EFFICIENT PREDICTIVE MODELLING FOR ARCHAEOLOGICAL RESEARCH

A. Balla¹, G. Pavlogeorgatos², D. Tsiafakis³, G. Pavlidis³

¹31st Ephorate of Prehistoric and Classical Antiquities, Archaeological Museum of Avdera, 67100, Avdera, Xanthi, Greece
²Department of Cultural Technology and Communication, University of the Aegean, University Hill, building of Geography, Off.2.14, 81-100, Mytilene, Lesvos, Greece
³ATHENA Research Centre, University Campus at Kimmeria, 67100, Xanthi, Greece

Received: 28/03/2013
Accepted: 30/10/2014

Corresponding author: George Pavlidis (gpavlid@ceti.gr)

ABSTRACT

The study presents a general methodology for designing, developing and implementing predictive modelling for identifying areas of archaeological interest. The methodology is based on documented archaeological data and geographical factors, geospatial analysis and predictive modelling, and has been applied to the identification of possible Macedonian tombs’ locations in Northern Greece. The model was tested extensively and the results were validated using a commonly used predictive gain, which proved the efficiency of the model’s predictive ability and its capability in providing answers to a series of questions related to archaeological research issues.

KEYWORDS: archaeology, predictive modelling, GIS, archaeological research, Macedonian tombs
1. INTRODUCTION

The application of technologies in humanities and specifically in archaeology attracted an increased attention during the recent years. Several techniques, mostly based on information technology have already been successfully applied in archaeological research and some of them are already considered as common practice. As the nature of the archaeology research involves the analysis of spatial and temporal data, it has soon been realised that GIS technologies could provide useful tools. The application of GIS technologies in archaeology over the past recent years has yielded important expertise that can be successfully exploited by archaeological research and, moreover, according to some scholars, is expected to be an integral part of archaeological practice, in interpreting and understanding the socio-economic structure of the past (Harris and Lock 1995).

However, the application that set GIS as a mainstream tool in the field of archaeology is predictive modelling (Gourad 1999), namely “a technique that, at a minimum, tries to predict the locations of archaeological sites or materials in a region, based either on a sample of that region or on fundamental notions concerning human behaviour” (Kohler and Parker 1986). The term “predictive models” was adopted internationally to describe tools for projecting patterns or known relationships into related areas of unknown patterns or relationships (Warren and Asch 2000). Van Leusen (2002) suggests that PM can be conceptualised as a specialised form of location-allocation analysis, where the aim is to allocate suitable locations to specific types of human activity and its archaeological remains. In this analysis, the criteria for suitability are derived by location analysis, namely by the generation of the behavioural norms from the observations of the way people behave or have behaved in the past.

The basic principle upon which this scientific field was based is that the selection of human activity locations in the ancient times was related to the current period environmental and geographical conditions. Based on these conditions that characterize a location, repeating patterns can be identified. These, compared to patterns of other areas with similar geographic features found at the same period, may result in identifying new locations that may also have been occupied by similar human activities (Hatzinikolaou and Hatzichristos 2004). Namely, predictive modelling aims at establishing a causal relationship between certain environmental parameters and known archaeological site locations, in order to create a statistical model based on that relationship that can be applied to unsurveyed areas (Gourad 1999).


In this work we propose a simple and efficient methodology that could generalize the predictive modelling approaches in archaeology. This methodology has been tested in ancient Macedonian tombs’ locations and proved highly successful in locating areas of possible archaeological sites’ occurrence. Models based on the proposed methodology would enrich archaeological knowledge about ancient culture and would contribute to the study of ancient topography, as the discovery of new sites can result in finding yet undiscovered settlements, roads, burial monuments etc. Furthermore, the proposed model can be used as an efficient solution to the lack of
funding by minimizing the number of trial excavations and by indicating specific areas that are of high probability to result in finding undiscovered archaeological remains.

In the following sections, we describe the proposed methodology and the model development process along with the experimental results and the evaluation of the model.

