PREDICTIVE MODELING OF SETTLEMENT MOUNDS (9000-5500 B.C.) IN THE LAKE DISTRICT REGION AND ITS IMMEDIATE ENVIRONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

TUNA KALAYCI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
SETTLEMENT ARCHAEOLOGY

SEPTEMBER 2006

Approval of the Graduate School of (Name of the Graduate School)

Prof. Dr. Sencer Ayata
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Numan Tuna
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Vedat Toprak
Co-Supervisor

Assist. Prof. Dr. Geoffrey Summers
Supervisor

Examining Committee Members

Prof. Dr. Ali Uzun (METU,STAT)

Assist. Prof. Dr. Geoffrey Summers (METU, SA)

Prof. Dr. Vedat Toprak (METU, GEOE)

Assist. Prof. Dr. Evangelia Ioannidou-Pişkin (METU, SA)

Assist. Prof. Dr. Mehmet Lütfi Süzen (METU, GEOE)

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name :Tuna, KALAYCI

Signature :

ABSTRACT

PREDICTIVE MODELING OF SETTLEMENT MOUNDS (9000-5500 BC) IN THE LAKE DISTRICT REGION AND ITS IMMEDIATE ENVIRONS

Kalaycı, Tuna

M. Sc., Department of Settlement Archaeology

Supervisor : Assist. Prof. Dr. Geoffrey D. Summers

Co-Supervisor: Prof. Dr. Vedat Toprak

September 2006, 185 pages

This study aims to construct a predictive model that investigates patterning of settlement mounds by employing environmental variables. The results then will help to search for unknown sites of the same age. The methodology is applied to the Lake District of Anatolia for the period of 9000B.C. – 5500B.C.

Four main sets of data are used in this study. The first set is the settlement data, which includes the names, coordinates, and periods of the sites. The sources of independent datasets are topography, lithology and soil. The study starts with the straightforward procedure of plotting the sites in the region. Then the layers (independent variables), populated with their sub-fields, are included in the model in the GIS to construct a predictive model by using logistic regression.

Results reveal some high potential areas with no known occupation, as well as some zones which need more research. Also, hierarchy of environmental variables is detected, which affected the settlement patterning of the study area.

Keywords: Lake District, Predictive Model, GIS, Logistic Regression

ÖZ

GÖLLER BÖLGESİ YÖRESİ VE ETRAFINDAKİ HÖYÜK YERLEŞİMLERİNİN (M.Ö. 9000-5500) CBS ÜZERİNDE TAHMİN MODELLEMESİ

Kalaycı, Tuna
Yüksek Lisans: Yerleşim Arkeolojisi
Supervisor : Yrd. Doc. Dr. Geoffrey D. Summers
Co-Supervisor: Prof. Dr. Vedat Toprak

Eylül 2006, 185 sayfa

Bu çalışma, çevresel değişkenleri kullanan tahmin modellemesi ile höyük yerleşimlerinin yayılımlarını incelemektedir. Çalışma, Anadolu'daki Göller Bölgesi'ne uygulanmış olup M.Ö. 9000- 5500 dönemini kapsamaktadır. Sonuçların tespit edilememiş yerleşimlerin bulunmasına yardım edeceği umulmaktadır.

Çalışmada dört ana veri kümesi bulunmaktadır. Birinci küme, yerleşimlerin isimlerini, koordinatlarını ve dönemlerini içermektedir. Bağımsız değişkenler topoğrafya, kayaç ve toprak verisinden üretilmiştir. Yerleşimlerin bölgedeki dağılımlarının incelenmesinin ardından tabakalar (bağımsız değişkenler) CBS ortamında çeşitlendirilmiş ve lojistik regresyon ile tahmin modeli kurulmuştur.

Sonuçlar, üzerinde bilinen yerleşimlerin olmadığı bölgeleri ve daha fazla araştırma yapılması gereken yerleri ortaya koymuştur. Aynı zamanda, çalışma alanındaki yerleşimlerin yayılımına etki eden çevresel faktörlerin önem dereceleri belirlenmiştir.

Anahtar Kelimeler: Göller Bölgesi, Tahmin Modeli, CBS, Lojistik Regresyon

geçen zamana…

ACKNOWLEDGEMENTS

Very special thanks to:

Advisors, Geoffrey Summers and Vedat Toprak

Committee Members, Ali Uzun, Evangelia Ioannnidou-Pişkin, and Mehmet Lütfi Süzen

All of my friends that I have

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Advancement of information technologies has brought new dimensions and possibilities to many scientific and pseudoscientific applications. Increased processing power and storage capacities altered non-feasible activities to ordinary ones. Archaeological applications, and even archaeological thoughts, are affected by this progress. Today, people ask different archaeological questions and employ different techniques to solve them. On the other hand, it is clear that some questions have never changed, and they are waiting to be solved with classical tools of archaeology.

Predictive modeling was a concern of archaeology even before the extensive use of computers, but with different names, or even without any name. Now, it is turning out to be a standard procedure for any big scale archaeological project, yet it can be stated that modeling efforts are still in their infancy. If employed properly however, broad information and even useful knowledge can be obtained.

Any model is constructed with many assumptions and restrictions. For this reason, they should be supplied with a clear list of problems and possible alternatives of solutions. Those restrictions are not general, but case specific and they do depend on temporal and spatial domains of any study area. Thus, there is no single predictive model covering all archaeological questions.

1.1 Purpose and Scope

This study aims to construct a predictive model over a wide area by employing environmental variables obtained from various sources. Apart from its modeling aspiration, single environmental variable to site location relations are also sought after. These relations are specified via constructing quantified data layers.

Definitely, the main objective is to locate possible site locations in the study area. To accomplish this, formal statistical procedures are the best candidates. Then again, there are many statistical tools which can be used to construct the model, and obtain probabilities of possible site locations.

1.2 Study Area

Study area is located in so called the Lake District and western fringes of Konya Plain. It is comprised of eight drainage basins extracted through GIS analysis, which are also later on manually corrected. Each basin has a particular lake at its base, but there are more than eight lakes in the area which are used for modeling. The intention behind choosing such a characteristic area to apply a predictive model is to make use of potentials of area, such as palynological studies, environmental reconstructions and various others (Figure 3.1). Although they are not planned to be fully employed, the potentials are well defined and some conclusions are expected to be drawn.

One of the discussions in building a predictive model is the determination of boundaries of a study area. Knowing this, a considerable effort is given obtaining a well defined and proper boundary. To do this, both geographical and archaeological data are used. While the extent of the area is determined by archaeological knowledge, the exact boundaries of those extents are drawn on the basis of geological information.

Temporal domain of the study area occupies a considerable part throughout the study. Efficiency of any predictive model is also based on temporal coverage of the model. A wide timespan will produce a very general picture and a set of rules which are not applicable for the entire time period.
A very narrow timespan, on the other hand, will be too restrictive and the use of data for that particular period of time will be questionable. In this study, a time range of 9000-5500 B.C. is used to perform the analysis.

