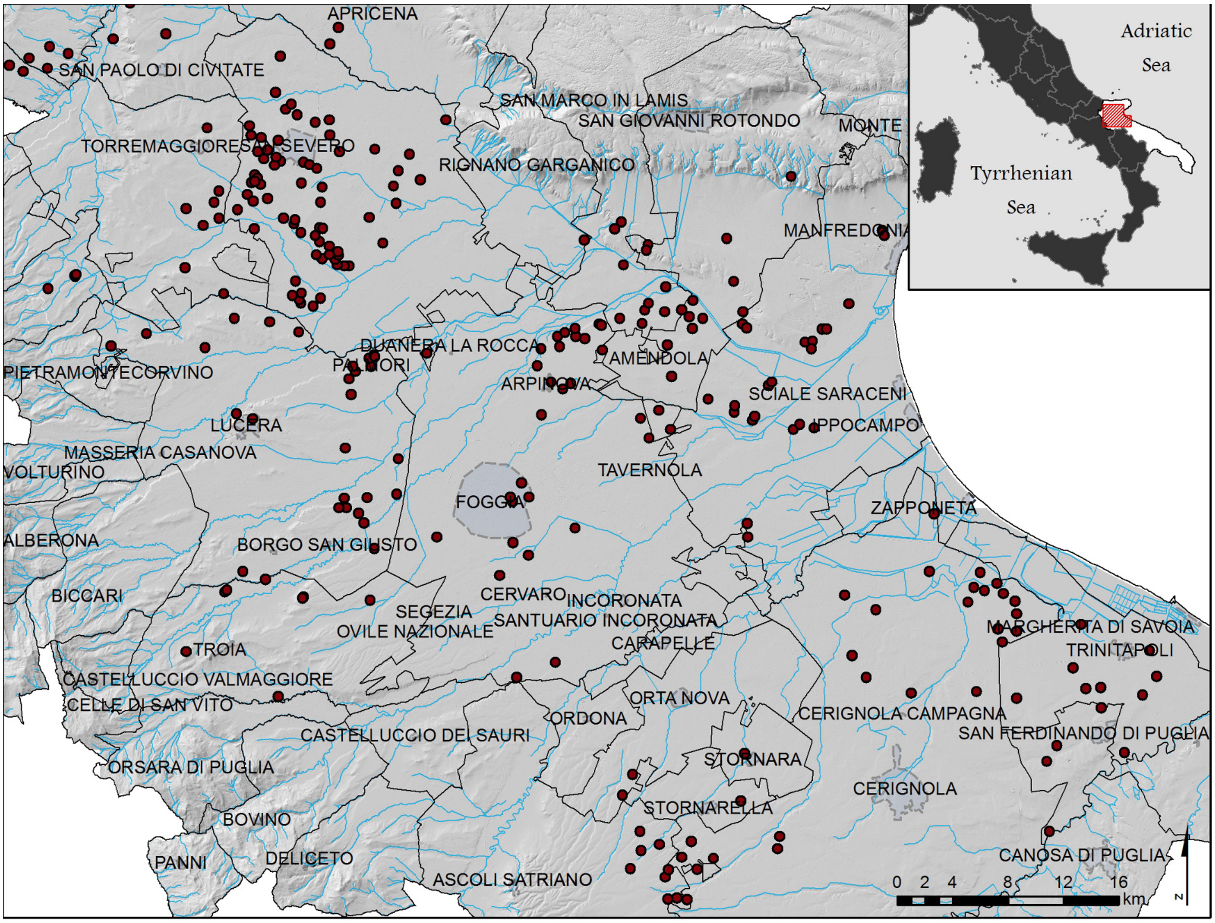

Figure 3. Neolithic sites of Apulian Tavoliere.

- areas significantly less than 4km away from rivers with a KDEN ¿ 0;
- areas with an elevation up to 350m;
- areas with a slope up to 4°;
- areas visible from at least one site

Aspect does not have strong effects on sites because classes are homogeneously distributed. The only meaningful result, coherent with slope statistics and with characteristics of Neolithic sites, is that the majority of sites are in the flat zone. For this reason, again, aspect has not been included in the analysis because it does not gives any additional information at the model. In addition, the lithology and land use layer have been reclassified. In terms of lithology calcareous dolomitic units, argillite with chaotic structure units, rocks succession with variable granulometry and composition units have been considered as excluding factors (fig. 5). In terms of land use, areas that are not considered suitable for Neolithic settlement are woods, bare rocks, falesias and outcrops. It is worth noting that during the Holocene Apulia was at the boundary between two climatic regions (eastern and western Mediterranean). As a consequence, the changes in climate that characterised the middle Holocene could have led to peculiar modifications in water availability for vegetation growth in this area [65]. The regional climatic variability is expressed by a mosaic of phytocoenoses corresponding to specific vegetation types, which in turn corresponds to the present vegetation covering used here as an including factor [66]. Thanks to archaeobotanical data resulting from plant remains of 35 multistratified sites, used in human activities (domestic fire or food processing), it was possible to derive palaeo vegetational (i.e. the meso-thermophilous/thermo-xerophilous ratio) and palaeo agronomical indicators: (a) the wheat/barley ratio (where a higher proportion of barley corresponds to relatively drier conditions), (b) hulled/naked cereal ratio (where a higher

