Landscape Archaeology between Art and Science
Guttmann-Bond, Erika, Kluiving, Sjoerd J.

Published by Amsterdam University Press

Guttmann-Bond, Erika and Sjoerd J. Kluiving.
Landscape Archaeology between Art and Science: From a Multi- to an Interdisciplinary Approach.

Amsterdam University Press, 2012.
Project MUSE. muse.jhu.edu/book/66303.

For additional information about this book
https://muse.jhu.edu/book/66303

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=2321066
Mapping the probability of settlement location for the Malia-Lasithi region (Crete, Greece) during the Minoan Protopalatial period

Authors
Ricardo Fernandes¹,², Geert Geeven³, Steven Soetens⁴,⁵ and Vera Klontza-Jaklova⁶,⁷

1. Leibniz-Laboratory for Radiometric Dating and Isotope Research, Christian-Albrechts-Universität, Kiel, Germany
2. Graduate School “Human development in landscapes”, Christian-Albrechts-Universität, Kiel, Germany
3. Department of Mathematics, Faculty of Sciences, VU University Amsterdam, Amsterdam, The Netherlands
4. Institute for Geo- and Bioarchaeology, Faculty of Earth and Life Sciences, VU University Amsterdam, The Netherlands
5. Research Institute for the Cultural Landscape and Urban Environment, VU University Amsterdam, Amsterdam, The Netherlands
6. INSTAP-EC, Pacheia Ammos, Crete, Greece
7. IIHSSA Priniatikos Pyrgos Project, Crete, Greece
Contact: rfernandes@gshdl.uni-kiel.de

Abstract
The current study considers a mixed environmental/historical statistical model to establish a probability map for settlement locations in Crete’s Malia-Lasithi region during the Minoan Protopalatial period. The work represents the continuation of previous research that focused on site location choices during the Protopalatial and whereby a comparison was made between the performances of a purely environmental over a mixed environmental/historical model. Statistical modelling consisted of fitting a logistic regression model using a Deletion/Substitution/Addition (DSA) algorithm for model selection. Model uncertainty was assessed through calculation of confidence intervals at the 95% confidence level and the results are presented as probability maps that show upper and lower interval endpoints for the study area. Assessment of the model’s predictive performance, on both the study area and on an independent validation area, indicates that the model is able to capture some underlying structure that determines preferences for site locations. Moreover there is a general agreement between the generated settlement probability map and many of the existing published survey results. The results obtained demonstrate the usefulness of the modeling approach and we expect that the existing model can be further improved in the future by incorporating more survey data.
INTRODUCTION

Archaeological predictive modelling (APM) is often used with the underlying meaning of predictive modelling applied to heritage management. For this reason predictive modelling is the focus of much controversy and debate. The main issue of this debate is whether predictive modelling should be used as a tool in establishing a heritage managing policy. The purpose of this paper is not to enter such a debate but rather to emphasise that the role of predictive modelling is not limited to heritage management.

Foremost predictive modelling provides a systematic approach to understand settlement location choices of past populations. With the use of a good theoretical framework based upon the existing archaeological data it then becomes possible to pursue appropriate statistical modelling. A major benefit of a statistical model is that it can be used to both systematically establish the major factors determining site location and to define probability maps that can subsequently be used to guide survey strategies which may further validate the statistical model.

The results presented in this paper represent an extension of previous work (Fernandes et al. 2011) that was strictly focused on the understanding of past location choices for Minoan populations during the Protopalatial period (1900-1650 BC) in central-eastern Crete. That research was able to establish a model with an excellent statistical fit and a sound archaeological interpretation. One of the major results was the establishment of a historical/environmental model in which the proximity to major centres (urban and agricultural) was introduced as a historical variable.

For the current paper the research area was extended as to include the region of Malia-Lasithi as defined by Knappett (1999). This region presents a certain degree of cultural uniformity when compared to the neighbouring regions (fig. 1). An independent validation dataset was reserved in order to assess the model’s predictive performance. The fitted model was used to create a probability map for the research area indicating the settlement location probability for the Protopalatial period. The quantitative assessment of model uncertainty based on confidence intervals, which is part of the approach that we present here, is usually lacking in settlement probability mapping.

