Predicting Archaeological Site Locations Using G.I.S. and Fuzzy Logic

Eugenia Hatzinikolaou
Department of Geography and Regional Planning
National Technical University of Athens
Technical University Campus, Zographou, 157 80 Athens
Tel.: +30- 10-7722746
e-mail: gina@central.ntua.gr

Thomas Hatzichristos
Department of Geography and Regional Planning
National Technical University of Athens
Technical University Campus, Zographou, 157 80 Athens
Tel.: +30- 10-7722757
e-mail: thomasx@survey.ntua.gr

Angelos Siolas

Department of Geography and Regional Planning
National Technical University of Athens
Technical University Campus, Zographou, 157 80 Athens
Tel.: +30- 10-7722769
e-mail: angelos@survey.ntua.gr

Eleni Mantzourani
Department of History and Archaeology
University of Athens
University Campus - Zographou, 157 84 Athens
Tel: +30-10-7277467
e-mail:emantzou@arch.uoa.gr

Abstract. Techniques for prediction of archaeological site locations have been developed during the past decades as practical tools based on experts' knowledge and software's evolution, aiming to an effective cultural resource management. Based on the assumption that the choice of a site's location can be explained by the presence of certain environmental parameters, knowledge from already known archaeological sites, forms the basis in order to predict where else within a region with certain geographical characteristics, phenomena of inhabitance with similar features may have occurred. The likelihood of a site's appearance is crucial in large scale projects' planning as well as in a cultural and historical aspect, as precise and small scale archaeological research may miss valuable large scale information and site surveying may be time consuming and not so effective all the time. In this paper, an attempt to use Fuzzy Logic in order to predict possible locations of prehistoric sites is made. The methodology is presented through a case study in Melos, an island of the Western Cyclades - Greece. The data sets extracted from G.I.S analysis are incorporated in a Fuzzy Logic System in order to give the output predictive zones.

Keywords. Prediction, G.I.S., Fuzzy Logic

1. Introduction

Archaeological research has always been valuable for the preservation of a country's history and culture.

Due to its nature, this kind of research must be precise and cautious in order not to destroy what nature has kept so carefully within the years. Small-scale investigation may last for years over spot-areas to reveal all buried artefacts and structures. On the other hand, the constant expansion of our modern world, moving faster day by day, tends to eliminate free spaces not affected by human activities. Moreover, prehistoric archaeological sites are most of the times not visible or easily distinguished since they are buried or mostly destroyed.

Based on the above, knowledge of archaeological sites not yet documented, seems crucial in nowadays planning in order to achieve an effective cultural resource management.

The need of a methodological tool able to indicate zones determining the potential of a site's occurrence, has lead to the generation of several methodologies of archaeological prediction.

New technologies and software's evolution have played an important role to this. Especially, the outburst of Geographical Information Systems and their world wide application have given archaeological analysis a new perspective.(Allen *et al.*, (eds.), 1990; Lock and Stancic, (eds), 1995).

Up to nowadays, several models have been built attempting to classify parts of our world in "sites" and "not sites", opening the way to what has been called "predictive modelling" and is actually "...the practice of building models that in some way indicate the likelihood of archaeological sites, cultural resources, or past landscape use across a region..." (Kvamme, 2001).

Generally, these methodologies use a G.I.S package in order to handle geographical data and measure several environmental

variables, which are tested either by regression analysis or statistical tests for their contribution on a "sites" and "non sites" sample. The strongest among them are chosen to participate in the final model's equation, based on which, the rest of the geographical space is going to be classified in "sites" and "non sites" accordingly. (Wescott and Brandon, (eds), 2000)

Recently, other promising methods, as Multi-Criteria Analysis (Krist, 2001) and Neural Networks, have also entered the predictive modelling world.

In this paper, an attempt to use the new science of Fuzzy Logic in archaeological prediction is made. Fuzzy Logic, has been welcomed in many applications. The use of this new science in archaeology seemed a challenge, worth to be taken no matter the result.