This period is claimed to be the last period of human history which is greatly influenced by environment, and first period in which material culture is visible through substantial elements.

1.3 Methodology

To construct the predictive model, logistic regression is used. It is selected as a tool since it requires fewer assumptions than other statistical techniques, and it is widely employed for predictive modeling applications.

A dichotomous dependent variable, *Site Presence,* is the main focus, where probabilities are found in between 0 (No Site) and 1 (Site). As covariates, environmental variables are used. Those variables can be divided into four categories.

First category of variables is topography. Basic source of this set is Shuttle Radar Topography Mission (SRTM) imagery. This satellite data is imported and combined in GIS platform to obtain primary coverage, and secondary coverage is obtained through GIS analysis. Primary coverage was also used to draw the exact boundaries of study area.

Second category of variables is soil. Data used to construct the model was already digitized and provided with an attribute table. Each attribute is decided to be a potential variable for the model. On the other hand, querying data screened some of those attributes, and they are withheld from the analysis.

As a third category, rock data is decided to be used. It was also available in digital format, and directly imported to GIS.
Data has been reclassified both for the ease of visual improvement and statistical excellence, but this classification is geological rather than archaeological, so that criticism can be raised.

Fourth category is comprised of hydrological data. This set of data is obtained from Elevation data through GIS analysis. As a very dynamic entity, use of this layer requires attention. In knowing this, social use of water is considered rather than the domestic use. Thus only lakes and higher level rivers are used in the analysis.

Variables obtained from those major categories are then used to create a predictive model, which can serve both to academic purposes in terms of understanding the determinants of settlement patterns for the given time period, and management purposes for planning future land development.

1.4 Organization

Total predictive modeling effort is not solely to produce a formula, but to understand the results of it. Thus, there is a need far more than a mathematical integration of datasets. To accomplish this, study is divided into many components, or chapters, for an easy understanding of whole concept. Chapter 2 is devoted to a survey of predictive modeling definitions, tools and theoretical controversies. By this way, the tool to be used in this study is chosen. Chapter 3 deals with archaeology of the area. Although there are some excavated sites in the area, total amount of information is not enough when surveyed material is also considered. Then each piece is important for grasping a general idea on response of human to nature. Chapter 4 is on geography and ecology of the area. Information obtained from this part is used to interpret the environmental variables in an efficient way. Chapter 5 is devoted to a discussion on determining the boundaries of study both in space and time. While doing this, understanding obtained from previous chapters is used. Chapter 6 introduces data which will be used to construct the model. Some graphs and descriptive statistics are also given for a better understanding.

Chapter 7 is the core of this study. Basic understanding of site environmental variables is obtained through the modeling efforts. Chapter 8 is a general survey on settlement patterning of the sites in concern. Chapter 9 provides a general conclusion on predictive modeling of settlement mounds of the Lake District Region and its immediate environs.

CHAPTER 2

PREDICTIVE MODEL: A SHORT REVIEW

2.1 Definition

A predictive model is defined by Kohler (1988; 33) as

> A simplified set of testable hypotheses, based on either behavioral assumptions or empirical correlations, which at a minimum attempts to predict loci of past human activities resulting in a deposition of artifacts or alteration of the landscape.

Then archaeological predictive modeling is a method for estimating the probability of archaeological site occurrence within a specific geography (Moon, 1993; 2). The casual link between site locations and natural independent variables is considered to be multivariate- that is, people positioned their sites with respect to an optimal combination of all resources in which they were tested (Altschul, 1988), or in a simple manner, predictive models are tools for projecting known patterns or relationships into unknown times and places (Warren and Asch, 2000).

Less formally, a predictive model identifies "patterns in spatial relationships between sites and their physical locations and thus indicate potential relationship between the natural and social environment and the locations of past human activities" (Moon, 1993). Then whole concern can be reduced to the basic question of whether or not a location contains any archaeological material (Wheatley and Gillings, 2002; 166). On the other hand, it should be clear that predictive model hypotheses *potential of being* rather than strongly asserting, or pinpointing, a site is placed at a particular place (Ejstrud, 2002; 2).

Although there are various methodological differences that exist in constructing a predictive model, the ultimate aim is to identify the location of human activities which

can be claimed as the responses to environment from various aspects such as time, space, and function (Kincaid, 1988, 550). Although such an approach attracts much criticism, the idea is still a base for constructing a predictive model.
While a model aims to locate sites, it also investigates the reasons for those responses. Apart from its fashionable use, predictive modeling can also be used for predicting the type or quality of remains, and as a database of sites, the current damage status of sites from a resource management perspective (vanLeusen, 2001; 2).

Different cultural groups, behave differently in response to different environments at different time periods (Moon, 1993), or conversely, but more correctly, landscapes are perceived and used in different ways. This perception is reflected and can be observed as a patterning in a region, which should also be measurable (Fry et al., 2004; 98). Thus there is no single predictive model for the whole. In other words, strategies determine the use of environment, which is reflected in archaeological data. This reflection is also observed in the correlation of locations of sites with other variables, or so-called independent variables, from a formal modeling perspective.

Predictive modeling is not new in archaeology, and in fact each survey and regional study is carried out with an implicitly defined set of rules about where a site might be located. What is different from the earlier studies is the advent of computational power and sophisticated tools to conduct such an analysis, such as Geographical Information Systems, or GIS (Wescott, 2000; 2).

A well defined study by Williams et al. (1973; 3) tries to identify presence/absence of sites in a specific study area with given decision rules:

The locus should be on a ridge or a saddle.
The ground should be relatively flat. (<5% slope)
The locus should be in the low foothills. (<250 m above the valley floor)
The locus should be within the modern pinon-juniper ecotone (<1000m)
The locus should be near semi-permanent water source (<1000m)
The locus should be some minimal distance from this source (>100m)

The formal definitions of those rules were the initial steps for the future of formalized GIS predictive modeling studies. In fact, this set of rules can be used as the layers of a GIS.

2.2 Use of the Model

Many sites are being destroyed, especially in developing countries, in the name of land development (Warren and Asch, 2000; 6). Since such *progress* is claimed to be irreversible then it is vital to distinguish between sites and non-sites. On the other hand, there is no consensus on the definition of an archaeological site and the definition changes through time and space. Thus, if a settlement distribution is studied over the landscape, a clear definition has to be set for the definition of settlement. It can be claimed that it is any area which contains human activity, but to be more restrictive it is the place where people live or were living (Sollars, 2005; 253). The difference between the workshops, caves, hilltop settlements, mounds should be explicitly defined. This can be done by setting clear definitions, and/or putting strict assumptions. Moreover, those definitions and/or assumptions might not be valid for a different period of the same region and can not be generalized for different places in same era. In other words, settlements evolve, change, progress or regress, thus a new explanation has to be made.