The settlement probability mapping and model uncertainty assessment provided in the current paper should provide an opportunity to test the statistical model and thus provide a more profound insight into the settlement location choices for the Malia-Lasithi region in Crete during the Protopalatial period. Furthermore, archaeological validation is also needed to establish that the introduction of historical parameters, namely the distance to major centres, is valid or that such parameters are actually the result of an existing surveying bias.
HISTORICAL BACKGROUND

The Protopalatial Minoan period in Crete (Middle Minoan IB–IIB, c. 1900-1700/1650 BC) followed the long Prepalatial period (Early Minoan I-Middle Minoan IA, c. 3500-1900 BC). The Protopalatial period is marked by a gradual social differentiation with the gathering of population in villages culminating in the rise of the first major Minoan urban centres. After a series of earthquakes before 1700 BC, these urban centres were destroyed and rebuilt during the beginning of the Neopalatial period (Middle Minoan IIIB-Late Minoan IB, c. 1650-1480/1425 BC, after Warren & Hankey 1989, 1750-1500 BC, after Rehak & Younger 2001). Further information on Minoan chronology and terminology can be found in Watrous (2001) or Shelmerdine (2008).

As a whole the Protopalatial period represents a period of increased prosperity and tremendous population expansion (Hayden et al. 2004; Watrous 2001). Settlement hierarchy is suggested with information retrieved from architectural elements, high status artefacts and documentary information (Driessen 2001). Settlement hierarchy is connected to the formation of more or less centralised states, linking a city to its hinterland and perhaps extending its influence on a more regional level (Haggis 1999).

STUDY, SAMPLE AND VALIDATION AREAS

An area in central eastern Crete (fig. 1), as assessed by Knappett (1999), was selected for the current study, guided by the premise that material culture reflects a certain degree of unity when considering political, economic and ideological aspects. Knappett (1999) established the area borders, based mainly on studies done on the regional distribution of artefact styles (most notably pottery, and fine tableware). Within the area assessed by Knappett the most problematic area is the Lasithi plateau, with some authors asserting

Figure 1. Study, sample and validation areas.
its independent status (Nowicki 1991, 1995) and others referring to the agricultural importance of Lasithi for the Malia state (van Effenterre 1980).

The study area has two interior areas, a sample area (also referred to as sampling area) used to train the statistical model, and a validation area used to verify the model’s predictive performance (fig. 1).

GEOLOGICAL, GEOMORPHOLOGICAL AND CLIMATIC ATTRIBUTES OF THE STUDY AREA

At the core of the study area is the Dikti mountain range, one of the three large Cretan mountain complexes. Crystalline limestones make up the mountain range with phyllites-quartzites dominating to its north-east and north-west. The Dikti mountain range is extensively karstified, with common occurrence of caves, basins and several plateaus. The plateau basins consist of a red-brown Quaternary alluvium (Fassoulas 2000; Watrous 2001). The Lasithi plain, to the north of the Dikti massif, is the largest high altitude plateau on the island. The limestone mountain range rests upon a stratum of schist, which protrudes from the lower slopes around the edge of the plain. In the areas where the schist is exposed the water seeps out in the form of springs. The coastline is made up of alternating sandy beaches and steep rocky cliffs. The south-eastern part of the study area, roughly from the city of Agios Nikolaos to the Ierapetra basin, is formed mainly by a thick sequence of Neogene marine sediments, with great extents of Pleistocene deposits (Gaki-Papanastassiou et al. 2008). The extension directly south of the Dikti Mountain is formed by alternating areas of flysch, phyllite-quartzites, Neocene sediments and coastal alluvium. Crystalline limestones predominate in the north-eastern corner of the study area. The coastal area to the north of the Diktian massif is characterised by a gentle slope moving upwards to the south of the city of Malia. A phyllitic-quartzitic series, in the mountain slopes south of Malia, acts as a hydrological impermeable substrate of the region (Lambrakis 1998), where Upper Triassic-Upper Eocene karstified limestone and dolomite constitutes groundwater-bearing carbonate sequences that rest on impermeable layers (Lambrakis 1998).

