Archaeological Predictive Modelling for Highway Construction Planning

Zoran Stančič, Tatjana Veljanovski, Krištof Oštir and Tomaž Podobnikar

Spatial Information Centre, Scientific Research Centre at the Slovene Academy of Sciences and Arts

Gosposka 13, 1000 Ljubljana, Slovenia

e-mails: zoran@zrc-sazu.si, tatjanav@zrc-sazu.si, kristof@zrc-sazu.si, tomaz@zrc-sazu.si

Abstract

The main objective of this paper is to present some results of the application and methodology of archaeological predictive modelling in Slovenian highway network planning. Recent analysis of the Slovenian highway construction budget has shown a considerable and constant increase in archaeological fieldwork expenses. According to state concern with cost diminution, we were asked to design a simple and effective methodology for archaeological site prediction. The results of a predictive model were used for consideration of the possible changes of highway location and optimal planning of archaeological fieldwork methodologies.

Key words: archaeological predictive modelling, highway planning, remote sensing

1. Introduction

Predictive modelling is fairly well established in North American archaeology. Interest in predictive modelling in Europe lags behind American trends by about one decade. Generally, predictive modelling in Europe tends to come within the purview of basic pragmatism. While the theory and methodology underwriting predictive modelling is conventional, practical applications of archaeological predictive modelling today can largely be considered within the context of cultural resource management. A great deal of predictive modelling work has already been carried out in the United States (Kvamme 1988, 1992a, Warren 1990, Kohler 1988) and Canada (Dalla Bona 1994), whilst significant applications have occurred in the Netherlands as well (Kamermans and Vansleeben 1999, Van Leusen 1996).

The strength of predictive modelling is that the methodology itself enables exploration and evaluation of locational factors. Therefore the knowledge of possible site location determinants can be tested and improved. What we propose here is that archaeological predictive models can also support highway planning. Integration of predictive models in highway planning is opportune in the early stages of planning (whilst selecting a highway corridor), here the results of predictive models provide immediate information for damage estimates to archaeological monuments. Secondly, predictive models can also be used during the construction stage, their main objective being integrate archaeological fieldwork methodologies with construction. Thirdly, results from predictive models can be used as a consolidated planning tool to define required fieldwork methodologies. Each of these propositions can play facilitate the efficient use of time, and promote cost efficient practice.

The Slovenian national plan seeks to complete and link highways in a network across the country. Prolongation of the building programme affects the public purse detrimentally. A recent examination of the Slovenian highway construction budget has shown a constant and considerable increase in overall costs, but it also disclosed that the principal increase in costs was related to archaeological fieldwork. The reasons for this increase are costs associated with unforeseen archaeological work which cause further delays in construction.

Concern over costs led led to an invitation to design a methodology for archaeological site prediction for a test section of a motorway corridor in the Pomurje region (along the Mura River), in northeastern Slovenia. The corridor measured approximately 11.2 km in length and archaeological prospection had already been carried out in the area including; extensive archaeological fieldwork, geophysical survey and air photo interpretation). The results from archaeological fieldwork could therefore also be used for evaluation of the predictive methodology.

2. Background

2.1. The Pomurje landscape

The Pomurje region lies in northeastern Slovenia, on the western edge of the Pannonian Plain. Here, the surface is gently dissected, with extensive lowlands surrounding the Mura River. The landscape opens to the east; and is interspersed with gentle terraces to the north and small hills towards the southwest. The area could be described as a monotonous landscape and this is reflected in its vegetation. Black alder woods flourish on, mainly, brown alluvial soils, while occasional hornbeam trees are found away from the flood plain. It is noteworthy that a very fertile plain, which is today almost entirely cultivated, lies between the two larger rivers – the Mura and Ledava. Distinguished from the other parts of Slovenia, this region is characterized by its stable geological structure. Non-agglutinate alluvial sediments belonging to the Quaternary and Pliocene compose the drift geology. Consequently, the area is dominated by gravels, sand and clay. Toward the hilly landscapes to the north clay based Holocene sediments emerge, and some light soils can be found in specific areas.