Known, excavated sites provide the necessary knowledge base, which is the basic element in a Fuzzy Logic system in order to give the final output, which, in this case, is prediction.

G.I.S are necessary as they offer a powerful set of tools for the input, the maintenance and the presentation of the data, while the use of Fuzzy Logic is based on the need for appropriately treating environmental phenomena, which are not exact or precise but rather fuzzy. (Karkazi *et al.*, 2001)

2. Using G.I.S and Fuzzy Logic for the Prediction of Archaeological Site Locations

The basic assumption in the suggested methodology is that there are archaeological sites in the geographical space that have attributes: they are characterized as settlements or observatories or agricultural units or mining units, or other, and each category is dominated by certain geographical features. However, there is a class overlap, some of the features may exist in all categories to some degree, so certain areas could be either determined as settlements or as agricultural units or else, to a degree.

Undoubtedly, only by carrying out an archaeological excavation one can reveal the true identity of a site. By using Fuzzy Logic, this multiple ability of an area to belong in more than one classes can be properly treated. This gives archaeological prediction a new potential, as by defining different types of sites this method can contribute more efficiently in cultural resource management and in an hierarchically driven archaeological survey as well.

2.1 G.I.S: An analytical tool

G.I.S. are a box of tools for inserting, handling, and presenting geographical data (Hatzichristos, 1999). In order to predict archaeological site locations, G.I.S. are necessary for the collection of data and the generation of the final data set comprising of all the environmental parameters which will take part in the predictive analysis.

As prediction implies a degree of fuzziness itself, the use of Fuzzy Logic for the determination of areas that have a strong potential to be archaeological sites seems ideal as this approach is most suited to applications where decision criteria are not rigid, where the boundary between two regions is gradual. Inexact boundaries or class overlap appear to be more the rule than the exception in geographical problems. (Openshaw, 1997)

2.2 Fuzzy Logic: A New Perspective

Classic Boolean logic is binary, meaning that a certain element is true or false, an object belongs to a set or it doesn't. Fuzzy Logic introduced by Zadeh in 1965 permits the notion of nuance. Apart from being true, a proposition may be anything from almost true to hardly true (Kosko, 1991). In comparison with the Boolean sets, a fuzzy set does not have sharply defined boundaries. The notion of a fuzzy set provides a convenient way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables.

As mentioned, a significant fact about statistical logic is the defect that each point of a set U is unequivocally grouped with other members of its group and thus bears no similarity to members of other groups. One way to characterize an individual point's similarity to all the groups was introduced in 1965 by Zadeh. The key to Zadeh's idea is to represent the similarity a point shares with each group with a function (termed the membership function) whose values (called memberships) are between 0 < m < 1. Each point will have a membership in every group, memberships close to unity signify a high degree of similarity between the point and a group while memberships close to zero imply little similarity between the point and that group. Additionally the sum of the memberships for each point must be unity.

Although in Boolean logic A and not A are unique, in Fuzzy Logic the following equation is true:

$$m_{notA} = 1 - m_A$$

Fuzzy degrees are not the same as probability percentages. Probabilities measure whether something will occur or not. Fuzziness measures the degree to which something occurs or some condition exists. Crisp sets are a subset to fuzzy sets. Only when an object belongs 100% to a group fuzzy sets are identical to crisp sets.

In order to solve a problem with a knowledge-based fuzzy system it is necessary to describe and process the influencing factors in Fuzzy terms and provide the result of this processing in a usable form. The basic elements of a knowledge-based fuzzy system are:

- 1. Fuzzification
- 2. Knowledge base
- 3. Processing
- 4. Defuzzification

These elements are described in detail, in the following paragraphs.

Every continuous math function can be approximated by a fuzzy set.