Water is made plentiful available by the existence of a great number of springs and wells, collecting water from aquifers in the porous limestone.

Climatic conditions vary significantly from coastal to mountainous areas. The climate for coastal areas is typically mild Mediterranean, dry and warm. At higher altitudes, as for instance the Lasithi plain at 800m altitude, the climate is clearly of a non-Mediterranean type, with an average winter temperatures from 5°C to 10°C (Watrous 1982).

Data
The data used in this paper forms part of the Digital Crete Project database (Digital Crete 2008). For the current paper Protopalatial settlement locations were incorporated into a Geographical Information System (for the current study the software package ArcGIS® 9.2 was used). Sampling and validation areas were selected due to the higher quantity and quality of settlement data available, with respectively 132 and 118 selected sites.

Different GPS sensors were used to determine the site location with an accuracy of up to three metres. An ASTER-derived Digital Elevation Model (DEM) with a 30 metres resolution was employed. Supportive evidence for the sufficiency of the DEM resolution is provided by the Vrokastro survey (Hayden
which included detailed information on site dimensions for the Protopalatial period and for which the calculated average for site diameter was of 70 metres.

Further details on the methods of data acquisition, satellite imagery, alongside other information on vector and raster datasets can be found in Sarris (2005) and Digital Crete (2008).

**Variables**

Seven independent variables (hereafter also referred to as predictors) were considered (table 1). The independent variables were selected from existing available digital data (Digital Crete 2008) as possible indicators of economic, environmental/climatic and cultural factors that would determine the site location. Including these variables resulted in a model with good predictive performance and we expect that further improvements can be realized by integrating more digital data that will become available in the:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude ($X_1$)</td>
<td>Continuous</td>
<td></td>
</tr>
<tr>
<td>Slope ($X_2$)</td>
<td>Continuous</td>
<td></td>
</tr>
<tr>
<td>Spring Density ($X_3$)</td>
<td>Continuous</td>
<td></td>
</tr>
<tr>
<td>Cost distance to coastline</td>
<td>Continuous</td>
<td></td>
</tr>
<tr>
<td>Cost distance to major centres ($X_4$)</td>
<td>Continuous</td>
<td></td>
</tr>
<tr>
<td>Surface Geology ($X_5$)</td>
<td>Categorical</td>
<td>alluvium (1), tertiary deposits (2), limestone colluvium (3), hard limestones (4), mixed flysch (5), argillaceous flysch (6), schists (7), peridotites (8), granite (9), deposition cones (10), dolines (11), not identified (12), river beds (13), schist colluvium (14), gneiss (15)</td>
</tr>
<tr>
<td>Soil depth ($X_6$)</td>
<td>Categorical</td>
<td>deep (1), shallow and deep (2), bare and shallow (3), deep and shallow (4), shallow and bare (5), shallow (6), deep and bare (7), bare and deep (8), bare (9), not identified (10)</td>
</tr>
</tbody>
</table>

**Altitude**: Climatic differences in the island of Crete are closely connected with altitude. Climate during the Middle Minoan period in the eastern Mediterranean was more humid and colder than at present (Moody & Rackham 1997; Moody 2000; Issar 1995; Issar & Makover-Levin 1996), most likely implying a decrease in the probability of finding high altitude settlements.
Slope: Minoan settlements are often located in a sloped area neighbouring an agriculturally valuable plateau, e.g. Lasithi plateau. This type of placement, when only considering data from a single cell location, would lead to wrong conclusions, which are overcome by considering cost surfaces or de facto site locations (see section Least cost path analysis).