The Pomurje environment most probably always consisted of broad-leaved trees and widely dispersed damp meadows. These native environmental characteristics should be considered when attempting to understand the dynamics of past settlement in a region. Archaeological surveys have demonstrated that both the lowlands and uplands were settled. The distribution of settlements appears to represent a complex and extensive utilisation of the area.

2.2. The archaeology of the Pomurje region

The archaeological heritage of the region was not well known till the mid 19th century. Following this period, chance finds and sites were discovered and recorded in the local archives. Unfortunately, documentation of sites in the area was sporadic, and often lacks geographic coordinate attributes. Archaeological topographical efforts were initiated just after WW II, although a systematic approach to fieldwork was not then implemented. In 1975, the Institute of Archaeology at the Slovene Academy of Sciences and Arts published the first national Register of archaeological sites in Slovenia. Further archaeological research in the 1970s and 1980s confirmed that the Prekmurje region was continuously settled, although settlement intensity was highly variable (Šavel 1991). The national database of archaeological sites ARKAS (Tecco Hvala 1992, Modrian 1994) is currently being updated with recent records. Unfortunately, the level of preservation of the sites is not good. For instance, many prehistoric barrows in the region have been looted, and further damage has also been caused in recent decades by the use of machinery for cultivation.

Some 90 archaeological sites were included within this study. The chronological attribution of the sites ranges from the Bronze Age to the Early Medieval period. The types of sites also vary from isolated finds to settlements**.** In the Pomurje study, site locations were broken down into five distinct types to allow locational characterisation, and to allow us to recognize diversity in their distribution throughout the landscape.

2.3. The site's predicting methodology

A great deal of information on predictive modelling methodology, concepts and application have already been published (Judge and Sebastian 1988, Dalla Bona 1994, Westcott and Brandon 2000). It is important to emphasize that within the context of the Pomurje study, our goal could not be oriented towards the development of a model that would aid our understanding of the known distribution of the sites. Our aim was to develop a consistent model, which would assist in the process of highway planning.

Instead of presenting a theoretical background on predictive modelling methodology, we prefer to explain briefly what guided us in our choice of the method. Considerable problems often occur when the sample of sites analysed is small, as is the case in our study and are particularly obvious when multivariate statistics are used (Stančič and Veljanovski 2000). Some statistical evidence may be hidden in this process and the significance of specific variables also lost. Another issue issue is that most of the variables used are, naturally, based on the topographic characteristics of the area. An archaeological predictive model is most useful and reliable when it is applied in a region with a diverse natural environment. Surface topography plays a special role in predictive models as a consequence of its association to past land use. As the Prekmurje case study area is almost flat, it would be difficult, and misleading, to extract any rules, or predict potential, solely on the basis of the relief-based distribution of the sites. This situation applies not only to sites that vary in terms of their chronologies, but also between sites of different types. Despite the fact that a powerful predictive model can be made with a limited number of strong location factors,

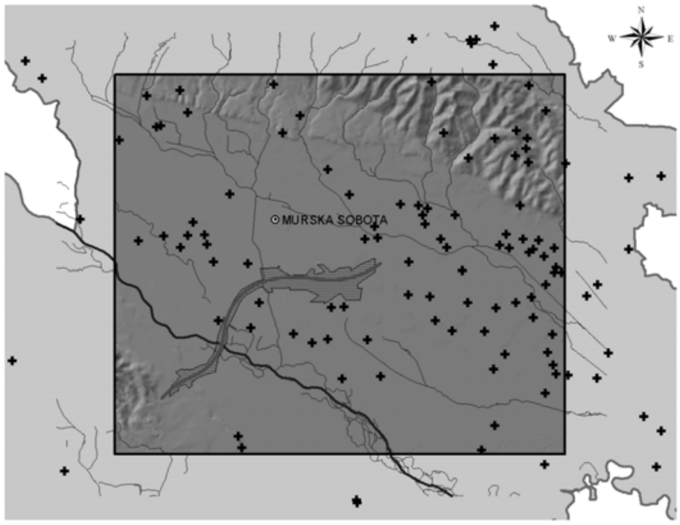

Figure 1: The Pomurje region and working environment for predictive model development together with the distribution of known archaeological sites (from National Database ARKAS) and anticipated location of highway section.

much effort was expended in the search for a range of other, potentially, relevant variables.