Overall effort is about modeling a very complex system, thus it is nearly impossible to locate each response of the ancient people to the environmental system. As a set of tools, predictive modeling aims to understand the human locational behavior. (Ejstrud, 2002; 18), but it can not be replaced, at least for the moment, by a proper archaeological survey. On the other hand, it might be helpful to conduct such an analysis over an area which has not studied before, or to show the places that need more research (Ejstrud, 2002; 19).Also, it will decrease the effort given to a survey and, optimistically, will give help in developing an intensive survey strategy (Moon, 1993; 26).

The popularity in developing predictive models is a result of the ‘identification, protection and management of increasingly threatened cultural resources’ (Duncan and Beckman, 2000; 33). Then again, it is clear that modeling should be performed in an effective manner, both in terms of time and resources. With the advent of GIS, manipulation of huge quantities of data as well as the flexibility of the analysis is possible, which makes predictive modeling an efficient tool, if constructed properly (Duncan and Beckman, 2000; 34). Moreover, a predictive model provides determinants of a location of a site (Warren and Asch, 2000; 8, Trigger, 1968), a schema for the protection of the area (Warren and Asch, 2000; 8).

2.3 Variables used in the Model

To construct a model, a number of variables are introduced to the model, and a number of results are obtained according to the questions asked. This is well defined by Wheatley and Gillings (2002; 168) as a distinction between inputs to the model, and outputs from the model. The inputs are:

- Spatial parameters (spatial behavior of observations, clustering, randomness)
- Physical environmental characteristics (for example, slope, aspect, distance to water, surface roughness.)
- Economic features (palaeoeconomic reconstruction)
- Cultural features (road networks, central places)

This categorization is not unique, but any of the variables might fall into one of the categories. In an ideal case, a model should include each of the input components stated above, but since social and cultural aspect of locations are unknown for most of the cases, researchers tend to use biophysical properties of any location and exclude or make proper assumptions on cognitive variables (Moon, 1993; 10), which is in fact the main source of criticism of predictive modeling.

The outputs, based on inputs can vary from simple to complex.

- Presence/absence of a site. Output is binary
- Site class. Output is categorical, where the model decides whether the location is a hunting camp, work shop or a base camp.
- Densities of sites/artifacts. Output is ordinal or real number.
- Site significance. Classification of the landscape according to the perceived importance of the archaeological remains.
- Site probability. A probability is assigned to the pixel on the presence/absence of the site.

2.3.1 Archaeological Data

Archaeological data for predictive modeling studies, or in other words, locations of sites, are manifestations of human locational behavior. The selection of the location of a mound, for instance, represents an initiative that effects future generations of that past, assuming that there was no severe environmental change that made the successive people to abandon the area. In fact, there are some clear examples of such cases of abandonment (Farrand, 1964, Bordaz, 1968; 44, Mellart; 1970; 8). Workshops, on the other hand, were set up for specific purposes such as the manipulation of raw material, butchering functions, and various others. It can be assumed that those locations do not necessarily reflect vital conditions, but aim to minimize the effort to access certain sources. The determinants for the locations of such workshops can then be, for instance, proximity to resources or migrating animal routes. On the other hand, such kind of ‘least cost’ approach is an end-product of industrialization, and it is suspicious that such methodology can be applied to a Neolithic society, where a numerous practices could be in contradiction. Also, according to Kohler and Parker (1986) oversimplifying the phenomenon results in deficit models, where even simple foraging systems can have differentiated site functionalities.

There should be at least three basic concerns in using archaeological data in a predictive model (Ejstrud, 2002; 5):

- Context: The way a site makes use of the landscape. Different land use characteristics should be defined separately and explicitly in a model.
- Dating: The sites should belong to a main cultural period, since cultural characteristics and technologies might result in different land use characteristics. Moreover, a wide timespan will include severe environmental changes, which might in turn affect the locations of the sites. Thus, integration of temporal GIS with predictive modeling might produce better results (Yuan, 1996).

Predictive modeling studies do not consider time in great detail and sometimes the concept is totally missed. Even when it is considered to some extent, minor details are overlooked. Time as a concept, creates norms and such norms can sometimes surpass the current conditions and lead to ‘illogical’ decisions. Moreover, even if some ‘logical’ decisions were given for locating a settlement, those criteria can change form and evolve to other set of rules, and can even be totally be forgotten (Kohler and Parker, 1986).

- Location: The locations of the sites must be exact, because modern predictive modeling applications require accurate coordinates rather than ambiguous phrases like ‘near’. On the other hand, in order to include the old survey data into models elaborate media should be developed.

Locations of sites should be approached critically since they can be approached from two counter aspects. One, landscape can be viewed from an individual site, and two, landscape can be viewed where it is occupied by sites. Different approaches result in different methodologies (Perkins, 2000; 133).

2.3.2 Modeling Data

To summarize, data may include clusters of spatial parameters, physical environmental characteristics, economic features, and cultural features, which can be also stated as biophysical properties of the location, the social subsystem, and the 'cultural' information (Moon, 1993; 2), but it can be claimed that most models only include environmental variables. This is not just because of the ease of finding the environmental determinants of the location, but due to the asperity of obtaining 'social' data (Brandt, Groenewoudt, Kvamme, 1992; 269, vanLeusen, 1993; 107).

Data itself provides information, and thus creates knowledge, but it is the context that changes knowledge to understanding (Tosta, 1991). Social context is a highly speculative issue, and a predictive model using a social context will create a biased end product. Environmental variables are measurable, and this is why they are widely used. On the other hand, they are used with the assumption that the environmental variables of the past are represented, at least indirectly, in the existing data (Warren and Asch, 2000; 6; Dalla Bona, 1997; 17).

In this study, as in most others, social variables are not included to the model. Rather great emphasis is given to environmental variables.

2.3.2.1 Environmental Variables

Environmental (or independent) variables of predictive modeling studies stem from primary coverage. Then there is a need for discussion of the origins of data. On the other hand, as being discussed, it should be kept in mind that very few variables are directly used, but most others are proxies for other real archaeological determinants. For instance, elevation, apart from its direct use, is a proxy for the growing season, the amount of summer rainfall or winter snowfall, and vegetation type.

On the other hand, the use of those proxies without discussion can be misleading because the meaning can be opposite in two different models (Kohler and Parker, 1986; 415). Moreover those embedded meanings of proxy variables can create implicit correlations of variables and distort the model.

2.3.2.1.1 Geomorphology

Geomorphological data is one of the least shifting agents through time used in a predictive model. For this reason, it is widely used (Warren and Asch, 2000; Duncan and Beckman, 2000). On the other hand, this shift considers natural changes rather than human induced changes.
Although this is much more rapid today than in prehistoric times, terracing, leveling or similar practices are known to be applied even in the early phases of prehistory.

