ACTA GEOGRAPHICA SZEGEDIENSIS
Tomus XXXVIII (2004)
http://www.geo.u-szeged.hu/acta

EXPERIMENTAL APPLICATIONS OF PREDICTIVE MODELING IN ARCHEOLOGY

FEJES, Csaba

Abstract

As Boast put it righteously "there is no basic difference in the way of human thinking between primitive and civilized cultures" starting off from the first intentional deeds, through the man of the medieval times up to the modern man of today. Settlements are generally established in areas where the environment is capable to support human needs. Thus the environment is an important influencing factor on the presence and distribution of humans, and this special relationship must be also quantifiable, analyzable in a given way.

The major aim of the present study was to shed light onto various phenomena and their interrelations, the collective appearance or presence of which may refer to the existence of a prehistoric, mainly archean archeological site nearby with great certainty in an area with measurable physical properties or parameters. With the help of the exposed relationships a category map can be prepared for a less investigated area, which depicts the probability for the presence of a human settlement or archeological site. The gained results may be a good starting point for planning future excavations or any other kind of research.

Thanks to the large-scale variety of data and procedures applied in the model, the collective usage of different analytical and statistical methods and approaches had to be utilized in our work ranging from the fields of geostatistics (logarithmic regression analysis) through geoinformatics (digital elevation model) to remote sensing (Landsat TM indexes).

Keywords: predictive modeling, GIS, geography, archeology

Introduction

There has been a really spectacular increase in the demand of getting to know our past and cultural heritage especially for the last few years. On the other hand, the available financial resources for widening our information background on the subject have experienced a relative drop.

This is truly alarming, as the detailed survey and analysis of even a small area may require huge sums of money. Another major problem comes from the fact that only a small portion of the known, plus the not yet discovered but existing archeological sites have been documented by archeologists so far, yielding some sort of data loss complemented by a continuous perishment of artifacts causing a similar problem.

All these alarming phenomena called for the development of special methods and approaches, which may on the one hand reduce the amount of money necessary for detailed field surveys via increasing the number of known sites on the other hand.

One such efficient approaches might be the construction of predictive models.

The practice of predictive modeling is quite common from the middle part of the 1980s, initially used only in medicine and business planning. However, it has recently becoming more and more widespread in archeological research as well. These models allow for the spatial and temporal expansion of known patterns and interrelationships. Thus they might be extremely useful in archeology as well.

Archeological predictive modeling is based on the assumption that human settlements generally develop in areas offering ideal natural endowments to meet the demands of human cultural groups on the one hand. On the other hand, it also presumes that traces of the former environmental conditions can be relatively well assessed even today as well, even if it requires the application of indirect methods mostly. The major purpose of the constructed model was to reveal these traces and those differences between them, which might have had a decisive role regarding the development of a human settlement or archeological site. In other words, the major task of archeological predictive modeling is to offer a reliable identification method of those natural and social factors, which had fundamental influences on human activities and regard the settlement as a feedback given to these factors by human communities.

The clear distinction between settlements and other areas is fundamental, giving the basis of the algorithms used in the model for the statistical classification of the studied parts of a pilot area via the measurable parameters of the surrounding environment and noting the possibility for the presence or lack of settlement sites in the pilot area. Consequently, these models are capable to predict the probability of the presence of such settlement sites in a given pilot area via the quantitative analysis of the environmental parameters. One of the major strength of the method is the utilization of explicit data and variables enabling the adequate reproduction and verification of the received results.

One of the practical benefits of the model is its applicability to extensive, less investigated areas, where the majority of archeological sites or former human settlement sites are not yet. Knowing the predicted distribution probabilities is useful for several reasons. With the help of the results of the model, one can get a clear view of not only the probabilities for the presence of archeological sites in a pilot area, but can receive useful information on the parameters, which might have actually influenced the development of the settlement itself as well. Furthermore, these models might be useful in planning regional management and development measures via highlighting the areas worthy for protection. As from 2001 during the course of a general land resource assessment (filing a conception of urban development, local construction regulations) the preparation of a detailed survey regarding the presence and distribution of cultural heritage values is compulsory in Hungary. Models with such scope may be extremely useful in these applications as well, since as one chapter of the bill says "all areas where the potential presence of archeological sites can be justified or presumed should be considered as archeological areas worthy of protection, including all natural and artificial ditches and watercourses as well". This is exactly what we are doing with the help of our predictive models. Furthermore, the gained results may be also useful in other research applications as well.