Figure 4. Descriptive statistics for distance from rivers, elevation and slope of known archaeological sites.

proportion of the naked form indicates wetter conditions), and (c) the relative abundance of legumes, weeds and minor cereals (where the ecophysiological characteristics of the species are used as cropping period indicators). The analysis of plant remains indicates diachronical changes in the exploitation of edible species (cereals and leguminosae) and fuelwood throughout the Neolithic period; it was possible to identify two dry periods (one between 5000 and 4600 BC and another peaking c. 4000 BC) and two wet periods (one between 6200 and 5500 BC and another peaking c. 4400 BC) [66]. Here we have reported an example of research that is very useful, but only concerns 35 sites. As it is impossible to calibrate the model based on a few sites we are forced to bypass this lack of data and of homogeneity, as anticipated at the end of section 3.1 by only considering the data that is available and certain.

3.4. Research of including and excluding factors II: the study of spatial autocorrelation

To study if any spatial autocorrelation exists between neolithic sites, and to calculate the average modular distance between sites, first the Nearest Neighbor method has been applied. The following values have been found: NNOD=1356m; NNED= 2236m, NNI= 0.61. The value of the NNI, less than one, shows that there is spatial autocorrelation in the distribution. NNED has been instead used to calculate a KDE representing areas where, for the effect of positive autocorrelation, the probability to find a new site increases. Other parameters used inside the KDE are: the cell size (8m), the kernel of Epanechnikov [66], chosen because it is smooth and it has finite boundaries. As intensity only the occurrences of points have been used. Results obtained from KDE have been classified in ten quantile classes and the first nine classes have been used as including factors.

3.5. The final model

The calculated KDE and the environmental factors have been finally reclassified as including and excluding factors and combined with map algebra (a summarizing flow chart has been drawn in figure 7). A value of 1 has been assigned to including factors and a value of 0 has been assigned to excluding factors. They are combined according to the following formula, in order to determine areas suitable for neolithic settlements:

$$K*1000000+R*100000+E*10000+S*1000+L*100+U*10+V*1 \quad (5)$$

where K is a result of KDE, R is distance from rivers, E is elevation, S is slope, L is lithology, U is land use and V is viewshed.

The result is a raster with a binary number in each pixel. This binary number consists of 7 digits, corresponding to the number of parameters used in the model. Therefore, each digit represents one parameter and shows if the pixel is suitable (digit equal to 1) for neolithic sites according to that parameter or if it is not (digit equal to 0). The raster has been named KRESLUV to remember the reading order of factors and which is the digit that represents each factor, according to the expression (5). In this way by querying the KRESLUV raster it is possible to know how many and which parameters are suitable for Neolithic sites. As an example, if the pixel selected has a value of 1010011, it means that (K=1, R=0, E=1, S=0, L=0, U=1, V=1) it is suitable in terms of KDE, elevation, land use and viewshed, but not for distance from rivers, slope and lithology. Finally KRESLUV has been classified in order to show areas from 0 to 7 factors suitable for neolithic sites by grouping together pixels with a number of digits from zero to seven equal to 1. In this way 8 areas with an increasing probability level to find new sites have been obtained.

4. Results

In total, 111 different combinations of parameters have been found (Tab. 1). Figure (9) is useful for understanding how the area is distributed into the first sixty largest classes.

The following result has been obtained (fig. 10): the greatest percentage of sites fall in areas with 7 and 6 suitable factors.

With these values the validity of the model has been tested by calculating Kvamme's gain index [67] expressed by (6)

$$\text{Gain} = 1 - \left(\frac{\text{\% of total area covered by the model}}{\text{\% of total sites within model area}} \right) \quad (6)$$

As gain approaches 1, the predictive utility of the model increases. In table 2 the calculation of gain is reported for the percentage of total sites within different probability areas. The most relevant gain found is the value calculated for sites within areas exhibiting 7 suitable parameters; this means that the model has a good ability in predicting 75.90% of sites.