Spring density: With rare examples of perennial rivers, modern Crete water supply relies heavily on pumped water from its several aquifers (Lambrakis 1998). With a colder and more humid climate during the Middle Minoan, the question arises whether past wells and springs represented also the main water source. In many cases there is a close proximity between a well or spring location and the site location. Hayden et al. (2004) reported for the Vrokastro area during the Protopalatial, that there was a close relationship between the site location and proximity to a well or spring. Archaeological evidence shows that water for the Malia cisterns was collected from springs and small mountain streams using a series of canals (Viollet 2003). Knossos was supplied both by a series of wells and by its aqueduct that initially connected to a spring (Angelakis et al. 2007). Seismic activity can affect aquifer conditions greatly and this in turn leads to variations in well levels, and on spring existence (Gorokhovich 2005), which might explain the lower percentage of springs neighbouring the settlements in the Malia region. The fluctuations in spring locations make it inadvisable to use distance to spring as a parameter. Therefore two assumptions were made. Firstly that there is a close relationship between spring and well locations, and secondly, that spring locations in current days still bear some relationship to the location of springs during the Protopalatial period. In order to parameterise the importance of spring locations, spring density was calculated for each map cell using Kernel density estimation (Silverman 1986) with a search radius of 1250 metres.

Cost distance to the coastline: The major urban Minoan centres are typically in the vicinity of the coastline, providing easy access to sea resources and maritime trade (Agouridis 1997; Cline 1994). Crete’s coastline acquired approximately its present shape during the Lower Pleistocene (Moody & Rackham 1997; Lambeck 1996). Since the Bronze Age till present eustatic sea level changes, tectonic movements, and depositional/erosional processes have further induced changes in the coastline (Rapp & Kraft 1978). An isostatic model indicates that the sea level rose in Crete in the past 2000 years about 1.5 metres (Pirazzoli 2004). Seismo-tectonic movements have produced larger changes, with events between 4000 BP and 15000 years BP, consisting of both sudden small sea-level rises (around 25cm), and of sudden emergences like those registered in Western Crete (between 2.7m and 9m) (Moody & Rackham 1997). Further complexity is introduced when assessing effects of sedimentation/erosion at a local level. It should be expected that the Cretan coastline, since the Bronze Age, might have undergone changes up to several hundred metres, with variations along its extent. A probability map should be analysed at a regional scale that mitigates considerations into variations in ancient landscape. A cost distance analysis (see next section Least cost path analysis) was performed in order to define the least cumulative cost from each cell in the study area to the coastline.

Cost distance to major centres: Within a political/cultural unity it is assumed that communication routes would link settlements according to their hierarchical order. This is especially relevant within the Minoan context where Palace complexes probably played an important role in a form of centralised economy (Driessen 2001). Four locations were selected due to their expected historical importance and hierarchical level both within and in the vicinity of the study area, Malia, Lasithi plain, Myrtos Pyrgos, and Vrokastro/Gournia. A previous study (Fernandes 2009) considered the distance to least cost paths linking major centres as a parameter. However, given the limitations of the D8 spreading algorithm used in
ArcGIS evident especially at larger scales, an alternative was procured by considering the least cumulative cost distance from each cell in the study area to the previously defined major centre locations (see section Least cost path analysis). This approach is less sensitive to the determination of a specific route while providing an improved estimate over simple Euclidean distances of the costs of transposing the areas under consideration.

Surface geology: Surface geology serves as an indicator for available natural resources, namely agricultural potential (e.g. rich alluvial areas). With the overall stability of Cretan landscape during the Holocene (Moody & Rackham 1997; Moody 2000), and apart from coastal localities, it is not expected that since the Protopalatial important geological modifications have been operated at the regional scale that this study reflects.

Soil depth: Soil depth is an important indication of agricultural potential (Bulte & Soest 1999). This variable is however far more sensitive than surface geology to changes operated in the last four thousand years (Dietrich et al. 1995).

LEAST COST PATH ANALYSIS

A simple Euclidean distance does not reflect the real difficulties of traversing the area under study. Therefore we present here a more realistic approach that uses a least cost path analysis. This type of approach has already been done in the context of Minoan Crete (Siart et al. 2007; Soetens et al. 2003; Soetens et al. 2008).