The present study is a clear example of the first experimental application of the above mentioned modeling method to local Hungarian examples. The results are mainly applicable to prehistoric sites.

Material and methods

The pilot area

The area chosen for detailed analysis is located in Békés county, at the interface of the southern marginal areas of the basin of the Hármas-Körös and the Békés-Csongrád Lowland, covering an area of about 130 km2 north of the city of Békéscsaba. (**Fig. 1.**) The most important geomorphological forms are two Pleistocene abandoned riverbeds of the ancient river Maros, presently signified as Hajdú-(Kamut) valley, and Kondoros brook, located east of the former. These channels were active riverbeds even during the Oldest Pleistocene as well (from 2.5 MA to 650 ka) corresponding to the northernmost margin of river migration as well.

The channels were not active during the Holocene, receiving water supply only from the rainfall and groundwaters forming a continuous water surface in these inactive riverbeds. It must be noted that during times of high floods, when floodwaters also managed to reach these channels, they might have been turned into active watercourses, acting as drainage channels on the floodplain.

The geomorphology of the Pleistocene was fundamentally determined by the alluvial fan deposits of the river Maros, composed of coarse sands and gravels. This setting fundamentally determines the general morphology and view of the referred landscape even today as well.

The former loess steppes, constituting the original natural vegetation have been replaced by extensive arable lands. Traces of the original vegetation are observable in some protected areas only, like the Cumanian burial mounds. And their former extents are known only from written historical sources.

Data utilized

The following types of data groups have been used in our work:

- Topographic data
- Spectral data
- Soil data
- Attributes

As a first step a digital elevation model depicting the relief conditions of the area was prepared. For this stereographic maps with a scale of 1:10000, frequently used by archeologists were scanned and transformed into the national uniform projection system (EOV). Afterwards, contour lines and elevation points were digitized and the received layers were used for the construction of the DEM using the software pack Arc/Info 8.1.

The archeological sites identified by archeologists during the course of field surveys were depicted on the scanned maps, so they just had to be transformed into a digital vector or raster format from the corrected digital maps.