For example the criterion "distance from a road" can be approximated from the following membership function:

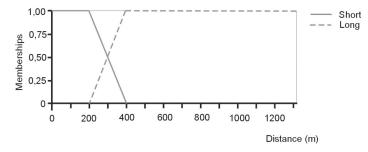


Figure 1. Membership function for the "distance from a road".

Several types of membership functions can be utilized (Bezdek 1981; Burrough, 1996). The membership function reflects the knowledge for the specific object or event.

The assignment of a membership function to every variable of the problem is called fuzzification process. During this process crisp subsets are transformed to linguistic subsets such as short or long distance (Fig. 1). The concept of the linguistic variable illustrates particularly clearly how fuzzy sets can form the bridge between linguistic expression and numerical information.

The second step in the fuzzy systems methodological approach is the definition of the rules, which connect the input to the output. These rules are based on the form "if ...then and". The knowledge in a problem-solving area can be represented by a number of rules. For example, if the output set "suitability" is comprised by two subsets called "poor" and "appropriate", the rules could be:

- If the distance is short then suitability is poor
- If the distance is long then suitability is appropriate

Experts with general knowledge on the specific field, usually accomplish the task of rules definition. There is no need for assigning weights in the criteria used. The weights are indirectly taken in account through the rules defined.

The next step is the processing of the rules. This step is also called inference. It comprises the three stages, aggregation, implication and accumulation. Aggregation provides the degree of fulfillment for the entire rule concerned. All the Boolean algebra operations (like intersection, union, negation, etc) can be easily extended to fuzzy set operations (Kandel, 1986) and they can be used in this stage. In implication the degree of fulfillment of the conclusion is determined. Accumulation brings together the individual results of the variables used. Details for this process can be found in Bezdek (1981).

The result of rules processing can be transformed, if necessary, back into a linguistic expression or a crisp value. This second process is called defuzzification and there are several methods to achieve it (Bezdek, 1981).

5. Case Study: Predicting Prehistoric Site Locations on the Island of Melos

Melos is a Greek island stuated at the western Cyclades, a complex of islands in the south Aegean Sea.

Special environmental parameters characteristic in all Cycladic islands, as climate, landscape - dominated by the mountains and the sea – and natural resources, are present also in Melos. Being a largely volcanic island in addition, Melos played a very significant role in past societies, as it has been the most important obsidian source of the Aegean, supplying not only the neighbor islands but also the mainland of Greece.

Moreover, on Melos the town of Phylakopi was discovered, which is up to now the best documented prehistoric site of the Aegean (Renfrew and Wagstaff, (eds), 1982).

The island of Melos has been chosen among the other Cycladic islands to realize this pilot research, due to the fact that on Melos an extensive site survey has been carried out during 1974-77 by the British School of Archaeology, offering a valid reference in order to check the results of the method proposed.

The main assumption forming the basis of this project, was that the choice of a site's location by prehistoric people must have been based on logical thoughts about the degree it could fulfill certain needs of its inhabitants. Those needs were affected by the period (prehistory) as it concerned natural fortifications, protection, easy or difficult accessibility, by the island's characteristics (obsidian mining, external trade, agriculture) and certain sociological parameters as the dominating presence of Phylakopi and the relations deriving from it.

The sites of the island were grouped in two major categories, according to the classification of the site survey ¹ in order to be able to compare the results:

- Settlements and
- Special purpose sites, including agricultural units, mining units and observatories.

In order to achieve the desired output which is the prediction of prehistoric archaeological sites of the two above-mentioned categories with the use of Fuzzy Logic two things were crucial:

- The accurate choice of the input data layers corresponding to the criteria set by the analysis
- The experts' knowledge on prehistoric Cycladic sites, on the geography of the Cyclades and on the method used

The input data layers used (scale 1:50.000) were chosen based on the criteria set by the analysis to be crucial on a site's existence (Table 1). Most of the layers were digitized in vector format and some of them were processed in order to give the desired output layer. The software used was the ARC/INFO 7.2.1 and ARCVIEW 3.2. for the G.I.S. analysis, in Unix and Windows 2000 Professional Environment.