With regard to the situation described above and the objective of the case study, complex methods based on multivariate statistics failed to suit our data. Therefore it was finally decided to apply a simple traditional method of Boolean overlay to create a predictive model of site location in the region. As insufficient numbers of variables was a potential pitfall of the analysis, we additionally investigated the potential of remotely sensed data to the case study.

3. Data and modelling

The case study focused on a very flat part of the Pomurje region with little geomorphological change. Elevation across the study area as a whole ranges from c. 165 m in the lowlands to c. 315 m on the upland terraces. In the 11.2 km long highway test section, however, variation in elevation is only 16 meters. The lowlands sweep gently along the Mura River, and the terraces are crossed by a series of gently sloping valleys; some of these are also important sources of water. The study area comprises part of the northern upland terrace and extends about 21 km in an east-west direction and 18 km north-south, all together incorporating about 384 km^2 (figure 1). The basic units of analysis for the development of the model were regular square cells measuring 30 m.

We incorporated elevation, slope and aspect as variables representing the morphology of the landscape. As the natural environment is rather homogeneous, we decided that an accurate digital elevation model (DEM) would prove important, and we created a high-resolution DEM. A high resolution and accurate DEM are required for modelling settlement distribution in flat areas (Stančič and Oštir in press) and a DEM with a cell size of 25 by 25 metres, with a height accuracy of 3 m and a positional accuracy of 10 m was generated using radar images and radar interferometry (Oštir and Stančič in press).

Geological structure and data relating to the quality of soils can also serve as relevant variables representing the natural characteristics of the region. While geological factors are considered trustworthy, soils should be handled with more caution, as the

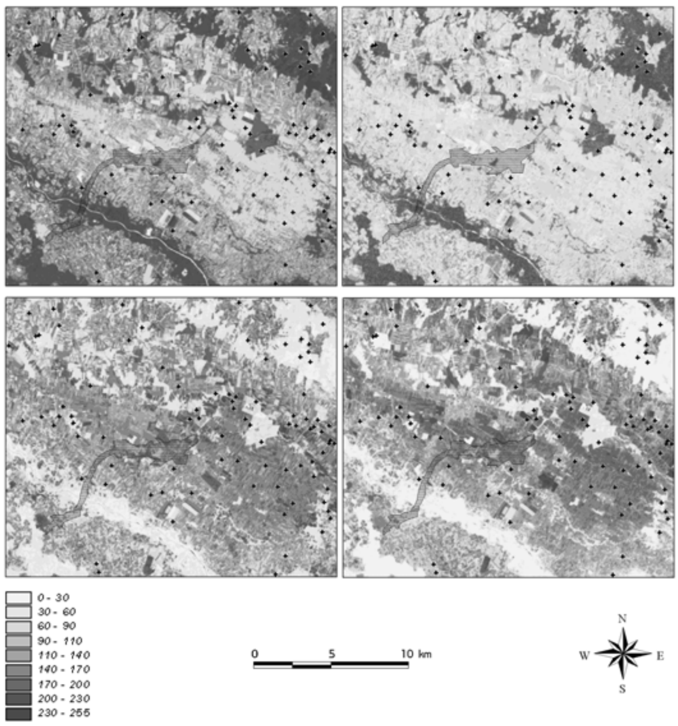

Figure 2: Vegetation index (upper left), Clay index (upper right), Ferrous Minerals index (lower left) and Ferric Oxide index (lower right) with distribution of archaeological sites and anticipated highway section.

data available usually expresses the most contemporary situation in the region. In addition to these data, a remotely sensed land use layer was created. Hydrology is also relevant. Given the case study goals, artificial ditches were not excluded from the hydrology vector layer, and these were considered "streams". Following this the distance to the nearest stream and distance to the main river within the region were calculated. Using these two layers, our aim was to enable an evaluation of the importance of the Mura River as a possible communication route.

The above variables are familiar and have been used in a number of locational studies (Kvamme 1992b). Given that our study area is a large plain, differences observed in variables such as elevation, slope and aspect could never play significant roles. Hence we were primarily concerned the variables that could provide a description of the lowland itself and this we felt was likely to be obtained from remotely sensed data.