Due to the use of KDE the model is bound to the spatial and geometric properties of the site distribution. With this aim to control the model behaviour, it has been recalculated by subtracting each time a site in a random way,

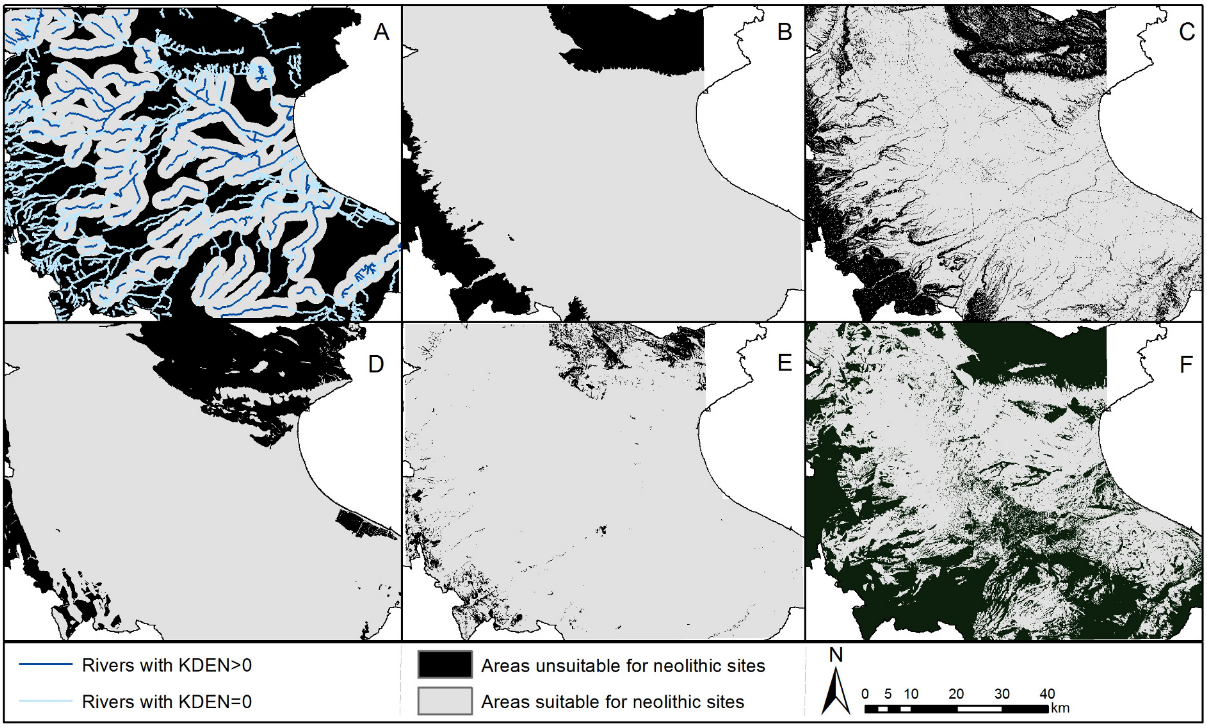

Figure 5. ncluding and excluding factors for the different environmental parameters considered: A) distance from river combined with KDEN; B) elevation; C) slope; D) lithology, E) land use and F) viewshed.

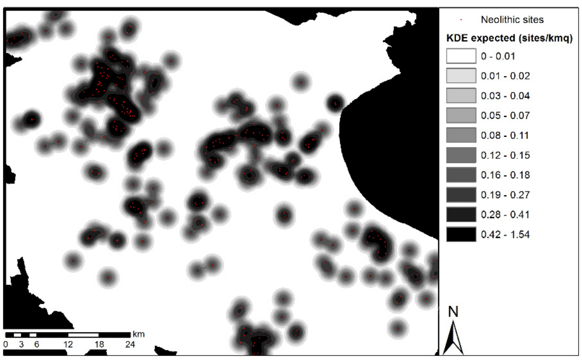

Figure 6. Classification of KDE combined with NNED returns areas with a greater probability to find new sites.

KDE + NNE	i.f. (1)	up to 9th quantile	$K*1000000+$
	e.f. (0)	10° quantile	
Hydrography	i.f. (1)	Distance<1500m & KDEN>0	$R*100000+$
	e.f. (0)	All the remaining areas	
Elevation	i.f. (1)	<350m	$E*10000+$
	e.f. (0)	>350m	
Slope	i.f. (1)	**Slope** <4°	$S*1000+$
	e.f. (0)	Slope >4°	
Lithology	e.f. (0)	- Calcareous dolomitic units - Argillite with a cahotic structure units - Rocks succession with variable granulometry and composition units	$L*100+$
	i.f. (1)	Remaining lythological classes	
Land use	e.f. (0)	- Woods - Bare rocks, falesias, outcrops	$U*10+$
	i.f. (1)	Remaining land use classes	
Viewshed	e.f. (0)	Visibility=0	$V*1$
	i.f. (1)	Visibility>0	

Figure 7. Flow chart summarizing the model. Here i.f. is including factor, while e.f. means excluding factor.

evaluating in which class of the new calculated model the subtracted site we forecast to find it and finally by looking the threshold of validity. The following cases could happen:

1. Isolated sites: after the recalculation of the model without the isolated site (far from other sites more than the NNED), its location falls in a lower probability area.