The experts' knowledge on prehistoric Cycladic sites and on the geography of Cyclades was provided by the group carrying out this project consisting by Geographers, G.I.S specialists and an Archaeologist, all having special knowledge on the specific field.

3.1 G.I.S Analysis

The criteria decided to be determining for a site's existence corresponding to the primary layers which had to be digitized, are listed in the following table along with their primary data sources.

All vector coverages were converted in raster format. The pixel size was decided to be 12,5m X 12,5m corresponding to the necessary accuracy (1/4 mm of the scale 1:50.000).

¹ See Cherry F., "Apprendix A: Register of Archaeological Sites on Melos", in Renfrew C., Wagstaff M., eds., 1982, *An Island Polity. The archaeology of exploitation in Melos*, Cambridge University Press

Criterion/ Layer	Primary Data Source
Streams	Topographical Maps of the Hellenic Military Service (Scale 1:50000)
Springs	Topographical Maps of the Hellenic Military Service (Scale 1:50000)
Slope	Original Films, Brown Color (contour lines) of the Hellenic Military Service (Scale 1:50000)
Agricultural Land	Corine.shp, Hellenic Chartographic and Cadastral Organization
Obsidian sources	X,Y coordinates (GPS)
Hills	Topographical Maps of the Hellenic Military Service (Scale 1:50000)
Bays	Topographical Maps of the Hellenic Military Service (Scale 1:50000)
Capes	Topographical Maps of the Hellenic Military Service (Scale 1:50000)
Phylakopi	X,Y coordinates (GPS)

Table 1. Criteria and Primary Data Sources.

For the criteria "Streams", "Springs", "Agricultural Land", "Obsidian Sources", "Bays" and "Capes", Euclidean distances were calculated, as proximity to these features was the crucial factor.

The final grids (Fig. 2) were combined in order to produce the final output file to be inserted in the Fuzzy Logic System.

3.2 Fuzzy Logic Analysis

For the Fuzzy Logic Analysis, the software used was the Data Engine 2.0, of the M.I.T. University.

The first step, fuzzification, was actually the assignment of a membership function for all the criteria.

The membership function for every criterion as well as its linguistic expression based on the experts knowledge are illustrated on the following table:

Criterion	Linguistic Expression	Membership function			
Streams	Distance Short(X) Long(X)	Long(x)= -	0, if distance(x)<800 Distance(x)-800/700 if 800 [distance(x) [1500 1, if distance(x) > 1500		

Criterion	Linguistic Expression		Membership function	
Spring	Distance Short(X) Long(X)	$Long(x) = \left\langle \right.$	0, if distance(x)<1500 Distance(x)-1500/2500 if 1500[distance(x) [4000 1, if distance(x) > 4000	
Slope	Smooth(X) Steep(X)	Steep(x)=	0, if steep(x)< 20% steep(x)-20%/30% if 20% [steep(x) [50% 1, if steep(x) > 50%	$\left. \right\}$
Agricultur al Land	Distance Short(X) Long(X)	Long(x)=	0, if distance(x)<500 Distance(x)-500/1500 if 500[distance(x) [2000 1, if distance(x) > 2000	}
Obsidian sources	Distance Short(X) Long(X)	Long(x)=	0, if distance(x)<1500 Distance(x)-1500/2500 if 1500 [distance(x) [4000 1, if distance(x) > 4000	$\left. \begin{array}{c} \\ \end{array} \right\}$
Hills	Hills	Hills(x)=	0, if no hills 1, if hills	}
Bays	Distance Short(X) Long(X)	Long(x)=	0, if distance(x)<500 Distance(x)-500/2500 if 500 [distance(x) [3000 1, if distance(x) > 3000	}
Capes	Distance Short(X) Long(X)	Long(x)=	0, if distance(x)<100 Distance(x)-100/400 if 100 [distance(x) [5000 1, if distance(x) > 500	$\left. \begin{array}{c} \\ \end{array} \right\}$
Phylakopi	Visibility	Vis(x)=	0, if not visible 1, if visible	}

Table2. Linguistic Expressions and Membership Functions for every Criterion.