Besides land use and a vegetation index, original information relating to surface soil type was attained through the generation of mineral delineation layers. These indices are extensively used in mineral exploitation and vegetation analyses, essentially because they can indicate minute differences between various rock types and vegetation classes (ERDAS Field Guide 1997). In many cases, judiciously chosen indices can highlight and enhance differences which cannot be observed in the original colour bands displays from satellite imagery. Therefore we produced three layers suitable for GIS analyses representing Clay Minerals, Ferrous Minerals and Ferric Minerals (Iron Oxide). Actually, these indices represent the presence of minerals on the surface (figure 2). In calculating these indices we hoped to obtain evidence for significant variation in the lowlands which was not be anticipated from the geomorphic variables.

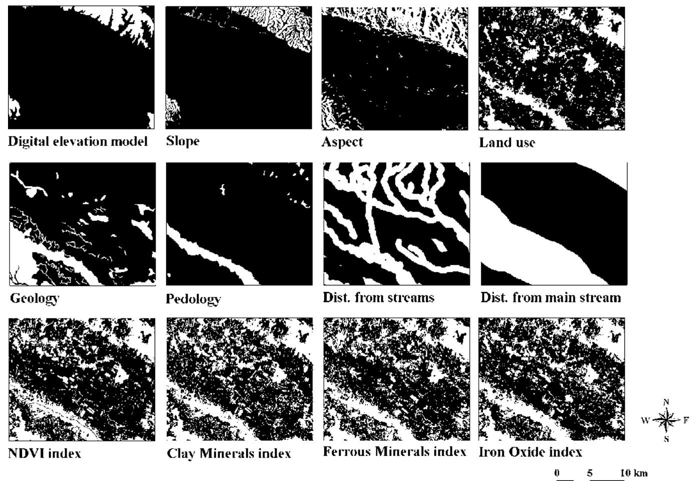

Figure 3: Binary layers representing prehistoric settlements proneness to twelve variables.

All map images were converted to a grid so as to create a series of primary layers portraying elevation, water streams, geology, soils, land use and archaeological site distribution. Some of the primary layers were later manipulated for derivative layers including aspect, slope, distance to the stream and distance to the river, as we hoped that these would help demonstrate how settlement behaviour related to the natural environment.

The majority of these variables used were in ratio scale. Some were later transformed into classes for further correlation testing. Several variables, such as geology, land use and quality of soils, were represented as a nominal scale.

Exploration and statistical analyses were carried out for five archaeological groups: prehistoric settlements (9), Roman and Early Medieval settlements (11), barrows (29), undated barrows (18) and isolated finds (mainly prehistoric axes, 15). A few of archaeological records were omitted from further processing, as it proved impossible to include them amongst these specified groups.

Univariate statistical tests (χ^2 and the Student *t*-test) were used to test the correlation between site locations and each of the variables. The results of univariate statistical tests indicate that different site types in the study area tend to occur in different environmental settings; however, from a statistical point of view, the correlation was quite tentative. Nevertheless, all the variables that were initially examined for significance were later included in the model, if they fulfilled "threshold" conditions. As two scale types were 5nv63ved, two criteria were introduced: a) for continuous variables, the interval between minimum and maximum values for sites was defined, and the circumscribed area considered as having potential, b) for nominal variables the area covered by classes associated with sites was treated as having potential. Twelve binary layers were prepared for each archaeological sample.

The initial model was designed as a simple sum of binary layers. Cells that fulfilled the entry condition for a particular variable were assigned a value of 1 and the rest a value of 0 (see figure 3). As a result, we attained five models for different site types, with a potential ranging from 0 to 12. Evidently, the score was a combined measure of the potential of the cell. For the purposes of overall compliance, the predictive model for the region was then obtained by merging the five model outputs.

On the basis of the preliminary analysis, we decided not to apply linear or logistic multivariate regression to define the relative weights of the variables. The weights are supposed to represent the relative relationship of locational factors within the past settlement system. As opposed to weights usually obtained through complex statistical examination, we believed that the weights most appropriate for a model of this kind should be based simply on the "sum of potential". Criteria were based on the assumption that a good predictor isolates highly specific and, therefore, a close or discrete signal in comparison to the values of the background. If a variable interpreted as a potential indicant also covered a large part of the study area then a smaller weight was assigned to that variable. Consequently, weights were applied in proportion to the observed power of the variable. As three mineral indices and one vegetation index seemed to be highly correlated we decided to give these additional weights. Each of their powers was diminished by a quarter and ultimately they represented – in their combined form – a single variable.