2. MODEL-BUILDING PROCESS

The proposed methodology is based on the following procedures: through archaeological research and data aggregation, assumptions related to the location of the sites of interest are formulated, resulting in the selection of criteria considered to have influenced their siting. At the core of the proposed methodology, a multi-criteria analysis on GIS data is being employed. A hybrid inductive-deductive approach of the criteria is proposed, whereas fuzzy logic is applied for the criteria normalization and quantification. The basic workflow of the methodology consists of the following steps:

1. Archaeological research-Data selection
2. Selection and theoretical approach of the criteria
3. Quantification of the criteria
4. Selection of importance and calculation of criteria weights
5. Criteria data aggregation

The model created can be tested under various combinations of parameters related to the criteria. The results are evaluated by using a commonly used predictive gain, which can test the predictive ability of the model for different cases. In the following paragraphs the basic steps of the modelling process are being discussed.

2.1 Archaeological research-data selection

The first concern for the development of the model is to collect all the involved archaeological data. This stage includes an extensive archaeological literature research, followed by field survey, in order to locate the archaeological sites indicated by the literature references. However, it might not be possible to locate and acquire the geographical coordinates of all sites, as some of them are today damaged or even lost. The next step of the data collection phase involves gathering the appropriate data that can be visualized through GIS software, and moreover, referred to the time periods, in which the sites of interest are dated (e.g. Classical period). In many cases one has to take into account possible geomorphological changes through time of the selected region. The final procedure of the data collection phase involves the selection of the most appropriate Digital Elevation Model (DEM), which includes literature research on resource identification and data quality evaluation.

2.2 Selection of the criteria

Archaeological predictive models use multi-parametric spatial analysis of geographic and archaeological data in order to identify areas of possible archaeological interest. Despite the differentiation in the approach of the analysis, they practically follow the same procedure: their creation is based on the correlation of the environmental parameters with the known archaeological sites, whose statistical analysis correlates, based on specific decision making rules, the spatial characteristics of the site with other similar areas of possible archaeological interest. Namely, the data in all archaeological predictive models are a combination of geospatial and cultural features related both to the study area and the subject of study. Typical environmental parameters encountered in predictive models in Archaeology are altitude, slope, orientation, geological and soil data, topography, hydrographical network distances from water bodies, even vegetation (in cases where it is assumed that there has not been greatly changed from past times). More complicated is the use of cultural parameters, which are sometimes difficult to be quantified. Such parameters arise from the observation that the under study archaeological site is located close to other
important or central cultural features in the landscape, such as settlements, sanctuaries, roads etc. It is clear, however, that not all these environmental and cultural parameters can be used as input data in all types of archaeological sites. For example, the socio-economic factors and features such as topographic relief, distance from water bodies or soil cover type that also had an important role in the locational processes of ancient settlements (Bauer et al. 2004, Duke 2003, Fletcher 2008, Kvaamme 1992, Stancic and Kvaamme 1999, Vanacker et al. 2001, Warren 1990, Willey 1953, Williams 1956, Williams et al. 1973) cannot be the same with the factors that led to the locational decision making for burial mounds or sanctuaries. Furthermore, it is clear that those factors-criteria can vary even for the same type of archaeological site, as they may be related to a specific time period, region or specific cultures.

2.3 Quantification of the criteria

A prerequisite for the implementation of a predictive model is to quantify the selected criteria. The first stage of this process includes the specification of certain areas around dimensionless data (hereinafter called “zones”), such as vector map data (points and lines). In the second stage of the quantification process, all criteria are normalized so that they could be referenced in a common scale, and therefore the criteria aggregation can be performed on a common basis. To this end, the criteria values are normalised using fuzzy logic membership functions.

2.4 Selection of importance and calculation of criteria weights

The next stage of the model-building process includes the calculation of the criterion weights, namely the importance of each criterion. One of the most popular methods used in a multi-criteria analysis is the Analytical Hierarchy Process (AHP) (Saaty 1980), a structured technique for organising and analysing complex decisions. In AHP one can attribute different weights to the criteria or sets of activities, depending on the degree of their significance, by making pairwise comparisons of the criteria, based on the decision makers’ judgments about their relative meaning and importance.