Basic elements of concern are slope, aspect, and elevation. On the other hand, geomorphological elements are environmental variables rather than archaeological ones. Although they are measures of terrain, they implicitly define exposure to sun and rain, defensive values and other survival conditions (Church, Brandon, and, Burgett, 2000; 143).

2.3.2.1.2 Soil

Soil data to be used in a predictive model are obtained from soil maps. Most of the time, soil data is not suitable for the analysis and thus they are reclassified and combined in order to create other soil related variables and much more related with human locational behavior, such as permeability, drainage, flood frequency, and landform type (Warren and Asch, 2000; 15). Unfortunately, this is an ideal case as well, and due to methodological constraints this reclassification does not end up being the perfect case, even sometimes they are hardly reliable (Leveau et al., 1999). Quantitative data, on the other hand, can be reclassified without much subjectivity (vanDalen, 1999; 122).

A modern soil map can hardly be representative of past soil distributions. For this reason, past 'soilscapes' should be reconstructed. In order to accomplish this task, archaeological excavations, geomorphology and sedimentology, palynology and biological indicators, and edaphology can be used (Leveau, Trement, Walsh, Barker, 1999). On the other hand, such reconstructions depend on the scale of study, and can hardly be feasible.

As a general observation, it can be stated that sandy soils are easy to cultivate but holds less nutrition and water, whereas clayey soil, on the other hand, is rich in water and nutrients, but hard to cultivate. Such preliminary acceptations are first steps for reconstructions and land use determinations (Bintliff, 1992; 76)

Erosion is one other important concept to be discussed under the title of soil. Erosion is a cause of destruction of settlements. Also, this process buries them under alluvium (Zangger, 1992; 135). Although, erosion can occur naturally, it is the human affect that boosts the course of action. Attica near Athens showed great landscape stability in the Holocene, with minor erosion instances, but with the evidence of increased land use, the level of erosion is immensely increased. It has been estimated that in Greek Argolid of c2500 BC the level of erosion is 40cm in mountains and 100 cm in lowlands. The ancient city of Eretria was abandoned due to catastrophic erosion in the vicinity. Also historic evidences state the immense breakdown in Attic crop production (Bintliff, 1992; 127).

Climatic change or human induced devegetation does not only affect the soil loss but also the agricultural potential. On the other hand, introduction of new technology and/or new species can resist this change or loss.

2.3.2.1.3 Hydrology

There is no doubt that water is the key element for life. It can be stated that it was much more important in the early times of prehistory, since there were limited transportation methods for water, but as technology progressed the impact and manipulation of humans on their environment was brandished by artificially diverting water courses, digging holes to reach the water table, and building aqueducts.

Although it cannot be totally generalized, a case study shows that there is a tendency of movement away from water as human technology progressed. The hunter-gatherers of Eastern Jutland, Denmark, shows a distribution along rivers, lakes, and shorelines, but by the Neolithic direct contact was lost, and by the Late Bronze Age, the contact with open water is totally broken (Ejstrud, 2002; 19). Also according to Sollars (2005; 258), no permanent water supply is found within the periphery of the Cypriot Late Bronze Age sites of Pyla Kokkinokremos or Maa Palaeokastro.
On the other hand, it was not clear if a source has vanished or there was really no water source, but in any case human activities of wells, river diversions, terracing are then for minimizing the environmental stress (Raikes, 1967; 5).

On the other hand, a distinction should be made between the domestic use and other uses of water. The use may change, and it should be explicitly defined in a model, where water is used for drinking as well as for irrigation (Wilkinson, 2003; 45). Small streams might have been used for domestic purposes, whereas larger ones might have been used for other purposes such as transportation, fishing, and other economic and ritual activities (Ejstrud, 2002; 4). While it is true that the variable concerned is the same, the results are entirely different, and should be interpreted in a different way.

Springs and streams should have been the main sources of water. Thus any research based on potentials, and behavior of those sources can shed light on the locations of ancient settlements. Not surprisingly, they are affected by climate and topography so that they are dynamic entities. Thus using them as variables in a model requires extreme attention and critical thinking.

Springs can be expected to occur at the valley sides where the massive horizontal bedrock is suddenly cut.

Also, rocks with joints enable circulation of ground water, but there is little movement of the ground water, thus not many springs are observed, but in a valley forms can be eroded to the level of an aquifer which enables the springs to be revealed. To sum up, one can expect to find a spring along the side of a hill or in the walls of a valley or at a fissure or a crack, due to joins or faults (Lobeck, 1939). On the other hand, ground water is also affected by the permeability of the bedrock.

Streams should also be examined and used wisely in models since they are not static, but dynamic entities where the abundance of the rain affects the effectiveness, duration, strength, etc. of the stream. Even when rain is abundant, if the soil is porous and/or heavy forest is on the soil the river system will not be complex.
VanLeusen (1993; 110) gives an algorithm for simulating the hydrological behavior of the terrain so that reconstruction of past flow can be done.

Lithology is also a determinant of the type, in fact even of the existence of streams. For instance limestone landscape results in springs but not streams. Arid regions can have streams due to minimal vegetation, but the system is also minimal in terms of volume, number, and permanency.

Not only the lithology but also soil type is highly correlated with streams. If in a drainage basin most of the precipitation is added to the stream system as surface run-off, then the effect is erratic, but instantaneous and not long-lived. This is the case if soil is clayey, which does not permit penetration inside, or the drainage basin is full of bare rock with few breaks and/or joints. On the other hand, if the soil is sandy, or covered with loose soil or humus, then the precipitation penetrates to underground, and joins to stream system, resulting in sustainment even in periods of drought (Lobeck, 1939). Thus, the variables to be used in a predictive model are not independent of each other, but highly correlated.

Water availability depends on factors other than direct human affect on water sources, but it should be clear that most of the factors are dependent on hydrological factors or "at least have a predictable affect" (Raikes, 1967; 8). The climate, soil type, landscape and very many factors determine the availability of water. On the other hand, the availability is not the sole factor, but duration and quality is also important (Church, Brandon, and, Burgett, 2000; 143).

The quality of the water is effected by the character of the bedrock, the style of weathering, the hydrogen ion content (pH) the oxidation/reduction conditions (Eh), and locally, magmatic conditions. Also hard water has an excess of bicarbonates, which is due to the existence of limestone, or other carbonate rocks, or of sulphate derived from gypsum (Watson, 1983). The quality of water can also be used for determining the choice of irrigation as well, such as spring-fed irrigation versus rain fed flood, where alkalinity and salinity analysis can be used to detect to choice (Barker et al., 1999; 278).

The ordinary reaction to the end product of a predictive model is seeing the highest potential along the water sources, but in fact the distribution is as varied as the landscape itself. The high possibility of selection along the shorelines and riverbanks might be true, but there are also places near water bodies where it is impossible to settle down (Dalla Bona, 2000; 86).