The final model (figure 4) was based on archaeological data attained from the national database and environmental data obtained from maps and satellite images. The social environment was poorly represented, as data , for example with respect to communications, were not reliable enough to allow prediction.

4. Results and testing

The validity of the resulting weighted model was tested by measuring the accuracy of its predicted potentials using two different samples,. The first was a sample of all archaeological sites that were used for the development of the model. This type of testing is known as internal testing and was carried out as a control of the process. Its results indicate that the model's prediction of potential was correct by about 81 %, if "high potential" incorporates areas with potentials ranging from 70 to 100 (figure 5). The remaining 19 % of sites fell within the limits of "medium potential" (40 to 70); no sites fell into the "low potential" range.

The data used for secondary testing were obtained through archaeological survey in the area of the proposed highway corridor. As most of this area has undergone cultivation for some time, survey was restricted to cultivated fields with good ground-surface visibility. The sites discovered were on the basis of artefact scatters located on the surface. Where a scatter was deemed promising the ground was intensively searched for artefacts and other traces of occupation sought through excavation of several 1 x 1m test trenches. Geophysical survey and aerial analysis of the area were also carried out prior to detailed work.

The second subset of archaeological data was therefore an actual independent testing sample as the sample was gained from recent archaeological survey. As these data were not at our disposal until the model was developed and tested with our own data, we can consider this second sample as valid for use for external testing of the model. As no excavations were carried out in the area during the period of development the extent of potential sites was not defined, and as a consequence we could only extract a centroid from the various areas that surface artefact scatters were located. These centroids are presumed to be representative of the these locations. As a scatter's area is sometimes quite extensive, two test samples were made: one with 13 principal centroids from 13 "sites" and the second one with 25 centroids derived from specified artefact densities within 13 "sites". It is important to note that

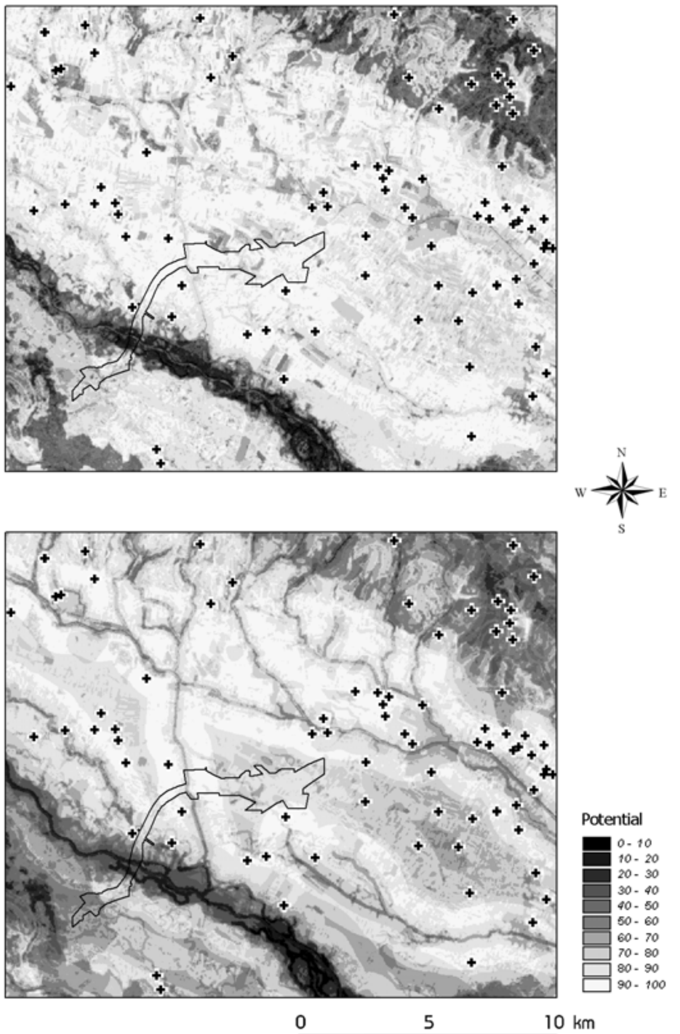

Figure 4: Archaeological predictive model for the Pomurje region: a) unweighted model (upper) and b) weighted model (lower) with the distribution of the sites upon which the model was developed.