Table 1. Areal extension of classes found in KRESLUV. Classes with existing sites are highlighted in gray.

Class	Area (km^2)	Class	Area (km^2)	Class	Area (km^2)	Class	Area (km^2)
0000000	147,21	0011100	2,45	0111000	0,01	1100000	0,01
0000001	14,13	0011101	0,38	0111001	0,11	1100010	4,94
0000010	259,02	0011110	518,20	0111010	7,22	1100011	3,90
0000011	34,63	0011111	462,92	0111011	15,88	1100101	0,00
0000100	21,49	0100000	1,98	0111100	1,51	1100110	1,70
0000101	3,59	0100001	0,13	0111101	1,09	1100111	2,04
0000110	105,40	0100010	9,33	0111110	388,34	1101010	0,98
0000111	29,53	0100011	4,51	0111111	497,78	1101011	0,63
0001000	15,36	0100100	2,53	1000110	0,77	1101110	0,32
0001001	0,14	0100101	0,09	1000111	0,49	1101111	0,15
0001010	67,48	0100110	16,75	1001110	0,09	1110000	0,18
0001011	4,21	0100111	2,07	1001111	0,21	1110001	0,39
0001100	2,49	0101000	0,03	1010000	0,01	1110010	1,68
0001101	0,18	0101001	0,00	1010001	0,00	1110011	11,98
0001110	33,91	0101010	2,52	1010010	1,24	1110100	0,79
0001111	11,28	0101011	0,29	1010011	3,37	1110101	1,45
0010000	3,30	0101100	0,07	1010100	0,0001	1110110	20,85
0010001	5,38	0101101	0,00	1010101	0,21	1110111	67,37
0010010	8,06	0101110	1,95	1010110	2,66	1111000	0,01
0010011	20,54	0101111	0,29	1010111	9,61	1111001	0,01
0010100	3,28	0110000	0,13	1011000	0,01	1111010	5,39
0010101	0,75	0110001	1,83	1011001	0,00	1111011	22,73
0010110	123,92	0110010	1,09	1011010	11,07	1111100	0,60
0010111	68,62	0110011	6,46	1011011	27,93	1111101	1,34
0011000	0,12	0110100	0,51	1011100	0,0004	1111110	204,44
0011001	0,13	0110101	0,71	1011101	0,13	1111111	923,08
0011010	15,07	0110110	41,23	1011110	52,48		
0011011	42,95	0110111	46,37	1011111	229,94		

Table 2. Gain index obtained with different percentage of sites predicted.

	Sites within areas from 4 to 7 suitable parameters	Sites within areas from 5 to 7 suitable parameters	Sites within areas from 6 to 7 suitable parameters	Sites within areas with 7 suitable parameters
Percentage of Total area covered by model	79,34	63,78	41,46	19,66
Percentage of total sites within model area	100	98,9	95,4	75,90
Gain	0,21	0,36	0,57	0,74

2. Sites located on the edge of a cluster of sites: due to NNED being used as bandwidth, if just one site at the edge of the cluster is removed the level of probability does not change and the model is still valid; instead, if two or more sites along the edge are removed its location results in a lower level of probability.

3. Sites located inside a cluster of sites: the result of the predictive model does not have any change.

Finally the effectiveness decreases with the numerousness of the basic dataset. It follows that 1) the more a sample dataset is complete, the more our predictive model will improve 2) the model is less valid outside clusters.

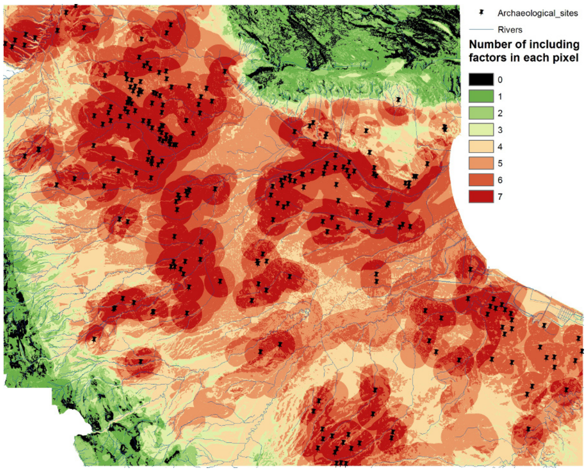

Figure 8. The predictive map of Apulian Tavoliere.