2.3.2.1.4 Geology and Lithic Sources

Lithology is static like geomorphology, which makes it easier to model. As a raw material, lithic sources can be seen as a determinant in location selection, but this may only be true for some special use sites (Church, Brandon, and, Burgett, 2000; 145).

An example of how geology can be used for archaeological understanding is given by Zangger (1992; 142), where slight erosion of limestone creates ridges, whereas soft marl and flysch are vulnerable.

In the same manner, limestone landscapes are often characterized by clay rich terra-rossa soil; highly fragile but fertile as well. Also marl and flysch dominated landscapes create gentle slopes, suitable for agriculture.

2.3.2.1.5 Vegetation

Although geomorphological variables are commonly used in predictive modeling, it is agreed that vegetation is also important, but it is again a dynamic entity rather than static (Dalla Bona, 1997; 6). The variable is important for two main reasons:

- Different types of vegetation contain different plants in different quantities.
- Borders of vegetation are attractive since it is possible to exploit two different zones, maximizing the quantity of plants (Moon, 1993; 38).

Vegetation is not only a source for food, but it is known that plants can also be used for medicine, or as a dye (Dalla Bona, 1997; 6; cf. Dalla Bona; 2000, 75 – Schermer and Tiffany; 1985, 220). On the other hand it is questionable that if such a parameter can be a major determinant for a selection of a location, depending on the site type.

Vegetation itself is a source for construction material as well as fuel. Thus it can be stated that it is important as a determinant of settlement patterning, like many others. Unfortunately, its use is not so easy since it requires reconstructions of the immediate environment, which is hardly possible in most instances.

2.4 Methodology

There is no clear agreement on constructing predictive models, which has resulted in a wide range of methodology. In fact a rigid theoretical framework of predictive modeling is not yet fully developed. A brief review will quickly reveal that early predictive modeling efforts (and modern ones as well) were based on dichotomies. This has been well described in a study of vanLeusen (2000; 5-4) as:

- Inductive vs. Deductive
- North America vs. Europe
- Cultural Resource Management vs. Academic
- Ecological vs. Cognitive

2.4.1 Inductive and Deductive Models (Theoretical Approaches)

The basic duality in predictive modeling stems from data itself. However, this duality does not originate from data manipulation but, rather, from the understanding of it. This problem, if it is a problem, is general and not specific for predictive modeling.

2.4.1.1 Inductive Models

The Inductive (Empirical, or Correlative) model helps to create some decision rules based on a sample of observations. Such decision rules are based on statistical methods and after extracting those rules, a model is constructed (Warren and Asch, 2000; 8). In other words, the researcher uses an available set of sites (data) to construct the model so as to find the unknown (theory) (White, 2002; 22). In fact, the term or the problem of induction is defined as "generalizing about the properties of a class of objects based on some number of observations of particular instances of that class of objects" (Wikipedia Dictionary, 2006a).

A correlative or inductive model shows where the sites might have been located, but not why they have been located to those particular places. For that reason correlative models are not totally predictive, but they are the projections from known to unknown, producing an explanation of the sample of sites but not the whole universe. This type of model is successful if applied to societies whose decisions are heavily determined by environment, such as hunter-gatherers (Moon, 1993; 15). Since any correlative model is based on an archaeological inventory, or site database, questions raised should be changed or modified as new data is introduced to the model, while in certain circumstances iteration of the model might be necessary (Moon, 1993; 18).

Since an inductively-derived model is generated from a database (Dalla Bona, 1997) it is vulnerable to biases in the database. French (1970; 139) gives a clear list of possible biases for archaeological locations:

1) Samples collected for dating the mounds does not necessarily reflect the age of the mound.
2) Terraces and valleys at the edge of the study area might be neglected as well as hill and mountain tops.
3) Some types of mounds, e.g. low or flat ones, may have been neglected.

Moreover, French asserts a critical assumption that all sherds found on a site indicate a settlement-occupation (1970; 142), which might not be always the case.

Furthermore, inventory elements might not be truly recorded, or even mistyped. The environmental data collected might be too coarse to be used in the model, or even missing. The information obtained can be outdated due to severe changes, but the most important error is the systematic error, and should be checked for at the very beginning of analysis (Dalla Bona, 1997; 5).

Ebert (2000; 133) heavily criticizes inductive models due to their inefficiency, and adds that there is a need to search for those factors that attracted ancient people to particular places. Thus an inductive model should be a search for theory rather than a mere exploration. Moreover, a general level of success is about 60-70% for an inductive predictive model, which is not low but, at the same time not completely satisfactory.

Despite the fact that inductive models suffer from various aspects, they are widely chosen. Processing information and turning it into knowledge does not take so much time and effort (Church, Brandon, and, Burgett, 2000; 136).

2.4.1.2 Deductive Models

Deductively-derived models aim to understand human locational behavior and then try to predict site locations accordingly (Kohler, 1988; 37; White, 2002; 22). This prediction is then tested according to the existing database.

According to Kohler and Parker (1986; 432), deductive models cover three considerations:

1. Inference about human locational behavior. It consists of (i) a process of decision making; and (ii) a result of the process
2. Specification about the determinants of decision making process
3. An opportunity to measure the determinants as variables, and test those according to a hypothesis.

By employing deductive models, some specific questions are asked about human locational behavior (Whitley, 2001). Although they are quite significant, they are rarely answered.

1. How any site selection process happened?

It is not just economic, but also psychological and/or sociological. It is not static but a dynamic process that cannot be reduced to ordinary generalizations (or laws).

2. How were the decisions made?

This is not merely a process, but also elements involved in the process. Neurons (individuals of the society) work on data in both parallel ways and hierarchically, they evaluate and compare. Also they learn from the experiences. Then if there is such an algorithm to find out the mechanism of locating settlements in the landscape, the result of the algorithm will produce one best place with superior conditions to any other place. On the other hand, due to constraints, there will be sacrifices from the perfect case. Moreover, since the environment is a dynamic phenomenon one best place can never have the same conditions forever, and there will be new ‘best’ places to locate as settlements.

3. How were the potentials evaluated?
4. How was evaluation implemented?

There are attractors (or determinants) which cause to locate a settlement. They can be seen as the "representation of the observable characteristics of the rules of the system." They can be classified into permanent point attractors (a lithic source), periodic attractors (migrating game) and strange attractors, where it is the response of the system to the other systems, which is not predictable but can be simulated (Whitley, 2001).

The discussion above can be defined mathematically to put into GIS (Wheatley and Gillings, 2002; 169). If the attractors are the variables of the decision making function then:

$$M=f(x_1, x_2, x_3,\ldots,x_n>0) \quad (2.1)$$

$$M'=f(x_1, x_2, x_3,\ldots, x_n<=0) \quad (2.2)$$

where M is the case representing site and M' is non-site, then f is called the multivariate discriminant function.