2.5 Criteria data aggregation

The process of criteria data aggregation can be achieved by using the method of Weighted Linear Combination (Voogd 1983), whereby each criterion’s value is multiplied with the value of its weight and the results are summed. Provided that the sum of all weights equals to 1, the result of the aggregation will have the same range as the one specified for the criteria. The aggregation process can be mathematically expressed by the following equation:

\[ S = \sum w_i x_i \]  

where \( S \) is the probability of archaeological site occurrence, \( w_i \) the weight of the criterion \( i \) and \( x_i \) the value of the criterion \( i \).

3. EXPERIMENTAL RESULTS AND MODEL VALIDATION – THE CASE OF MACEDONIAN TOMBS

Undoubtedly, burial mounds, tombs and cemeteries have been the subject in many studies, which, however, examine the correlation between topography and their location on the landscape (De Reu et al. 2011, Löwenborg 2010b), chronological estimations (Löwenborg 2009), viewshed and visibility (Fisher et al. 1997, Lageras 2002, Wheatley 1995, Woodman 2000) or simply included among other archaeological data, the locations of funerary monuments and cemeteries to map archaeological sites. The studies found in literature regarding exclusively the prediction of burial monuments or mounds are rare (Al-Muheisen & Al-Shorman 2004, Burns et al. 2008, Fry et al. 2004, Löwenborg 2010a). The noticeable lack of that kind of predictive models attracted our attention in the particular scientific field and triggered the effort of creating a predictive model for the detection of Macedonian tombs in Northern Greece.
In the case of Macedonians tombs the selection of the criteria was determined by the environmental and cultural factors that were considered to have influenced the choice of their location. In order to extract the criteria that led to the specific human decision rules it was necessary to study thoroughly the particular type of archaeological site. It was clear, based on the documented archaeological research on Macedonian tombs that features such as orientation, for example, of those funerary monuments could not be used as a criterion, because the archaeological research does not identify a certain pattern or certain principles (Lilimbaki-Akamati and Troxidis 2004, Miller 1993). Similarly, the distance from sanctuaries or temples could neither be used, because the archaeological literature to date does not substantiate any relationship between Macedonian tombs and religious places. Other variables, such as distances among burials themselves, or symbolic aspects related to other kind of tombs found in different places or dated in different time periods and therefore were associated with different cultures are not archaeologically documented as relative to the late classical and Hellenistic Macedonian culture or as influential factors regarding their location. Thus, by taking under consideration the literature research on all Macedonian tombs, and, also, based on the existing geographic data, we ended up with four environmental (altitude, slope, soil hardness, distance from rivers) and two cultural parameters (distance from settlements, distance from roads).

The proposed predictive model has initially been tested for cases where all six criteria were considered equally important and therefore shared the same weight. Additionally, by taking into account the archaeological research that documents the existence of Macedonian tombs near settlements and roads, the criteria “distance from settlements” and “distance from roads” were considered to be of higher importance than other criteria, and so they could be assigned higher weights. This could be quantified by varying the significance of these criteria using a multiplier (2x up to 6x).

Thus, in order to examine the predictive capability of the model, the uncertainty regarding the selection of the zones’ range and the sensitivity of the model, we have tested the model using a total of forty-two (42) combinations of the criteria weights (w) and the zones’ range (m). The different zones’ sizes are defined as multiples of the median distance from the exact location of the dimensionless data. Table I shows the selected combinations.

<table>
<thead>
<tr>
<th>Criteria weights (w) and Median multipliers (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
</tr>
<tr>
<td>w2</td>
</tr>
<tr>
<td>m0.15</td>
</tr>
<tr>
<td>w3</td>
</tr>
<tr>
<td>m0.15</td>
</tr>
<tr>
<td>w4</td>
</tr>
<tr>
<td>m0.15</td>
</tr>
<tr>
<td>w5</td>
</tr>
<tr>
<td>m0.15</td>
</tr>
<tr>
<td>w6</td>
</tr>
<tr>
<td>m0.15</td>
</tr>
</tbody>
</table>

The 42 different scenarios of the model resulted in the production of 42 maps of the region of interest with colour coding corresponding to different probabilities of archaeological sites’ occurrence. In Figs. 1-3 we present results of three scenarios of the model. On these maps we overlap the known Macedonian tombs’ locations as black dots, in order to provide a first impression about the predictive efficiency of the model.