The membership functions in graphics format for two of the above criteria are illustrated in Figure 3 and Figure 4.

The output variable of the fuzzy system is the site's existence comprised of two subsets, settlements and special purpose sites.

At the end of the process the whole island will be assigned with a value between 0 and 100 for

- the existence of settlements
- the existence of special purpose sites

In order to achieve the above, it is necessary to define the rules connecting the input values through their membership functions with the output subclasses.

The rules are formed in a "if...then" format and they reflect the knowledge of the experts. The percentage of certainty is assigned by the group of experts as well and is also reflecting their knowledge and experience not only on the archaeology of the island but on its geographical characteristics as well.

The rules, which have been formed, are:

- 1. If distance from a stream is short and distance from agricultural land is short and slope is smooth then \rightarrow it is a settlement with 60% certainty
- 2. If distance from a stream is short and distance from a bay is short and slope is smooth then → it is a settlement with 80% certainty
- 3. If is a hill and distance from a bay is short then → it is a settlement with 70% certainty
- 4. If distance from a spring is short and distance from agricultural land is short and slope is smooth then → it is a settlement with 60% certainty
- 5. If distance from a stream is short and distance from agricultural land is short and it is a hill and it is visible from Phylakopi then → it is a settlement with 90% certainty
- 6. If is a hill and distance from a bay is short and it is visible from Phylakopi then → it is a settlement with 80% certainty
- 7. If distance from obsidian sources is short and distance from a stream is short and distance from agricultural land is short then → it is a settlement with 80% certainty
- 8. If distance from a cape is short and distance from a stream is short and slope is steep then → it is a settlement with 75% certainty
- 9. If is a hill and distance from agricultural land is short and distance from a stream is short then → it is a settlement with 85% certainty
- 10. If is a hill and distance from agricultural land is short and distance from a spring is short then \rightarrow it is a settlement with 85% certainty
- 11. If distance from a spring is short and distance from agricultural land is short and distance from a bay is short then → it is a settlement with 80% certainty
- 12. If distance from agricultural land is short and distance from a stream is short then → it is special purpose site with 80% certainty
- 13. If distance from obsidian sources is short then \rightarrow it is special purpose site with 90% certainty

- 14. If distance from a cape is short and it is visible from Phylakopi then → it is a special purpose site with 85% certainty
- 15. If distance from agricultural land is short and slope is smooth and distance from a spring is short then \rightarrow it is a special purpose site with 85% certainty

The next stage, inference, which is actually the processing of the rules, was carried out using the following operators:

- Aggregation operator : Minimum
- Implication operator : Algebraic product
- Accumulation operator : Maximum

The product of inference was transformed into an ASCII file, which consisted of 12 columns and 98.000 rows. This file was joined with the INFO file of the final grids' combination in order to give the final output in the G.I.S. environment.

The results for settlements are presented in Map1 (Fig. 5) and the results for special purpose sites are presented in Map 3 (Fig.7). Pixels with values close to black (100) represent predictive areas where settlements or special purpose sites could be located. As pixels fade away towards white (0) the existence of the two above-mentioned categories becomes weaker.

In order to check the output of the suggested methodology, the results of the site survey held by the British School of Archaeology were used. The map with the sites² was geo – referenced and only sites belonging to the prehistoric period characterized as "habitation" or "special purpose", were digitized in two separate layers.

Results were overlaid and presented in Map2 (Fig.6) and Map4 (Fig.8).

It is obvious that there is a strong correlation of the two results, proving the accuracy of the method, which has a strong potential to be evaluated with the addition of more environmental criteria. Moreover, its application to a group of islands where factors as marine routes and visibility among them will alter the accuracy of the results.