There are (at least) two scales which determine the location of the settlement. The regional scale is more general and concerned with settlement strategies. The local scale is directly related with the minute affects of the environment, closeness to water or raw resources (Whitley, 2001).

Although deductive models seem much more elaborate and exact than inductive models, this type of modeling suffers from the theoretical orientation of the researcher. Such a bias is more destructive when compared to an incomplete model.

Moreover, there is no clear cut division between those two theoretical concerns. In fact, existing knowledge of determinants of settlement patterns comes from survey data, and surveys are performed with an implicit knowledge and orientation of the researcher.

Thus they are not totally separate, but feeding each other. The selection of one or the other is not theoretical, but a practical matter.

At this point it is logical to follow a methodology that considers constructing various types of models, and various variations of a single model. The advent of GIS to predictive modeling makes such kind of work applicable.

2.4.2 Approaches

Apart from theoretical considerations, there are also some methodological controversies. They are independent of theoretical dichotomy, and have their own bases. The first division is on approaches to methodology, namely numerical vs. graphical. The other dichotomy is about procedures of methodology, intersection method vs. weighted value method.

2.4.2.1 Numerical Approach

Numerical approach can be stated as an end product of the effect of statistical techniques to archaeology. As a branch of statistics, multivariate analysis is used in order to discover the associations between variables (Dalla Bona, 1997; 7). Those associations then are evaluated by the researcher for a proper discussion.

As in Inductive Models, a numerical approach is heavily affected from biases in the archaeological inventories, since the methodology directly makes use of it. Also, numerical approach models generally avoid using time as a concept. They do not address temporal considerations, whereas physical and cultural environment changes continuously change over time (Dalla Bona, 1997; 8).

Another problem with the numerical approach is its heavy reliance on statistics, where not all of the developers can use and understand statistical theory. Then the inferences might be incomplete and even sometimes wrong (Dalla Bona, 1997; 9).

2.4.2.2 Graphical Approach

The progress in information sciences enabled manipulation of vast data, as well as feasible handling. With this approach, different variables of a model are represented as in the same way as the layers of a GIS.

In this approach, statistics are used in evaluating the associations rather than discovering them. This approach does not investigate the unknown; rather it investigates more of the same, nor does it contribute to the system, but rather it validates known site distributions. On the other hand, this approach is helpful in discriminating between the majority of the sites having the same environmental characteristics, and the remaining sites (outliers), where these outliers might lead to new understandings (Dalla Bona, 1997; 11).

2.4.2.3 Intersection Method

If each variable of the model is considered as a layer (or a set) and if for each layer there are some preferred places for locating settlements, then intersections of those preferred locations depict the attractive places for locating settlements. On the other hand, such an assumption is not very accurate due to extremely complex human systems (Dalla Bona, 1997; 12).

In a deductive framework the intersection method, with the corresponding background information, can be mathematically defined as the following (Wheatley and Gillings, 2002; 169):

In a given location with given set of rules:
1) Sites tend to occur in flat areas. (Slope<10)
2) Sites tend to be close to fresh water sources. (Distance<1km)
3) Sites tend to be located on a particular soil type, A

4) Sites tend to be located to the south facing aspect. Then the rules will be:

$$M = (Slope<10) \cap (Distance <1) \cap (Soil = A) \cap (Aspect = South) \quad (2.3)$$

$$M' = (Slope>=10) \cup (Distance >=1) \cup (Soil \neq A) \cup (Aspect \neq South) \quad (2.4)$$

where M depicts the location with a site, and M' is the no-site.

2.4.2.4 The Weighted Value Method

If the assumption used in the intersection procedure is degraded then it can be rephrased as, the variables contribute in different weights to the decision of locating the settlements in a region. Weights are assigned to the classes or categories of map layers, according to some criteria. Then the sum is computed to reveal favorable lands for occupying the terrain (Brandt, Groenewoudt, Kvamme, 1992; 271).

The basic problem with this assumption is the lack of an agreement on assigning the weights. The weights are heavily influenced by the experience and understanding of the researcher (Dalla Bona, 1997; 12). Not surprisingly, different results can be obtained from the same predictive model by employing different weights.

The weights can be assigned by using ethnographic, ethnological and historical studies. Also, empirical studies can benefit from the definition of weights to be used in the model (Dalla Bona, 1997; 14), or the layer weighting approach can be coupled with the empirical data itself to create more objective weights (Brandt, Groenewoudt, Kvamme, 1992; 278).

2.5 Tools

There are various ways to perform predictive modeling. Each of them has both advantages and disadvantages over the other, and there is no clear rule of thumb for selection.

One or more ways can be used to construct the same predictive model with some success, but the performances of the tools are case specific (Ejstrud, 2002; 16).

There are four main categories of tools in the literature (Brandt, Groenewoudt, Kvamme, 1992; Dalla Bona, 1993, 1997; Whitley, 2001; Wheatley and Gillings, 2002). Some advanced extensions of the recent tools or some new tools will be introduced to the literature with the full integration of GIS with statistics software packages and advanced database structures.

The easiest, but surprisingly sometimes the most competent method is the Binary Addition. In this, areas having less sites than expected are assigned to 'zero' and the areas having more sites than expected are assigned to 'one', and those binary layers are added to each other for the result. This simplicity made the results easily understandable among archaeologists. It also permits the researcher to add his/her own variables as are thought to be important for the model (Ejstrud, 2002; 10).

The other tool with which to construct a predictive model is an extension of Binary Addition, called Weighted Binary Addition. As the name implies, the tool adds weights according to the importance of binary layers (Ejstrud, 2002; 10), but such weighting is not so easy and objectivity is hardly obtained while assigning the weights, as discussed earlier on.

Another tool is called the Dempster-Shafer, which is in fact base on Bayesian Inference. On the other hand, it contains a degree of belief as a *"...belief function* rather than a Bayesian probability distribution" (Wikipedia Dictionary, 2006b).

While constructing any model, an inexplicit assumption is made at the time that full information about the phenomenon can be set. In other words, for a particular location if there is 10% of the evidence favoring existence of a site, then the model will reveal 90% of the evidence for non-existence for the same area of concern.

Definitely such distinct assignments are not correct most of the time. By adding a level of *ignorance*, it is also then possible to deal with the incomplete data that cannot say much about a site (Ejstrud, 2002; 13).

By employing the Dempster-Shafer Theory, the model can make use of knowledge, and experience. An example given by White (2002, 23) can clarify the theory. The evidence of distance to water can be tested for (i) the sites near a water body (P[presence]), (ii) the sites are not near water body (P[absence]), and by employing Dempster-Shafer Theory a third hypothesis can be introduced as (iii) uncertainty about the evidence (P[presence, absence]).

An outstanding tool used in predictive modeling is called logistic regression, but this role is mostly due to its widespread use rather than its theoretical excellence. On the other hand, its consistency and feasibility over other tools makes it a popular tool.