![Figure 1. Map results showing probability zones and distribution of known Macedonian tombs for the combination w1/m1](image-url)
The simple, visual evaluation of the results indicates that a large number of Macedonian tombs are located in areas of high and very high probability for scenario w1/m1 (Fig. 1), where the zones equal to the median of the distances of the known Macedonian tombs from settlements, rivers and roads. The results of scenarios w3/m1 and w5/m1 (Fig. 2, Fig. 3) suggest that higher weight values of the criteria “distance from settlements” and “distance from roads” significantly reduce the area of high and very high probability and thus affect the identification of the known Macedonian tombs in all areas. Another important conclusion that emerges from the initial processing of all the test results is that increasing the zones’ sizes (m) leads to an increasing number of identified known tombs in the high and very high probability areas.

However, the fact that it has been possible to construct a predictive model does not in itself guarantee the accuracy of its predictions (Conolly and Lake 2006). The validation of the model must be examined with reference to the areas, which the model indicates as most likely to find Macedonian tombs. Kvamme (1988) proposed the validation of a predictive model, defining a predictive gain as follows:

\[
G = 1 - \frac{E_1}{E_2} \tag{2}
\]

where, in the case of our model, \(E_1\) is the percentage of the total area where tombs are predicted, and \(E_2\) is the percentage of known tombs identified within the area where they are predicted, for a given probability of site occurrence. It is noted that the gain \(G\) is calculated for a specified probability of archaeological sites occurrence. The gain ranges between -1 and 1, where a zero (0) value indicates no predictability, value one (1) indicates the highest predictive utility and minus one (-1) the highest predictive utility for the reverse of what is initially supposed.

In order to validate our model using the gain \(G\), we used the spatial coordinates of all known Macedonian tombs that have been collected. In an attempt to identify the areas of high and very high probability with a high allocation of known Macedonian tombs we have isolated the results of those tests that detect more than 75% of the known tombs (Table II).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Tombs’ detection</th>
<th>Zone area</th>
<th>Gain G</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1/m1.25</td>
<td>81.93%</td>
<td>38.08%</td>
<td>0.535</td>
</tr>
<tr>
<td>w1/m1.5</td>
<td>92.77%</td>
<td>41.40%</td>
<td>0.554</td>
</tr>
<tr>
<td>w1/m2</td>
<td>99.38%</td>
<td>47.17%</td>
<td>0.498</td>
</tr>
<tr>
<td>w1/m3</td>
<td>97.59%</td>
<td>55.16%</td>
<td>0.435</td>
</tr>
<tr>
<td>w2/m1.5</td>
<td>87.95%</td>
<td>16.55%</td>
<td>0.812</td>
</tr>
<tr>
<td>w2/m2</td>
<td>91.57%</td>
<td>23.05%</td>
<td>0.748</td>
</tr>
<tr>
<td>w3/m2</td>
<td>85.54%</td>
<td>18.38%</td>
<td>0.785</td>
</tr>
<tr>
<td>w4/m2</td>
<td>85.54%</td>
<td>18.36%</td>
<td>0.785</td>
</tr>
<tr>
<td>w5/m2</td>
<td>79.52%</td>
<td>18.12%</td>
<td>0.772</td>
</tr>
<tr>
<td>w6/m2</td>
<td>79.52%</td>
<td>18.03%</td>
<td>0.773</td>
</tr>
</tbody>
</table>

The maximum value of gain \(G\) (0.812) is achieved for scenario w2/m1.5, where 87.95% of the known Macedonian tombs fall in an area of 16.55% of the total survey area. On the contrary, the minimum value of gain \(G\) (0.435) resulted for scenario w1/m3, in which a great percentage
(97.59%) of the tombs was found, but in a large land area (55.16%). The latter result indicates that there is a threshold in increasing the zones (m). Any higher value above that threshold does not improve the predictive ability of the model (even though a large percentage of Macedonian tombs is detected), as it also increases the area of high and very high probability. Therefore, the best performance of the predictive model lies in the use of scenario w2/m1.5, obtaining a high percentage of Macedonian tombs in just 16.55% of the total land area of interest. A graphical representation of the results shown in Table II is provided in Fig. 4.