the predictive model of the region was "windowed" to extract an area coincident with the independent archaeological survey, which was constrained to the highway corridor itself (figure 5). The results from both tests were similar. In the first case, six out of 13 centroids fell within the zone of highest potential (90 – 100) and the remaining seven just in the lower zone. The second test located 14 centroids within the highest potential zone, 10 centroids within the second highest range, and a single centroid fell within a lower range. A comparison of the statistical parameters within these tests could also be used as a technique to confirm the validity of the process. There is a substantial shift of both "sites" average potential within these tests. The predicted potential of the cells of sample centroids rises to 88 against a background average of 70.

Both tests indicate that the model's predictions are accurate. Although it is difficult to ascertain from the final model why particular sites within the region are situated within a specific area or why the potential of another site is lower, the model still distinguishes between locations where sites are most likely to be present and locations where they are not. The model does not, however, allow us to isolate the most important environmental or social determinants of site locations. In the end what we are providing to

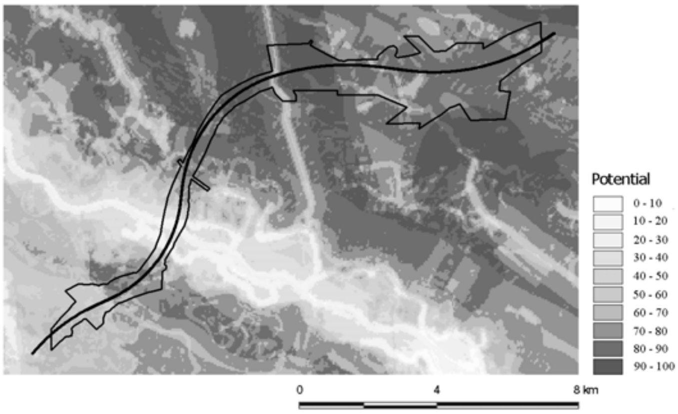

Figure 5: Predictive model of the Pomurje region (weighted model) showing area restricted to the highway location.

highway network planners is simply the expected distribution of archaeological resources.

Figure 5 illustrates the range of potential for site location. In studying this image it is important to note that in this "windowed" image, 63 % of the total land area can be coded as high potential, 26 % as medium potential and 11 % as low potential. This assumes that a range of 70 – 100 is considered as high potential, 40 – 70 is medium, and 0 – 40 has a low potential. Although there are clearly problems in this process the predictive power of such a model should still be useful for highway network planners. Ultimately, the relatively poor performance of the model may simply result from the geographical conditions that are far from suitable for predictive modelling. The manner of determining threshold conditions and the need to generate five smaller predictive models for specific site types may also mitigate against the model's efficient performance.

5. Conclusions

Although the results obtained from the model are good, several issues must be kept in mind when considering the results. The first issue, which is almost impossible to overcome, is the lack of comprehensive and reliable archaeological data. This is particularly problematic when creating an inductive model. The second point is that any generated prediction may be misleading because of the underlying dataset. The predictive model of the Pomurje region is ultimately based on the available, and essentially environmental, data. It does not take into account any of socio-political factors that must have played an important role in past settlement and land use decisions within the area.

Despite this the model does seem to provide highway planners with accurate information relating to areal archaeological potential. On this basis we believe that the model's predictive capabilities could be used in several ways:

- During the early stages of a highway planning (corridor study). Prior avoidance of a high-potential area could result in lower costs both for archaeological fieldwork and in terms of delay in highway construction.
- During the ongoing stages of highway construction, the model could be helpful in determining priority areas for intensive archaeological fieldwork. Although this would have to be in combination with results obtained through previous archaeological survey. The model could also identify low-potential areas where the construction process could be initiated with minimal archaeological supervision.