The cost of traversing each map cell in the study area was defined by using as a cost function the value of the inverse horizontal speed (1/v) determined from the terrain’s slope (m), with a, b, and c as parameters:

\[
\frac{1}{v} = a + bm + cm^2 (1)
\]

\[
a = 0.75 \text{ms}^{-1} \\
b = 0.09 \text{ms}^{-1} \\
c = 14.60 \text{ms}^{-1}
\]

Equation (1) was derived by Rees (2004) and it was established empirically by traversing mountainous foot paths. This type of cost function is adequate for the Cretan mountainous terrain where slope plays a major role and rivers do not constitute significant obstacles. It was also assumed that surface vegetation and surface geology play comparatively a lesser role in the determination of the cost function.

A Geographical Information System (GIS) implements algorithms and provides tools to determine the least cumulative cost of traversing all cells from a chosen cell location to a destination. For the current study least cost calculations were done using the Cost Distance tool in the Spatial Analyst extension of ESRI’s ArcGIS® 9.2 GIS software package.

Least cost path analysis was employed both as a distance measure and also to define cost surface areas. Site locations were defined as de facto site locations consisting of catchment areas or cost surfaces buffering each site location considering the equivalent in cost distance units of a Euclidean planar...
distance of 1250 metres. The distance of 1250 metres was selected after an informal assessment of the
model’s performance based on two main criteria: an accessible walking distance and limiting catchment
overlap to nearest neighbours. Under this approach we are interested in determining settlement choices
considering the site’s immediate hinterland. Hence the model’s results should be interpreted at an appro-
priate scale.

LOGISTIC REGRESSION

To analyse the data and build a model that can be used for prediction, we use a multivariate Logistic Re-
gression (LR) model. LR is the classical tool for APM, representing Kvamme’s (1983a, 1983b, 1988) integrat-
ed approach. LR is part of a statistical modelling framework known as Generalized Linear Models (GLMs).
GLMs (McCullagh & Nelder 1989; Dobson 1990) are natural extensions to ordinary least squares regres-
sion in which the response variable Y is related to predictor variables through some, possibly nonlinear,
link function. GLMs are among the most widely used models in applications of statistics and analysis rou-
tines are well established. LR is a generally applicable method that can be used for predictive modelling
and classification. It can naturally handle both categorical and continuous predictor variables. Since we
do not want our results to depend in any way on a particular, subjective chosen prior, which may be the
case in Bayesian inference, we use classical frequentist methods for model fitting and inference (Venables
& Ripley 2002). In this paper we are interested in modelling the probability that a given cell on the map
in the study area is a de facto site, given a number or relevant archaeological and geological covariates,
e.g. spring density or cost distance to coastline. Let Y denote this probability and let X be a p dimensional
vector of relevant covariates. The logistic model assumes a linear model for the logarithm of the odds of
site presence, i.e.

\[ \log \left( \frac{P(Y=1|X=x)}{P(Y=0|X=x)} \right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \] (2)

For a given model, where the relevant predictor variables are known, statistical inference such as point esti-
mates for relevant parameters and confidence intervals can be obtained using standard methods. Most rele-
ant to our application are point estimates and confidence intervals for the probability of de facto site
presence, rather than estimates for the individual \( \beta \)’s. However, since the relationship between site pre-
essence and the seven candidate predictor variables we consider is unknown, we need to implement a model
selection procedure to first select an appropriate model on which we can base our inference. Apart from
having a good predictive performance, which is a requirement in order to be practically useful, we need
our model to be archeologically plausible. This means that, if predictive performance can be only mar-
ginally improved with a model containing many terms including higher order interactions, we prefer a
smaller model without complex interactions for reasons of interpretability. In the final stage of model se-
lection, we keep this pragmatic two-fold aim in mind.
MODEL SELECTION

Given the observed binary response \( Y \), indicating site presence, and a covariate matrix \( X = [X_1...X_7] \) containing five continuous and two categorical predictor variables, our goal is to find a model for \( Y \) with a good predictive performance. Here, we use the Deletion/Substitution/Addition (DSA) (Sinisi & Van der Laan 2004) algorithm for model selection.