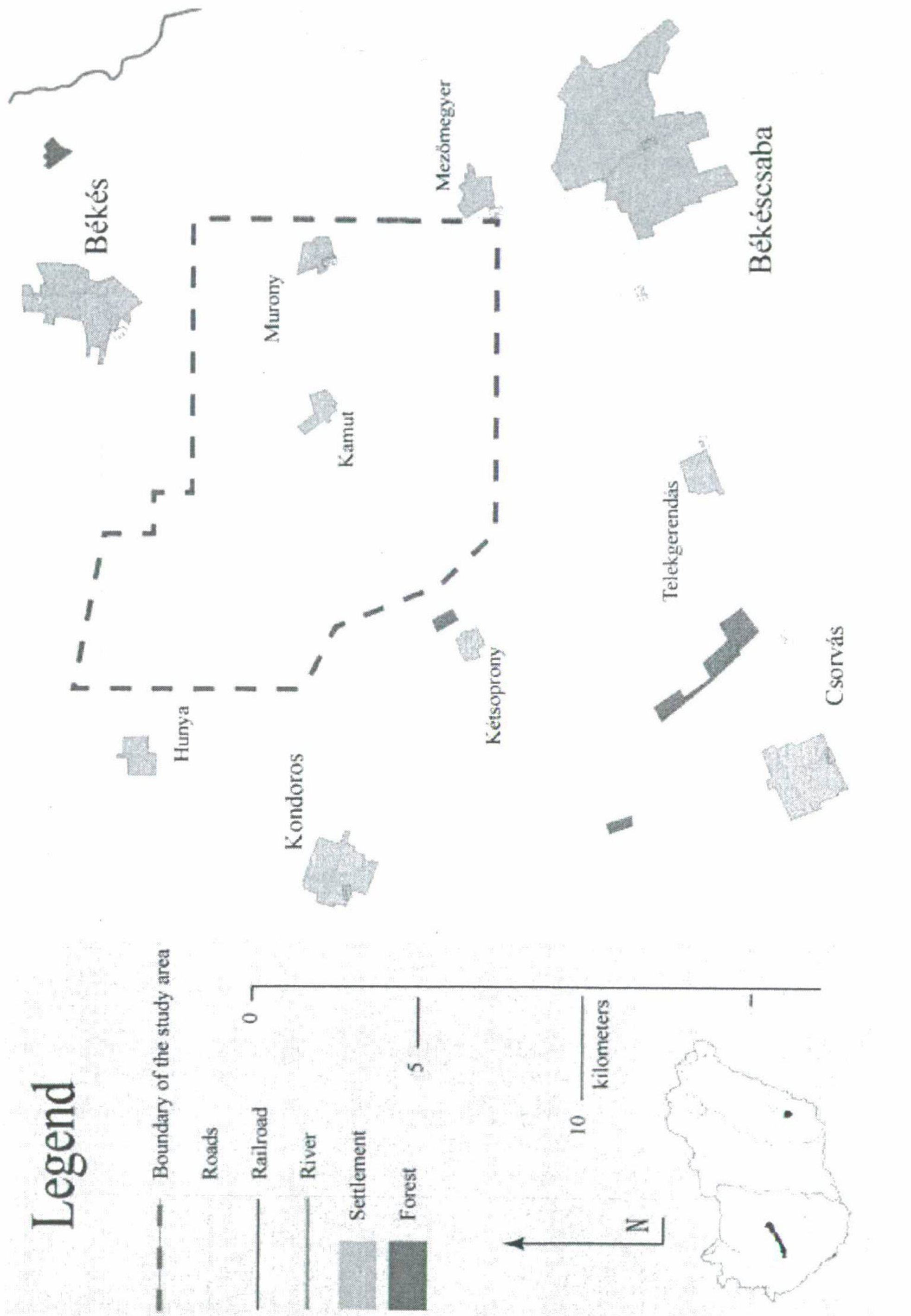

Figure 1. The location of the pilot area

The utilized spectral data, deriving from Landsat and TM satellite images correspond to such environmental parameters as soil temperature and humidity, as well as vegetation cover. These were received via the use of the ERDAS IMAGINE 8.4 software pack.

The low resolution (30 m) of the photos hampered the possibility of getting direct information on the location of the individual field objects. However, with the calculated indexes the individual areas could have been ordered regarding the probabilities for the presence of these objects.

The soil map layer was already in a digital format with a scale of 1:25000. The descriptive attributes of the model storing the name, age and code of the archeological sites derive from the archeological field reports and the volume entitled the Archeological Topography of Hungary.

Data processing was carried out in two steps. As a first step the raw data were transformed into a digital format, unless it was in that format, creating layers readable by the GIS software systems serving as starting layers for further analysis of the utilized environmental variables. These are the so-called primary layers.

The second step involved the construction of the so-called secondary layers, which contained only those variables deriving from the analysis of the primary layers, which directly participate in the modeling process.

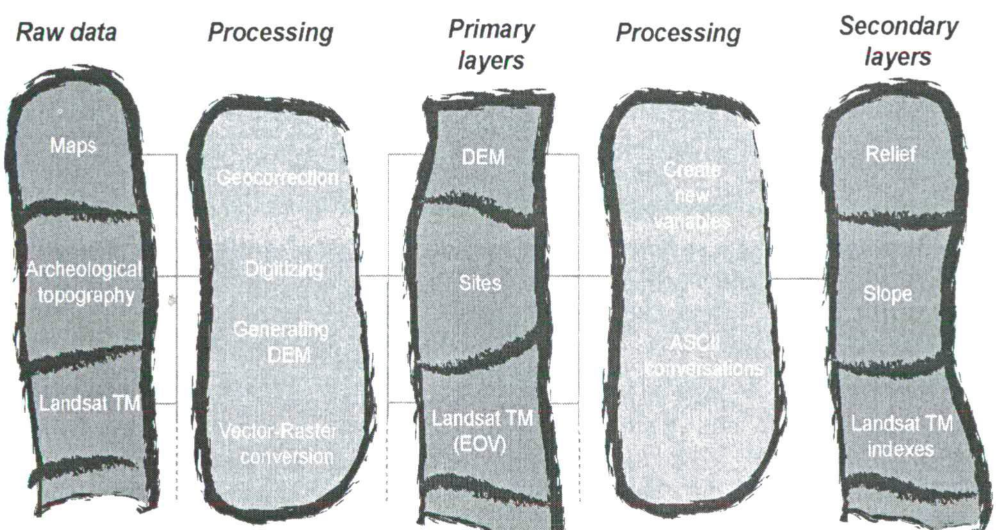

Figure 2.a The steps of predictive modeling (after Warren)