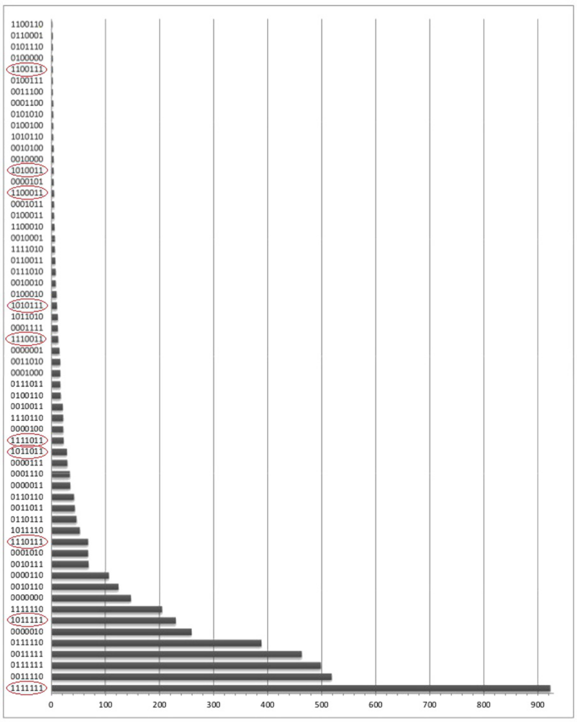

Figure 9. Histogram showing the areal distribution of the first 60 classes found in KRUSLEV raster. Classes with known sites are marked with a red circle

5. Conclusions

The predictive model proposed in this paper attempts to take into account both environmental layers (that characterize areas with Neolithic settlements) and the spatial relationships among sites using the spatial autocorrelation analysis. If from one side the use of spatial autocorrelation is a point of strength of the approach we devised, because it allows us to focus the research inside the contiguity belt outlined by NN methods and KDE, on the other hand, it can also be a point of weakness of the model because they are strongly linked. Another result obtained with the synergy between environmental factors and KDEN is the improvement in the estimation of areas with some probability to find new sites. Looking at fig. 7, in fact not all areas near rivers are in the probability level from low to high; some of them belong to the very low level, so that the model is sensible to the inverse proportionality existing between the hydrographical network and the number of sites. So these places may be regarded as less suitable for living in the Neolithic time and, in fact, areas with a branched network are more subject to erosion phenomena, so they are less cultivable and the water is wasted and less usable compared to conditions in the main stream. Finally the "KRESLUV" provides an instrument that can be especially useful for territorial or urban planning because it not only returns areas with different sensibility to the presence of neolithic sites, but also, by

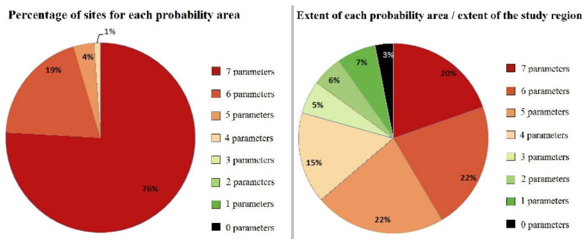

Figure 10. Pie charts for percentage of sites in each probability area (on the left) and for the percentage occupied by each probability area.

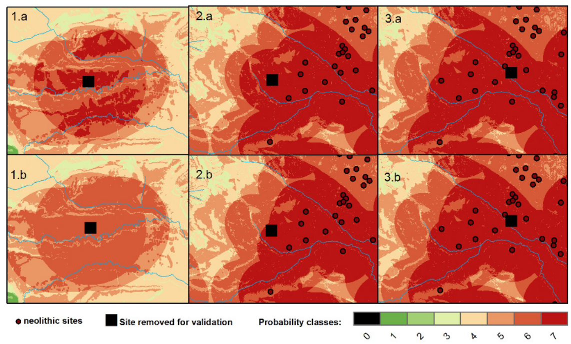

Figure 11. Validation of the model first (a) and after (b) it is recalculated by subtracting a site each time (with the black squared symbol) in a random way in the case of 1.a-1.b for an isolated site, 2.a-2.b for a site located on the edge of a cluster, 3.a-3.b for a site located inside a cluster.

querying the map, it is possible to determine which parameters influenced that sensibility as support for reasoning and decision making. Of course, this attempt to integrate spatial autocorrelation and the use of environmental parameters could be further improved: firstly because there are many other parameters that could be included in the analysis such as the communication networks between sites, distance from the sea and the variation of the location of settlements within the Neolithic period, and secondly because there is the need to go more into palaeoenvironmental data.