4. Conclusions

The use of G.I.S. and Fuzzy Logic for the prediction of archaeological site locations could be a powerful tool in modern planning and archaeological research. The most important advantages are:

The final output is gradual, based on logical assumptions, reflecting realistic conditions under which people in the past would have thought in order to make a choice on a site's location. General knowledge used by experts, is coming from excavated cycladic sites of the same period, which is rather different from just some statistical results of a site's survey sample.

The method reflects realism through linguistic variables

The final output is classified meaning that it is possible to specify the characteristics of the output (settlement, special purpose site or else) by running the procedure one single time.

² See Cherry F., "Apprendix A: Register of Archaeological Sites on Melos", in Renfrew C., Wagstaff M., eds., 1982, An Island Polity. The archaeology of exploitation in Melos, Cambridge University Press

The final output consists of more alternatives, which are lost in statistics' analysis.

The results are ranked hierarchically in the geographical space.

It is time saving as it provides a unique predictive result for the whole area, while traditional methods of site surveying last long, meaning that they cost in time and money, and many times could not be so effective as accessibility may be very difficult or even impossible. Of course the method has disadvantages which could be mostly characterized as drawbacks and not as defects of the method itself:

The lack of ready to use membership functions. Of course, the same membership functions used for this project, corresponding to another location, for the same period, could be used.

In order to use the method as described in this paper it is necessary to posses high level knowledge which is something gained and not easily found.

The incorporation of Fuzzy Logic Analysis in G.I.S packages is a challenge nowadays. This accomplishment could make the suggested methodology even easier for non-experienced users, taking as a fact that both software are user friendly.

The use of G.I.S and Fuzzy Logic in the prediction of archaeological site locations although evolutionary, could be an alternative method contributing significantly in archaeological research which needs such data especially in countries with a rich historical past.

References

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

BEZDEK C.J., 1981, Pattern Recognition with Fuzzy Objective Function Algorithms, New York, Plenum Press

BURROUGH P., 1996, "Natural Objects with Indeterminate Boundaries", in Burrough P. and Frank A. (eds.) *Geographic Objects with Indeterminate Boundaries*, Taylor & Francis, 3-29

HATZICHRISTOS T., 1999, Delineation of Ecoregions using GIS and Computational Intelligence, PhD Thesis, National Technical University of Athens (in Greek)

KANDEL A.,1986, Fuzzy Mathematical Techniques with Applications, New York, Addison-Wesley

KARKAZI A., HATZICHRISTOS T., MAVROPOULOS A. EMMANOUILIDOU A. and ELSEOUD A., 2001, "Landfill Siting Using GIS and Fuzzy Logic", *Eight International Waste Management & Landfill Symposioum*, Sardinia, October 2001

KOSKO B., 1994, Fuzzy Thinking, the new science of Fuzzy Logic, Flammingo Press

KRIST F., 2001, "Multi-Criteria Predictive Modeling: A Tool for Archaeological Problem Solving", GIS and Archaeological Predictive Modeling Conference: Large-Scale Approaches to Establish a Baseline for Site Location Models, Argonne National Laboratory, Chicago, March 2001, (Forthcoming)

KVAMME K. L.,2001, "There and Back Again: Views on the Development of Predictive Location Modeling in Archaeology", GIS and Archaeological Predictive Modeling Conference: Large-Scale Approaches to Establish a Baseline for Site Location Models, Argonne National Laboratory, Chicago, March 2001, (Forthcoming)

LOCK G. and STANCIC Z., (eds.), 1995, Archaeology and Geographical Information Systems: A European Perspective, Taylor & Francis

OPENSHAW S., 1997, Artificial Intelligence in Geography, Whiley, London

RENFREW C. and WAGSTAFF M., (eds.), 1982, *An Island Polity. The Archaeology of exploitation in Melos*, Cambridge University Press

WESCOTT K.L. and BRANDON R.J, (eds.), 2000, Practical Applications of GIS for Archaeologists. A predictive modeling kit, Taylor & Francis

ZADEH L.A., 1965, "Fuzzy Sets", *Information and Control8* 338-353.