Raw data: maps, archeological topography, Landsat, TM photos; Primary Processing: transformation, digitizing, DEM construction, vector-raster conversion Primary layers: DEM, archeological sites, Landsat, TM (EOV) Secondary processing: new variables, ASCII conversion Secondary layers: relief, slope, Landsat, TM indexes

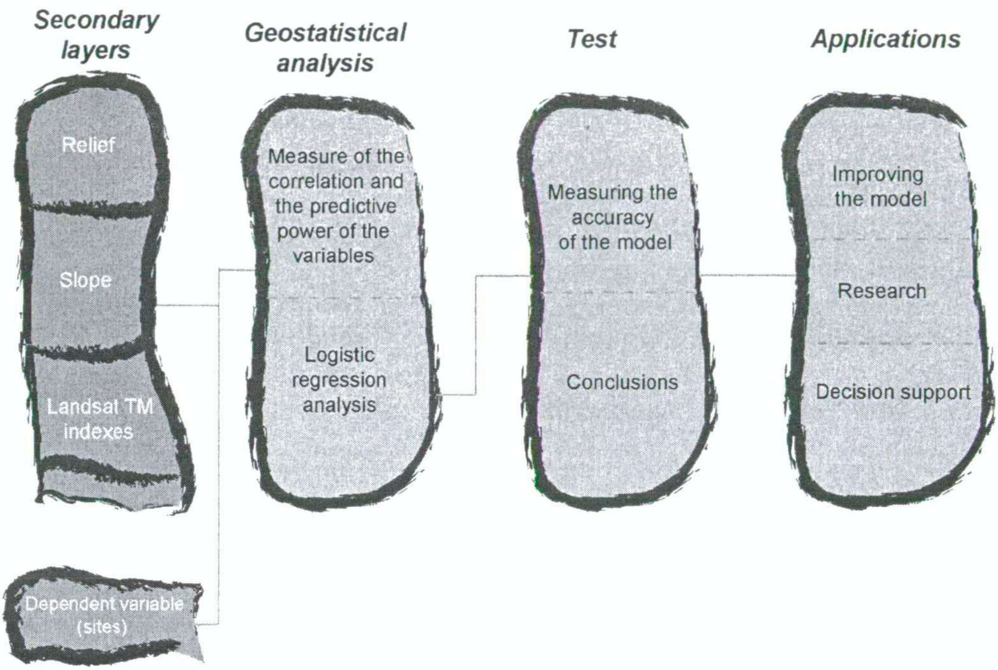

Figure 2.b The steps of predictive modeling (after Warren)

Secondary layers: relief, slope, Landsat, TM indexes, dependent variables (archeological sites) Geostatistical analysis: correlation and applicability, logistic regression analysis Testing: verification of predicted results, conclusions Usage: model improvements, research, support for decision making

Methods

The application of predictive modeling requires three major steps:

- Choosing dependent variables
- Choosing independent variables
- Assessing the relationship between these two variables

First let us have a look at what we can consider as dependent and independent variables.

Dependent variables were the raster layers containing information on the archeological sites, where the cells covering the sites were assigned a value of 1, while those cells not covering the sites received a value of 0. Cells located outside the pilot area had a value of NODATA.

The independent variables – in the factor of which we are examining the probabilities of the presence of sites- are the environmental parameters such as slope, exposure, relief, soil humidity, vegetation cover, genetic soil types, excess surface water inundation, and distance from watercourses.

Most likely, a lot of critics will let me know after the reading of this paper that there is not a single social or cultural component among these variables. And you know what they are absolutely right, at least, partly.

Since the cultural evolution of a human group is a very complex process, there is every reason to believe that nature and the parameters of the environment are important not only in the initial phase but during the whole process greatly influencing the shaping of thoughts and deeds. To put it in another way, in my opinion the characteristics of a culture as well as its demands are fundamentally determined by components of the environment giving the background of its birth as long as the culture is not modified by other external effects.