6. Acknowledgements

The authors thank Basilicata Region for supporting this research in the framework of the project PRO_CULT ("Advanced methodological approaches and technologies for Protection and Security of Cultural Heritage") financed by OP Basilicata ERDF programme 2007/2013 .

References

[1] Kamermans H., Van Leusen M., Verhagen P., Archaeological prediction and risk management. archaeological studies. Leiden University Press, 2009

[2] Varnes D.J., Slope movements, type and processes. In: Schuster R.L., Krizek R.J. (Eds.), Landslides-analysis and control: National Research Council, Washington, D.C., Transportation Research Board, Special Report 1978, 176, 11-33

[3] Danese M., Masini N., Murgante B., Archaeological risk and spatial analysis. How to compare urban sprawl and archaeological sensibility maps. EARSeL eProceedings 2013 (in press)

[4] Danese M., Masini N., Biscione M., Lasaponara R., Gis and archaeology: a spatial predictive model for neolithic sites of the Tavoliere (Apulia). In: Hadjimitsis D.G., Themistocleous K., Michaelides S., Papadavid G. (Eds), Proc. SPIE 8795, First International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2013), 2013

[5] Lasaponara R., Danese M., Masini N., Satellite-based monitoring of archaeological looting in Peru. Remote sensing and digital image processing, 2012, 16, 177-193, doi: 10.1007/978-90-481-8801-7_8

[6] Hadjimitsis D., Agapiou A., Alexakis D., Sarris A., Exploring natural and anthropogenic risk for cultural heritage in Cyprus using remote sensing and GIS. Int. J. Digital Earth, 2013, 6, 2, 115-142

[7] Anschuetz K.F., Wilshusen R.H., Scheik C.L., An archaeology of landscapes: perspectives and directions. J. Archaeol. Res., 2001, 9, 2, 157-211

[8] Ashmore W., Knapp A.B. (Eds.), Archaeologies of landscape: Contemporary perspectives, Blackwell Publishers, Malden, MA, 1999

[9] Sever T.L., Irwin D.E., Landscape archaeology: remote-sensing investigation of the ancient Maya in the Peten rainforest of northern Guatemala. Ancient Mesoamerica, 2003, 14, 113-122

[10] Vaughn S., Crawford T., A predictive model of archaeological potential: an example from northwestern Belize. App. Geog., 2009, 29, 4, 542-555

[11] Walker J.H., Recent landscape archaeology in South America. J. Archaeol. Res., 2012, 20, 4, 309-355, DOI 10.1007/s10814-012-9057-6

[12] Niknami K.A., A stochastic model to simulate and predict archaeological landscape taphonomy: monitoring cultural landscapes values based on an iranian survey project. Archeologia e Calcolatori, 2007, 18, 101-120

[13] Podobnikar T., Veljanovski T., Stan?i? Z., Oï'·tir K., Archaeological predictive modelling in cultural resource management. In: Koneï'·nï'· M. (Ed.), GI in EUROPE: integrative - interoperable - interactive. Proceedings of 4th Agile Conference on Geographic Information Science, 2001, 535-544

[14] Ciminale M., Gallo D., Lasaponara R., Masini N., A multiscale approach for reconstructing archaeological landscapes: applications in Northern Apulia (Italy), Archaeol. Prospect., 2009, 16, 143-153

[15] Gallo D., Ciminale M., Becker H., Masini N., Remote sensing techniques for reconstructing a vast Neolithic settlement in Southern Italy. J. Archaeol. Sci., 2009, 36, 43-50

[16] Brandt R., Groenewoudt B.J., Kvamme K.L., An experiment in archaeological site location: modeling in the netherlands using gis techniques. World Archaeol., 1992, 24, 2, 268-282

[17] Stan?i? Z., Kvamme, K., Settlement pattern modelling through boolean overlays of social and environmental variables. In: Barcelo J.A., Briz I., Vila A. (Eds.), New techniques for old times, CAA 98, BAR International Series , 1999, 757, 231-237

[18] Verhagen P., Testing archaeological predictive models: a rough guide. In: Posluschny A., Lambers K., Herzog I. (Eds.), Layers of perception. Proceedings of the 35th international conference on computer applications and quantitative methods in archaeology (CAA), Berlin, Germany, 2007, 10, 285-291.