Focusing solely on the areas of very high probability, we still get successful prediction results. Due to the restrictions imposed by adopting only very high probability areas we have to relax our requirements of identification percentage. As of this, the attention is directed towards the identification of those areas, in which the detection of known Macedonian tombs (in very high probability areas) is over 50%. Table III summarizes the test results for this case. A graphical representation of the results shown in Table III is provided in Fig. 5.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Tombs’ detection</th>
<th>Zone area</th>
<th>Gain G</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1/m1.5</td>
<td>62.65%</td>
<td>11.36%</td>
<td>0.819</td>
</tr>
<tr>
<td>w1/m2</td>
<td>78.31%</td>
<td>17.03%</td>
<td>0.783</td>
</tr>
<tr>
<td>w1/m3</td>
<td>93.98%</td>
<td>26.30%</td>
<td>0.720</td>
</tr>
<tr>
<td>w2/m2</td>
<td>55.42%</td>
<td>5.26%</td>
<td>0.905</td>
</tr>
<tr>
<td>w3/m2</td>
<td>55.42%</td>
<td>5.28%</td>
<td>0.905</td>
</tr>
<tr>
<td>w4/m2</td>
<td>56.63%</td>
<td>5.46%</td>
<td>0.903</td>
</tr>
<tr>
<td>w5/m2</td>
<td>56.63%</td>
<td>5.47%</td>
<td>0.903</td>
</tr>
<tr>
<td>w6/m2</td>
<td>56.63%</td>
<td>5.47%</td>
<td>0.903</td>
</tr>
</tbody>
</table>

The maximum value of gain G (0.905) is achieved for w2/m0.5, where 55.42% of the known Macedonian tombs fall in an area of only 5.26% of the total survey area. On the contrary, the minimum value of gain G (0.720) resulted for w1/m3, which detected 93.98% of the known tombs within an area of 26.30% of the total survey area.

Further processing of the test results led to another significant acknowledgement: In both cases (high and very high probability areas, solely very high probability areas) different weight values for “distance from settlements” and “distance from roads” can only influence the predictive ability of the model up to a certain degree. Therefore, there is no point in increasing the weights for those two criteria, above a certain value, as it is shown, in many combinations, that the increased weight values affect very little both the spatial identification of Macedonian tombs and the area of the corresponding probability zones (Fig. 6, Fig. 7).

3. CONCLUSIONS

This paper described a general methodology used to create a predictive model that can be applied to indicate areas of archaeological sites’ occurrence. The aim was
to identify small areas that are of high probability to find archaeological sites and therefore need further investigation. In this sense, the model can contribute to archaeological research and particularly to historical topography and to a possible cost reduction by minimizing the requirements for trial excavations, as it indicates areas that are of high probability to crown an archaeological excavation with success. The methodology involves archaeological research, GIS and predictive modelling.

The methodology has been successfully applied in Macedonian Tombs location identification and numerous tests proved the efficiency of the method’s predictive ability.

By conducting archaeological research on a specific type of archaeological site and by customising the input data of the model (criteria, geospatial features related to the site), the adopted methodology could also result in successful predictive modelling for other types of archaeological remains.

Figure 6. (a) High and very high probability zone area as a function of the criteria weight for different zone ranges (b) Detected known tombs in high and very high probability zones as a function of the criteria weight for different zone ranges

Figure 7. (a) Very high probability zone area as a function of the criteria weight for different zone ranges (b) Detected known tombs in high and very high probability zones as a function of the criteria weight for different zone ranges
REFERENCES