This must be applicable to hunting-fishing-gathering groups of humans as well as those dwelling on the steppes where the most common ornaments are all organic or naturalistic, referring to the former site and mode of the origin of that culture.

Of course, the non-natural parameters are also important. I just wanted to notify that their complete lack does not necessarily result in the total unreliability of our final results.

After the creation of the sufficient number of dependant and independent variables, their interrelationship should be assessed somehow in the next step of the analysis.

For this let us turn to the method of logistic regression analysis.

Logistic regression analysis

Several statistical methods have been utilized in the process of predictive modeling so far. However, the most popular and common of these is that of logistic regression analysis (ALLEN et al., 1990). This type of regression analysis developed from the method of linear regression. Regression analysis is the best tool for predicting the dependent variable values knowing those of the independent variables, in other words to reveal the relationship existing between two or more variables. In our case we were to shed light on the relationship between the archeological sites and the environmental parameters or components. In case of the logistic regression analysis the dependent variable is dichotomic – has dual value –, the best prediction is related to the probability of actual occurrence or happening, expressed as the logarithm of the quotient of the probability of occurrence and that of none occurrence (CSABA et al., 1997).

During this step our independent variables were introduced into an equation, which predicts the probability for the occurrence or non occurrence of an event in the studied area.

The equation of the process for several independent variables is:

$$p(B) = \frac{-1}{1 + Exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i)}$$

where p(B) is the probability of occurrence; Exp is the exponential function, α is the section constant, representing the value of the dependent variable, when x = 0; βi is the component of independent variables; xi is the independent variable for the suitable βi coefficient. The procedure uses the so-called maximum-likelihood method for calculating the values of α and β.

This analysis was implemented in the software environment of Arc/Info 8.1 yielding a raster map, where the value of the cells represented the probabilities for the presence of archeological site in that cell unit with values between 0 and 1.

For verification purposes before the second run about 40% of the archeological sites was left out of the analysis using a random function. Thus by applying these removed sites onto the received maps the accuracy of the predictions could have been easily observed.

Results

In the final maps depicting the prehistoric archeological sites the points with the highest probability values were almost all restricted to the areas of the ancient riverbeds (**Figure 3.**). These points managed to perfectly delineate the Kondoros creek. The area of the Hajdú valley was less precisely contoured, but could have been clearly identified.

This must be attributed to the fact, that the area of the Kondoros creek is in a relatively lower position compared to that of the Hajdú valley thus must have enjoyed more waters as well.

The actual sites relatively well correspond to the predicted values. Furthermore, with the help of these values the location of the sites, left out of the analysis for verification purposes, could also have been relatively "precisely" identified.

The scattered peak probabilities along the Kondoros creek must correspond to the former embayments as well as the deepest parts of the ancient channels. As the area was generally poor in surficial watercourses during the Holocene, open water areas must have been restricted to the deepest parts of the former channels only. So it's quite apparent why many sites turned up in these areas.

Calculated probability values, contour line of the sites

The calculated probability values mostly draw out the outline of the former channels and the once existing morphological features of the area (the correlation coefficient between the probability values and the relief values is 0.8).

Consequently, the natural endowments must have played a crucial role in the settlement strategy and location of the ancient prehistoric human communities.

The model was run for three other cases too, when sites of three distinct historical periods present in large numbers in the pilot area were individually depicted (the age of the Sarmatians, Avars and the Arpadian period).

However, only the presence of sites connected to the channels could have been clearly justified with this method. It must be attributed to the fact, that these cultures were not as much dependent on the natural endowments of the area as the formerly examined prehistoric archaic groups.

To improve the reliability of predictions, in these cases the social parameters must also be given consideration in the models.