[19] Garcia A., GIS-based methodology for Palaeolithic site location preferences analysis. A case study from Late Palaeolithic Cantabria (Northern Iberian Peninsula), J. Archaeol. Sci., 2013, 40, 217-226

[20] Canning S., BELIEF in the past: Dempster-Shafer theory, GIS and archaeological predictive modeling. Aust. Archaeol., 2005, 60, 6-15

[21] Zhongxuan L., Gang H., Cheng Z., Multi-fractal analysis on neolithic sites around the Sanxia Reservoir Area. RSETE, 2011, 491-494

[22] Alexakis D., Sarris A., Astaras T., Albanakis K., Integrated GIS, remote sensing and geomorphologic approaches for the reconstruction of the landscape habitation of Thessaly during the neolithic period. J. Archaeol. Sci., 2011, 38, 89-100

[23] De Mers M.N., GIS Modeling in raster. USA, 2002

[24] Tomlin C.D., Geographic Information Systems and cartographic modeling. Englewood Cliffs, New Jersey, Prentice Hall, 1990

[25] Tobler W.R., A Computer model simulating urban growth in the Detroit region. Econ. Geogr., 1970, 46, 234-240

[26] O'Sullivan D., Unwin D., Geographic Information Analysis. John Wiley & Sons, 2002.

[27] Gatrell A.C., Bailey T.C., Diggle P.J., Rowlingson B.S., Spatial point pattern analysis and its application in geographical epidemiology. T. I. Brit. Geogr., 1996, 21, 256-271

[28] Silverman B.W., Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1986

[29] Flahaut, B., Mouchart, M., Martin, E.S., THOMAS, I., The local spatial autocorrelation and the kernel method for identifying black zones a comparative approach. Accident Anal. Prev., 2003, 35, 991-1004.

[30] Borruso G., Network density and delimitation of urban arrears. Transaction in GIS, 2004, 7, 177-191

[31] Borruso G., Network density estimation: analysis of point patterns over a network. In: O. Gervasi et al (Eds.), Computational Science and its applications. Lecture notes in Computer Science (Berlin, Heidelberg: Springer-Verlag) 2005, 3482, Part III, 126-132

[32] Borruso G., Network density estimation: a GIS approach for analyzing point patterns in a network space. Transactions in GIS, 2008, 12, 377-402

[33] Downs J.A., Horner M.W., Spatial modeling pathways of migratory birds for nature reserve site selection. Int. J. Geogr. Inf. Sci., 2008, 22, 6, 687-702

[34] Downs J. A., Horner M.W., Characterising linear point patterns. In: Winstanley A.C. (Ed), Proceedings of the GIS Research UK Annual Conference, Maynooth: National University of Ireland, 2007, 421-425

[35] Downs J. A., Horner M.W., Network-based kernel density estimation for home range analysis. In: Demsar U. (Ed.), Proceedings of the 9th International Conference on Geo-Computation, Maynooth: National Center for Geocomputation, 2007

[36] Xie Z., Yan J., Kernel density estimation of traffic accidents in a network space. Comput. Environ.Urban Systems, 2008, 32, 396-406

[37] Okabe A., Okunuki K., Shiode S., The SANET toolbox: new methods for network spatial analysis. Transactions in GIS, 2006, 10, 4, 535-550

[38] Okabe A., Okunuki K., Shiode S., SANET: a toolbox for spatial analysis on a network. Geogr. Anal., 2006, 38, 1, 57-66

[39] Okabe A., Satoh T., Sugihara K., A kernel density estimation method for networks, its computational method and a GIS-based tool. Int. J. Geogr. Inf. Sci., 2009, 23, 1, 7-32

[40] Jones G.D.B., Apulia-Vol. I: Neolithic Settlements in the Tavoliere. Thames and Hudson, London, 1987

[41] Bradford J.S.P., Buried landscapes in Southern Italy, Antiquity, 1949, 23, 58-72

[42] Cassano S.M., La diffusione del Neolitico in Puglia. Convegno Nazionale sulla Preistoria-Protostoria-Storia della Daunia, San Severo, 1979

[43] Gravina A., Caratteri del Neolitico medio-finale nella Daunia centro-meridionale- 6° Convegno Nazionale sulla Preistoria-Protostoria-Storia della Daunia, San Severo, 1984

[44] Gravina A., Masseria Istituto di Sangro un insediamento del neolitico medio-finale nella Daunia, 8° Convegno Nazionale sulla Preistoria-Protostoria-Storia della Daunia, San Severo, 1986

[45] Gravina A., Masseria Santa Giusta. Un insediamento del Neolitico antico nella Daunia, 9° convegno nazionale sulla Preistoria - Protostoria - Storia della Daunia, San Severo, 1987

[46] Mazzei M. (Eds.), La Daunia antica: dalla preistoria all'altomedioevo. Milano, Electa, 1984

[47] Delano Smith C., Daunia vetus. Terra, vita e mutamenti sulle coste del Tavoliere. Amministrazione Provinciale di Capitanata, Foggia, 1978