Logistic regression creates a ‘prediction formula’ for the study area. It employs variables of ‘any scale’ of data, and provides a formula to correlate variables with site locations (Warren and Asch, 2000; 8). It can be divided into two: probability component, and score component (Warren and Asch, 2000; 18). Ratio-interval scale is the ordinary regression coefficients, but nominal- ordinal scale (categorical variables) operate on design variable codes.

If it is thought that the study area is divided into grids, then each grid (or cell or pixel) will have a potential score (assigned by logistic regression). This score is the integration of ‘attractiveness’ of the cell and the ‘disturbance’ to the cell (Duncan and Beckman, 2000; 36).

It was stated that the basic problem of the weighted value method was the formal determination of the weights. This problem can be reduced, if not totally solved, by logistic regression.

The weights can be adjusted with the use of logistic regression, and then can be used through map algebra (Duncan and Beckman, 2000; 42).

The use of regression analysis can also solve the deficits of other tools, where

1) Effect of the variables on the model is unanswered.
2) The extent of fitting of the model to the reality?
3) The extent to which the ordinal or continuous variables can be predicted (Wheatley and Gillings, 2002; 171).

Moreover, it (i) requires less assumptions, (ii) stands more powerful, and (iii) enables the use of mixed scale data (White, 2002; 24).

Logistic regression is stimulated from ordinary linear regression. Although there are many similarities, logistic regression is separated by employing dichotomous dependent variables, namely sites and non-sites. Then at least two assumptions of ordinary least squares regression analysis are being violated. First, the variance of error is not constant anymore, and secondly, errors in the model are not distributed normally (Allison, 2000; 8). Thus another model, logit model, is used in cases with relatively less restrictions as well as due to the capacity to analyze mixed scale data (Tabachnick and Fidell, 1996; 578).

Another concept, namely 'odds' should be explained before examining logistic regression analysis. Odds of an event or simply 'odds', is the ratio of the expected number of an occurrence of an event to the expected number of non-occurrence of the same event. Mathematically, if the probability of the occurrence of an event is 'p' then odds of an event will be:

$$O = p / (1 - p) \text{, or } p = O / (O + 1) \qquad (2.5)$$

In linear probability models, the probabilities are bounded by 0 and 1. While a transformation from probabilities to odds removes the upper bound, taking the logarithm (either natural or 10-base) of the odds removes the lower bound of the model (Allison, 2000; 13).

Then for one independent variable the model will be:

$$\ln(p/(1-p)) = \alpha + \beta x \qquad (2.6)$$

The equation 2.6 is also equal to ln(odds) or logit (p)

If the equation 2.6 is transformed then,

$$\exp(\alpha + \beta x) = p/(1-p) \qquad (2.7)$$

After leaving p alone, the equation 2.7 can be written as:

$$p = \exp(\alpha + \beta x) / (1 + \exp(\alpha + \beta x)) \qquad (2.8)$$

Then for a multivariate case the equation of 2.8 for presence of site will be:

$$p(y = 1 \mid x) = \exp(\sum \beta_i x_i) / (1 + \exp(\sum \beta_i x_i)) \qquad (2.9)$$

Although the tool is widely used, it has some basic problems. The first trouble is about non-sites. The prediction formula is heavily affected by the number of non-sites (absence), thus there should be an adjustment for it. Either a random selection of non-sites is obtained or this bias can be adjusted after modeling via taking the natural logarithm of the ratio of the sample sizes.

$$\alpha' = \alpha + \ln(n_2 / n_1) \qquad (2.10)$$

where α is raw constant; n_2 frequency of cases in large sample (non-sites) n_1 frequency of cases in small sample (sites) (Warren, 1990; 106).

Moreover, there is the problem of the assumption on the 0 value (non-site) pixels, where it is not always certain that a pixel coded as ‘0’ value is really free of archaeological material (Wheatley and Gillings 2002; 174, Warren, 1990). Furthermore, the interpretation of the coefficients of the logistic regression is not as easy as the multiple regressions (Wheatley and Gillings 2002; 175). Also, poorly presented categorical variables are problematic. The simple solution is screening them, and then weak categories can be combined to create stronger ones (Warren, 1990; 212).

Apart from the theoretical constraints, some practical problems put logistic regression in a deprived position. These problems can be listed as: (i) the lack of any check on the linearity of the regression model, (ii) neglecting the autocorrelation problem (Warren and Asch, 2000; 18), (iii) not performing a proper test of the model (Wheatley and Gillings 2002; 179). As a simple check, residuals of logistic regression can be examined for a patterning in order to see if the model contains systematic error (Warren, 1990, 211).

2.6 Testing

Predictive modeling of archaeological sites should be tested by two distinct procedures, one is the test of individual variables used to develop the model, and the other one is the general model checking.

2.6.1 Testing Variables

For testing variables, univariate statistical tests are used to compare environmental characteristics of site and non-site locations (Warren and Asch, 2000; 14). There is no single test for such a comparison.

Thus depending on scale, the Mann-Whitney rank-sum statistic test can be used for interval-ratio scale, and Chi-square or Fisher's exact tests can be used for nominal and ordinal scales (Blalock, 1979). There are also examples of the use of Kolmogorov-Smirnov two-sample test for the variables (Duncan and Beckman, 2000; 42; Thomas, 1986).

2.6.2 Testing and Validating Models

Comparison of frequency distributions of sites and non-sites on the axis of discrimination reveals the degree of separation between sites and non-sites (Warren and Asch, 2000; 19). This separation can be due to chance. In that case the log-likelihood chi-square value can be checked against a tabulated value to validate this. (Warren and Asch, 2000; 20) A simpler testing procedure is to construct a line plotted through points representing observed versus expected frequencies in which case it should approach a 45 angle with small residuals (Kohler and Parker, 1986; 431).

The test of the model accuracy is performed by running the training samples against sites and non-sites. Although this gives an estimation, another test should be performed for objectivity supplied by test samples (Warren and Asch, 2000; 20). In this particular case, training samples are obtained from the cells (pixels) that were used for constructing the model, whereas test samples are from the pixels of sites and non-sites which are intentionally withheld from the model while constructing it.

Another test procedure can be performed by applying the predictive model to a new study area, whilst it will suffer from the theoretical framework of the model, as discussed above.

There is another formal tool to evaluate the performance of the model, defined as Gain Statistic (Kvamme, 1988). The gain is defined as 1- %Area / %Sites, where the value is from 0 as the lowest value, to 1 as the highest. This statistic divides the model into two parts as high and low.

A measure of efficiency then is obtained and by measuring the proportion of sites and areas on the high part of the model (Wescott and Kuiper, 2000; 69; Ejstrud, 2002; 15).