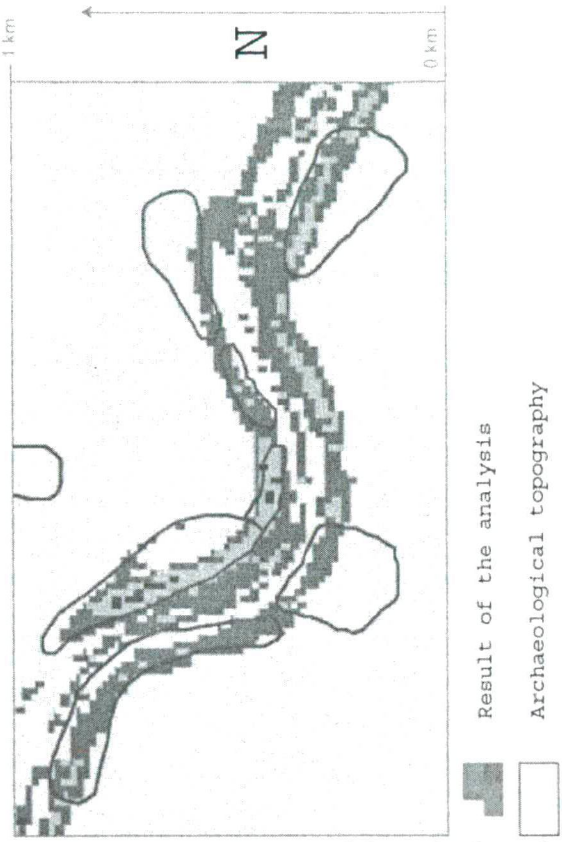

Figure 3. The final map with the location of the archeological sites depicted

Summary

On the whole the newly established predictive model was suitable for predicting the probability of the occurrence of archeological sites dated to the archaic prehistoric period. Thus the primary goal in the initial identification of these sites preceding archeological field surveys and understand the spatial characteristics of the identified settlement types was met for the period of the archaic cultures. However, in case of the younger cultural groups the results were not convincing and the model requires improvements.

There must be several reasons for this. It's possible that the quality of the raw data used for the calculation of the independent variables was not fully suitable for the task; like for example they derive from a Landsat image taken during a drought. Another possible source of error might result from the improper assignment of the sites during the field surveys, as in the practice of archeological topography even five pottery remains or fragments are considered to represent a site as well. Thus the exact delineation of such sites is utmost impossible via the application of the predictive method. Moreover, data on such sites are even harmful regarding the success of the outcome.

Despite all this, the utilized independent variables seemed to have yielded acceptable results in predicting the presence of archaic archeological sites. Since the actual physical parameters of the environment had the most decisive role in the settlement strategy during this time period.

In order to attain higher accuracy, the selection of the layers containing information on the natural environment should be carefully revised and tested. Furthermore, the introduction of the social and cultural parameters into the model may also enhance the reliability of the application rendering it suitable for use in the case of younger cultures as well in the future.

References

ALLEN, K.M.S. – GREEN, S.W. – ZUBROW, E.B.W. (eds) (1990) Interpreting Space: GIS and Archaeology. Taylor & Francis, London

ASCH, D.L. – WARREN, R.E. (2000) A Predictive Modell of Archaeological Site Location in the Eastern Praire Peninsula. In: Wescott, K.L. – Brandon R.J. (eds) Practical applications of GIS for archaeologists. Taylor & Francis, London

BÁRCZY, B. (1994) Differenciálszámítás. Műszaki Könyvkiadó, Budapest

CSABA, I. – GÁL, R. I. (1997) A bőség zavara: tökéletlen fogyasztói információ és verseny a háziorvosi szolgáltatások piacán. Közgazdasági Szemle, vol. XLIV.

KINCAID, C. (1988) Predictive Modeling and its Relationship to Cultural Resource Management Applications. In: JUDGE, W. – SEBASTIAN, L. (eds) Quantifying the Present and Predicting the Past. U.S. Government Printing Office, Washington.

KOHLER, T. A.(1988) Predictive Locational Modeling: History and Current Practice. In: JUDGE, W. – SEBASTIAN, L. (eds) Quantifying the present and predicting the past: Theory, method, and application of archaeological predictive modeling, edited by, U.S. Government Printing Office, Washinton, D.C..

JANKOVICH, B. D. (1987) Magyarország régészeti topográfiája. A békéscsabai járás.

FEJES, Csaba
University of Szeged, Faculty of Sciences
Department of Physical Geography and Geoinformatics
H-6701 Szeged, Egyetem u. 2-6., P.O.B. 653., Hungary
fcsaba@earth.geo.u-szeged.hu