The use of DSA offers an important methodological improvement over traditional stepwise model selection procedures. Stepwise procedures for variable selection in linear regression are known to be suboptimal and should be avoided. The DSA procedure systematically and progressively builds more complex models that contain more variables and interactions between them in an attempt to find an optimal model that fits the data well and has good predictive performance.

The DSA algorithm performs data-adaptive estimation through estimator selection based on cross-validation and the \( L_2 \) (‘squared error’) loss function. DSA generates predictors as linear combinations of tensor product polynomial basis functions. We believe polynomial regression is appropriate since it is not clear that predictors such as cost distance to coastline and spring density are linearly related to the \((\text{logit}\ of)\) the probabilities of site presence/absence. Allowing a predictor which is a linear combination of products of terms of different powers of the original predictor variables enables us to approximate the true relationship between predictors and outcome. The DSA algorithm generates a sequence of candidate models by minimising the empirical loss function over subspaces indexed by three user-defined parameters. These are:

1. \( \gamma_1 \): The maximum number of terms in the model.
2. \( \gamma_2 \): The maximum order of interactions allowed in the model.
3. \( \gamma_3 \): The maximum allowed sum of the powers of variables involved in a single term.

Starting from a model with only an intercept, the DSA algorithm searches the model space by making different moves. At each step, it chooses between an addition, deletion or substitution move, hence its name. After the optimal model in each subspace is identified, it selects the final model based on \( V \)-fold cross-validation.

The sampling area was randomly sampled using 10,000 points and the model was trained and statistical analysis performed using the programming language R (R Development Core Team 2007) and the ‘DSA’ package (Sinisi & Van der Laan 2004). The model was tested for its validity by considering 5,000 random points from the validation area.

RESULTS AND MODEL EVALUATION

The DSA algorithm was applied to the data from the study area, with default 5-fold cross-validation. We experimented with different settings of the \( \gamma \)'s. In the final run, we chose to set \( \gamma_2 = 2 \). Although a higher setting did result in (marginally) lower cross-validated risks, we find third order interactions unsatisfactory since the practical interpretation of the selected higher order terms is unclear. From an archaeological point of view, we prefer to keep the model simple and since adding more complexity by introducing higher order interactions does not significantly improve the predictive performance of the model, we
decided to include only pair-wise interactions between the predictor variables. Because it is unsure that, especially for the two continuous cost distance variables, the original variables are linearly related to the response, polynomial terms were included up to order three in order to be able to approximate the true relationship between predictors and response. Thus, the required parameters were finally set as follows: \( \{ \gamma_1, \gamma_2, \gamma_3 \} = \{8, 2, 3\} \). Note that the DSA algorithm generates candidate models that include candidate predictors which are functions of the original input variables.

The final model returned by the DSA algorithm for the sampling area was:

\[
\log \left( \frac{P(Y=1|X)}{P(Y=0|X)} \right) = 6.23 - 0.00335X_5 + 3.00 \times 10^{-7}X_5^2 + 0.946X_3 - 8.82 \times 10^{-5}X_4X_6 + 0.238X_2X_6 - 8.95 \times 10^{-12}X_5^3 + 4.16 \times 10^{-9}X_4^2X_7 \quad (3)
\]

The LR model fitted to the data from the study area can be viewed as a binary classifier. Given values for the predictor variables in the model at any location, the LR model outputs the probability that an archaeological site is present at that location. Given a specific threshold, these probabilities can be converted into site predictions. Given the truth, i.e. information regarding whether a site is actually truly present at each location, predictions made by the model can be classified as either TRUE or FALSE respectively depending on whether the predictions do or do not resemble the truth. Hence, positive predictions (which are predictions of site presence at a location) are classified as either True Positive (TP) or False Positive (FP) predictions. Analogously, negative predictions (which are predictions of site absence at a location) are classified as either True Negative (TN) or False Negative (FN) predictions. The performance of binary classifiers is often evaluated using Receiving Operator Characteristic (ROC) curves (Peterson et al. 1954), also used within an archaeological context (Finke et al. 2008). In these curves, TP rate (TPr) is plotted versus FP rate (FPr),