[48] Whitehouse R.D., Settlement and economy in Southern Italy in the Neothermal period. Proceeding of the Prehistoric Society, 1968, 34, 332-367

[49] Whitehouse R.D., Società ed economia nel Neolitico italiano: la problematica dei fossati. In: La preistoria del Basso Belice e della sicilia meridionale nel quadro della preistoria siciliana e mediterranea. Società siciliana per le Storia Patria, Palermo, 1994, 275-285

[50] Jarman M.R., Webley D., Settlement and land use in Capitanata, Italy.In Higgs E.S. (Eds.), Papers in economic prehistory, Cambridge University Press, Ca, Cambridge, 1975,177-221

[51] Sargent A., Changing settlements location and sub-

sistence in later prehistoric Apulia, Italy, Origini, 2001, 23, 145-168

[52] Brown K.A., A passion for excavation. Labour requirements and possible functions for the ditches of the ”villaggi trincerati” of the Tavoliere, Apulia. The Accordia Research Papers, 1991, 2, 7-30

[53] Brown K.A., Settlement distribution and social organisation in the Neolithic of the Tavoliere, Apulia. In: Herring R., Whitehouse R., Wilkins J. (Eds.), Papers of the Fourth Conference of Italian Archaeology, The Archaeology of Power , 1991, 1, 9-25

[54] Malone C., The italian Neolithic: a synthesis of research. *J. World Prehist.*, 2003, 17, 3, 235-312

[55] Skeates R., The social dynamics of enclosure in the Neolithic of the Tavoliere south-east Italy. J. Mediterr. Archaeol., 2000, 13, 155-188

[56] Volpe G., Gli insediamenti rurali dall’eta romana all’età tardoantca. Il mondo dell’archeologia, 1st. Treccani ed. Roma, 2002, 1, 911-917

[57] Sanderson D.C.W., Murphy S., Using simple portable OSL measurements and laboratory characterisation to help understand complex and heterogeneous sediment sequences for luminescence dating. Quat. Geochronology, 2010, 5, 299-305

[58] Radcliffe F., Paesaggi sepolti in Daunia - John Bradford e la ricerca archeologica dal cielo 1945/1957. Claudio Grenzi Editore, Foggia, 2006

[59] Bradford J.S.P., Ancient Landscapes: Studies in field archaeology, G. Bell and Sons, Ltd., London, 1957

[60] Tozzi C., Contributo alla conoscenza del villaggio neolitico di Ripa Tetta (Lucera). 6° convegno sulla Preistoria-Protostoria - Storia della Daunia, San Severo, 1984

[61] Biancofiore F., Note di antropologia economica delle comunit? neolitiche della Puglia centro-settentrionale, 5° convegno nazionale sulla Preistoria - Protostoria - Storia della Daunia. San Severo, 1983

[62] Gravina A., Le comunità neolitiche di Coppa Pallante, 5° convegno nazionale sulla Preistoria - Protostoria - Storia della Daunia. San Severo, 1983

[63] Caldara M., Pennetta L., Influenza del clima sul paesaggio neolitico del Basso Tavoliere di Puglia. In: Tinè V. (Ed.), Forme e tempi della neolitizzazione in Italia Meridionale e in Sicilia, Rubbettino, Soveria Mannelli, 1996, 559-568

[64] Coppola D., Indagini paletnologiche su un insediamento neolitico in località Le Macchie (Polignano a Mare - Bari), 3° convegno nazionale sulla Preistoria - Protostoria - Storia della Daunia, San Severo, 1981

[65] Gladstone R.M., Ross I., Valdes P.J., Abe-Ouchi A., Braconnot P., Brewer S., Kageyama M., Kitoh A., Legrande A., Marti O., Ohgaito R., Otto-Bliesner B., Peltier W.R., Vettoretti G., Mid-Holocene NAO: A PMIP2 model intercomparison. *Geophy. Res. Lett.*, 2005, 32, L16707, doi:10.1029/2005GL023596

[66] Fiorentino G., Caldara M., De Santis V., D’Oronzo C., Muntoni I.M., Simone O., Primavera M., Radina F., Climate changes and human-environment interactions in the Apulia region of southeastern Italy during the Neolithic period, , *Holocene*, 2013, 23, 1297-1317, doi: 10.1177/0959683613486942

[67] Epanechnikov V.A., Nonparametric Estimation of a Multivariate Probability Density, Theor. Probab. Appl.+, 1969, 14, 153-158

[68] Kvamme K., Development and testing of Quantitative models. In: Judge W., Sebastian L. (Eds.), Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modelling, Denver, 1988, 325-428.