Acknowledgements

The Slovene National Agency for Highway Construction (DARS d.d.) funded the project backing this paper. The authors would like to extend their gratitude for all data provided concerning the highway section location, and especially for the results from recent archaeological fieldwork analyses, which were carried out by the Ministry of Culture, the National Institute for the Preservation of Cultural Heritage. A significant opportunity for independent testing of the predictive model was enabled by means of these recent archaeological results. Special thanks go also to the Institute of Archaeology, Scientific Research Centre at the Slovene Academy of Arts and Sciences, for accessing the national database of archaeological sites ARKAS.

References

DALLA BONNA, L., 1994. *Cultural Heritage Resource Predictive modelling Project*. Volume 3, Methodological Considerations. Centre for Archaeological Resource Prediction, Lakehead University, Thunder Bay.

ERDAS Inc., 1997. *ERDAS Field Guide TM*. Fourth edition. Atlanta, Georgia.

JUDGE, W., SEBASTIAN, L. (eds.), 1988. *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modelling*. U.S. Government Printing Office, Washington, D.C.

KAMERMANS, H. and VANSLEEBEN, M., 1999. Predictive modelling in Dutch Archaeology, Joining forces. In Barceló, J.A., Briz, I. and Vila, A. (eds.), *New techniques for old times, CAA98*, Barcelona, BAR International Series 757.

KOHLER, T.A., 1988. Predictive Locational Modelling: History and Current Practice. In Judge, W. and Sebastian, L. (eds.), *Quantifying the Present and Predicting the Past*: 19-61. U.S. Government Printing Office, Washington, D.C.

KVAMME, K.L., 1988. Development and Testing of Quantitative Models. In Judge, W. and Sebastian, L. (eds.), *Quantifying the Present and Predicting the Past*, 325-429. U.S. Government Printing Office, Washington, D.C.

KVAMME, K.L., 1992a. A Predictive Site Location Model on High Plains: An Example with an Independent Test: 19-40. *Plains Anthropologist* 37 (138).

KVAMME, K.L., 1992b. Terrain Form Analyses of Archaeological Location Through Geographic Information Systems. In Lock, G. and Moffet, J. (eds.), *Computer Applications and Quantitative Methods in Archaeology*: 127-136. BAR International Series 577.

LEUSEN, van P.M., 1996. GIS and Locational Modelling in Dutch Archaeology: A Review of Current Approaches. In Maschner, H.D.G. (ed.), *New Methods, Old Problems: GIS in Modern Archaeological Research*: 177-197. Center for Archaeological Investigations, Occasional Paper No. 23, Southern Illinois University.

MODRIAN, Z., 1994. Kataster arheoloških najdišč Slovenije (ARKAS) (II. del). *Arheo* 16, Ljubljana.

OŠTIR, K. and STANČIČ, Z., in press. Interferometric generation of DEM for mobile telephone network planning. In *Proceedings of the 'Fringe 99' Workshop on ERS SAR Interferometry*, Noordwijk, The Netherlands.

STANČIČ, Z., OŠTIR, K., in press. Producing Digital Elevation Models with Radar Interferometry. In Computer Applications and Quantitative Methods in Archaeology. *Proceedings of the CAA conference*, *Dublin 1999*.

STANČIČ, Z., VELJANOVSKI, T., 2000. Understanding Roman Settlement Patterns Through Multivariate Statistics and Predictive Modelling. In Lock, G. (ed.), *Beyond the Map: Archaeology and Spatial Technologies*. NATO Science Series, IOS Press Amsterdam: 147-157.

ŠAVEL, I., 1991. *Arheološka topografija Slovenije, Topografsko področje XX (Prekmurje).* SAZU, Ljubljana.

TECCO HVALA, S., 1992. Kataster arheoloških najdišč Slovenije ali zgodba o nastanku neke računalniške baze podatkov (I. del). *Arheo* 15, Ljubljana.

WARREN, R., 1990. Predictive Modelling of Archaeological Site Location: A Case Study in the Midwest. In Allen, K.M., Green, S.W. and Zubrow, E.B.W. (eds.), *Interpreting Space: GIS and Archaeology*: 201-215. Taylor & Francis, London.

WESTCOTT, K.L., BRANDON, R.J. (eds.), 2000. Practical Applications of GIS for Archaeologists – A Predictive Modelling Kit. Taylor & Francis, London.