![Figure 2. Model evaluation by ROCs for the sampling area (left) and the validation area (right).](image)
for all thresholds in the range \((0, 1)\). The area under this curve (AUC) is a widely used performance measure. The AUC is a value between 0 and 1. An AUC of 1 corresponds to a ‘perfect’ predictor, whereas 0.5 is the expected AUC value of a ‘random’ predictor. Hence the predictive performance of our model was evaluated by plotting ROC curves and computing the AUC values (fig. 2).

**PROBABILITY MAP**

Through the establishment of a predictive model it becomes possible to create a map providing probability estimates for the location of archaeological sites. In this respect the approach here presented is no different from previously mentioned studies. However often lacking, within current research, is a probability error assessment which constitutes an important methodological deficiency. An analogy can be done, for instance, with the reporting of radiocarbon dates without indicating the type of distribution used or the associated standard deviation. Information on confidence intervals serves as a measure of uncertainty which can be very useful in the implementation of a survey strategy, namely by optimising the selection of preferential survey targets.

Given values for the five relevant geographical variables at any point of a map, the model can be used to estimate the probability of a site presence. Let, for any point \(k\) on the map, \(x_k = (x_{1k}, \ldots, x_{7k})\) be the vector of local geographical data. Let \(Y_k = 1\) denote the event that there is a site at point \(k\). We use the model to predict

\[
\hat{\eta}_k(x_k) = \log \left( \frac{P(Y_k = 1 | X_k = k)}{P(Y_k = 0 | X_k = k)} \right) = \beta_0 + \beta_1 x_{1k} + \ldots + \beta_7 x_{7k} \quad (4)
\]

Let \(Z \sim \mathcal{N}(0, 1)\) and \(z_{\alpha/2}\), be the quantile of the standard normal distribution defined by \(P(Z > z_{\alpha/2}) = 1 - \alpha/2\). Then, we use classical linear model theory which asserts that, under commonly made assumptions, that \([\hat{\eta}_k(x_k) - s_k z_{\alpha/2}, \hat{\eta}_k(x_k) + s_k z_{\alpha/2}]\) is a \((1 - \alpha)\%\) confidence interval for \(\hat{\eta}_k(x_k)\). Here \(s_k\) is the estimated standard error of \(\hat{\eta}_k(x_k)\), which can be calculated from the obtained fit. We use the confidence interval, with significance level \(\alpha = 0.05\), to quantify the uncertainty in the estimated probability at point \(k\) in the map (fig. 3). Figure 3 constitutes an illustration of the final results as these should be interpreted within a GIS environment,
where for the entire region the probability and associated confidence interval values can easily be obtained. The upper and lower intervals are not symmetric towards the probability value (readily visible in some map areas), thus presenting both the upper and lower limits constitutes the most informative form of representation.

**DISCUSSION AND CONCLUSION**

The assessment of the model’s performance using ROC curves (fig. 2) resulted in an AUC value of 0.97 for the sampling area which is indicative of an excellent fit (Hosmer & Lesmow 2004) and an AUC value of 0.67 for the validation area which represents a satisfactory result. This result indicates that the model is indeed able to capture some of the ‘true’ underlying structure that determines the preference for a site location. The lower performance of the model in the validation area is observed especially in the lower left corner of the graph of the ROC-curve (fig. 2, right), i.e. there are some locations with high probability values that are actually false positives. This may be due to differences in distributions of values for some of the predictor variables, i.e. overall differences in values for cost distance to major centres or surface geology between the sample and validation areas due to geological differences. Note that we did not re-estimate the coefficients of the trained model to fit the data from the validation area, which explains at least in part the reduction in the predictive performance of the model. We also point out that from a strictly archaeological point of view, the occurrence of some false positive predictions can be tolerated as long as the overall predictive performance is good and the model gives valuable insights. Under a theoretical perspective it can also be argued that a fully deterministic model is not be expected as several dynamic and cultural aspects are absent from model consideration, a critique often addressed within the post-processual movement (van Leusen et al. 2005).