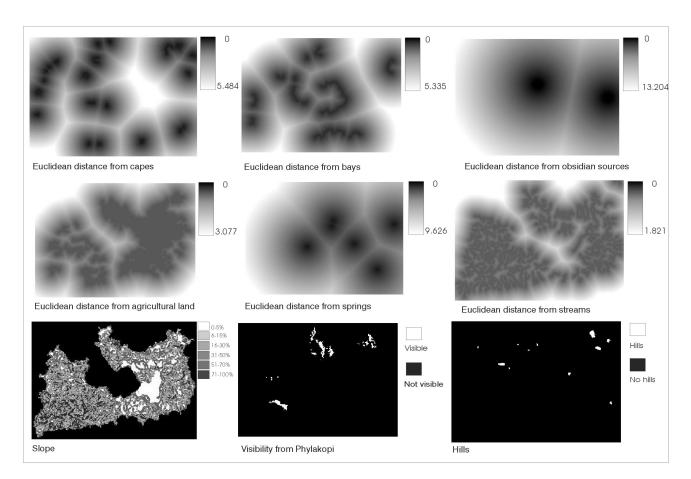


Figure 2. Final Grids

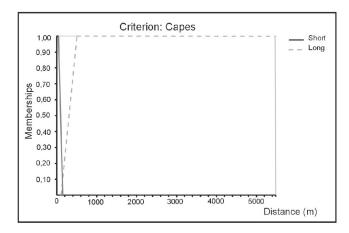


Figure 3. Membership function for the criterion "Capes"

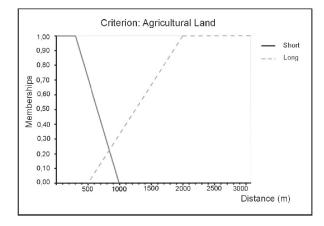


Figure 4. Membership function for the criterion "Agricultural Land"

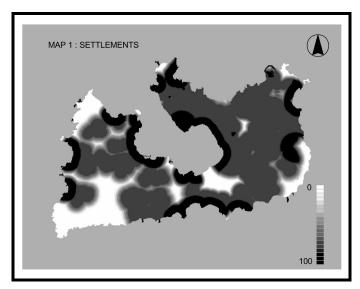


Figure 5. Predictive map for settlements.

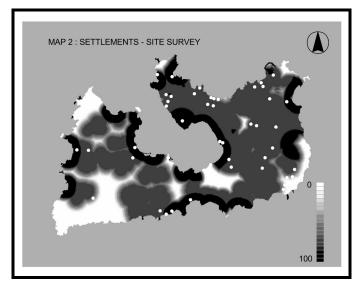


Figure 6. Predictive map for settlements compared with the results of the site survey.

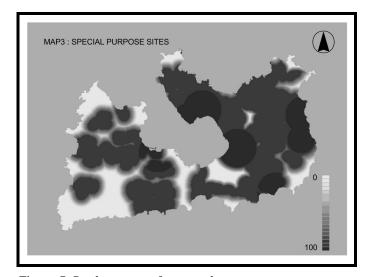


Figure 7. Predictive map for special purpose sites.

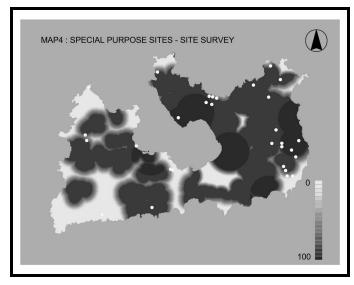


Figure 8. Predictive map for special purpose sites compared with the results of the site survey.