In relation to the distribution of known archaeological sites the probability map corresponds well with settlement locations that were identified in the field, i.e. the Malia survey (Müller 1996, 1998, 2000), the Vrokastro Survey (Hayden et al. 2004), the surveyed areas of the Lasithi high plain (Watrous 1982) and the extensively surveyed surrounding mountains by Nowicki (1998). The probability map indicates to the north-west of the Lasithi plains areas of high probability for settlement location. As such the region west of Lasithi, the Pediada is known for a dense site distribution (Panagiotakis 2006), as well as the area towards Kavousi (Haggis 2002). High probability areas extended well beyond the natural limits of the Lasithi plain. This is particularly significant for the east and north-east of the Lasithi plain as Nowicki’s surveys reported several archaeological sites along the valley of Amygdali to Mesa Potami (Nowicki 2000), which is the eastern natural passage that goes from Agios Nikolaos to the Lasithi high plain.

Protopalatial site locations of high probability outside sample and validation areas do also confirm a consistency with the archaeological reality, as there is an emphasis on the south coastal plain of Ierapetra, reaching as western edge of potential, the valley of Myrtos Pyrgos, from where higher potential is also predicted going north along the villages of Mythi and Males (Nowicki 1998) and even reaching Selakano. High archaeological potential is also defined further south in the area of the villages Meseleri, Prina, Kalamafka and Anatoli where archaeological sites have been reported by Watrous (1996) and Nowicki (1998, 2000).

The clear concentration of high probability areas surrounding the defined major centres is not sur-
prising given the high importance of the variable cost distance to major centres in the model. The main issue to consider is to what extent the results reflect a surveying bias that may be caused by training the model in the sample area that includes the major centres of Malia and the Lasithi plain. It should be noticed however that the model results for the major centres outside the sample area (Myrtos Pyrgos & Vrokastro/Gournia) are, at least partially, in accordance with existing surveying data (Nowicki 1998; Hayden et al. 2004).

It is hoped that the results presented here will serve as a guideline for further field research, both to consider less surveyed areas but also to perhaps reassess previous survey results. In the case of Crete Bonnefant (1972) reported that only 12.5% of the island had been surveyed with only about 5% of the results published. These numbers have been improved in recent decades but unpublished results remain a major problem. The nature of the surveys also varies in terms of intensity, and there are many associated problems. The finds’ visibility is often linked to the nature of the soil and whether ploughing has or has not disturbed the area under survey. Surveying is often made difficult by climate conditions or by terrains of hazardous access. It is often difficult to determine the site’s extent, site’s chronology, and to verify extent changes within a site’s chronology. Even in the cases of well studied sites there is sometimes reluctance in destroying later layers in the pursuit of earlier ones.

Of special interest should be the areas outside the immediate proximity of major centres for which there is an indication of both high and low probability targets. In this respect the evaluation of the error maps is especially relevant as it provides a better insight into model uncertainty and thus allows for an optimisation of a survey strategy.

This work should be viewed as having a continuously ongoing nature, whereby the model here presented can be reassessed and improved by an evaluation done in the field and by the update of existing data. Particularly relevant is information obtained from paleo-environmental/climatic reconstructions as this would allow assessing settlement location choices as based on present-day data. A specific example of this type of reconstruction is the integration of data concerning Minoan spring locations which due to past seismic activity have subsequently often become inactive.

ACKNOWLEDGMENTS

The authors would like to thank the Institute for Mediterranean Studies – FORTH for having kindly provided most of the raw data used in this study. We would also like to thank the editors and reviewers of the LAC2010 proceedings for their kind comments which have greatly benefitted the current paper.

REFERENCES