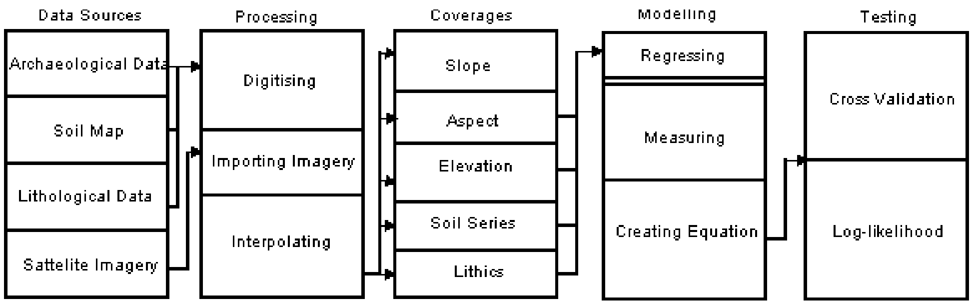
The outcome should always be checked by a close comparison with the data. It is not necessary to obtain valid results from formal test procedures.
The assumptions made while constructing the model can be wrong, or the necessary prerequisites of statistical tests can be missed, or any other uncontrollable error might have been introduced to the model.

2.7 Algorithm

A stepwise algorithm can be given to construct a predictive model. Although the given list below is not covering the whole set of procedures, the implementation is trivial. The list below is from Duncan and Beckman (2000; 36)

1. collection of primary data sets;
2. derivation of secondary data sets;
3. sampling of the environmental variables with site locations and random background samples;
4. exploration and statistical analysis of the two populations;
5. appropriate, the implementation of logistic regression analysis;
6. identification of significant variables to be used within the model;
7. creation of a model formula, which is a weighted sum of the significant variable values;
8. creation of the predictive surface from the formula;
9. internal testing of the model against the model training sample;
10. external testing of the model against an independent sample;
11. reiteration of the model formula and predictive surface given the testing results
12. continuous updating of the model given future discoveries.

The same algorithm can also be given as a flowchart, containing each step above, but with a more general idea. Although this flow is not the only way to construct a predictive model, it serves well and gives a clear idea of the development of such a model. Definitely, for each step some other sources and manipulations can be used, and tested accordingly (Figure 2.1).



Source: Warren and Asch (2000)
Figure 2.1 Simplified flowchart of a predictive model.

2.8 Constraints

As in all other modeling efforts, predictive models are deprived of perfect conditions, infinite sources, and complete datasets. Thus, before any attempt at modeling studies, constraints on those models should be explicitly defined.

2.8.1 Assumptions

It is stated that there are two basic assumptions of the system on which all other assumptions are based. Firstly, environmental factors shaped the selection of settlement locations in ancient times; and the secondly, those same environmental factors are represented today (Warren and Asch, 2000; 6; Dalla Bona, 1997; 17).

Another important point is made by Kohler and Parker (1986), where it is stated that the researcher is so ready to accept the assumption of locational behavioral is a multivariate function, but maybe very few proxies are enough to construct a model, and superfluous data distorts the model.

2.8.2 Problems

The problems of predictive modeling can be divided into two main aspects, problems of the tools that a model uses, and problems of the model itself, both related to data and to its interpretation.

The problems with the tools employed are basically recoverable and usually happened due to 'naive' use of statistics. The first problem is related with the assumption of normality, where a number of statistical tests are based on this assumption, and it is rarely checked (vanLeusen, 2000; 5-7). To overcome this, either data should be normalized, or some specific tests should be used which do not require normality assumption, such as logistic regression.

An important assumption of some statistical tools is the independence of observations. Depending on the data type, archaeological observations are geographically referenced. Thus use of spatial statistics is essential. On the other hand, some non-spatial tools employed are working with small problems, such as Spearman's rank size rule or principle component analysis (vanLeusen, 2000; 5-7). A criticism of the (inductive) predictive modeling is its use of sites as being independent of each other, whereas they are in fact parts of complex systems (Perkins, 2000; 133)
The choice of individuals is determined by local (independent, practical) concerns as well as global (system) concerns, such as defending, or competing (Ebert, 2000; 131).

The main problem of the predictive model is that each model makes use of limited number of variables based on environmental parameters (Dalla Bona, 1997; 15). Most of the time, quality of the model is determined by existing data. Although a researcher demands more data in order to make a full scale model, it is often impossible to obtain the complete set of data. Thus model turns out to be inadequate from the beginning. Then the model is subject to heavy criticism. On the other hand, sometimes such kind of environment based models can be informative, and can be seen as initial steps in a proper (or full) modeling.

In any case, overemphasis on environmental variables will create bias. For instance, although there seems to be a tendency for the locations to be close to water sources, there must be some cases where ancient people were keeping themselves away, due to erosion, regular floods or even the quality of water (Ebert, 2000; 132).

A big controversy in predictive modeling is environmental determinism. The first studies of this concept started with Montesquieu, Jean Bodin, Friedrich Ratzel, and Ellen Semple (Kohler, 1988). An informative discussion is provided by V. Gaffney and M. vanLeusen (1995; 367). In this text environmental determinism is given as the result of:

1) Limited availability of geographical datasets.
2) Limited functionality of GIS

An example of a dictation of environment over culture is provided by Meggers (1954; 809), where it is suggested that the decline in the decoration and manufacture quality of the pottery may be due to diminution of the environment where more people spent time searching for food. Such a diminution of a particular environment can be observed in a way that:

1) The advanced cultures around cannot diffuse into area.
2) The diffused cultures cannot preserve their positions.

The very basic ideology underlying environmental determinism is "…the situations we observe are not or arbitrary, but are caused" (Meggers, 1954; 805, Perkins, 2000). Thus the affect of environment can be mapped and measured to some extent in order to study the cultural differences or the behavior of a single one for a period of time. On the other hand, it should be clear that similar geographies might have different patterns of culture (Raikes, 1967; 8).

Then it turns to be the responsibility of the archaeologist, not the GIS itself to prepare (obtain) an appropriate dataset, containing geographically referenced variables as well as social variables. The problem is about defining space and adding it to a GIS (Llobera, 1996; 615).

The inverse case is, then, creating 'synthetic' landscapes away from the experience of an individual (Llobera, 1996; 613). Moreover, an archaeological interpretation should always be added as a result. Researcher should not merely map the soil and distance to water (Ejstrud, 2002; 16), where determinism in applications is implicit. Thus to break determinism, the first step should be to define it explicitly, and then to apply an interpretative (hermeneutic) approach to landscape studies (Llobera, 1996; 612).

There is also the question of to what extent humans respond to the environment, or challenges by transforming the immediate environs (Erickson, 1999; 641), where these transformations can be summarized as devegetation, soil loosening, soil water and groundwater changes, construction, and accelerated soil erosion (Butzer, 1982). With the general understanding of the problem and available dataset, as well as with the existing capabilities of the analysis systems, determinism cannot be broken, even cannot be avoided.

Another problem is the quality of data sources, whether they are environmental or not (Dalla Bona, 1997; 15). This also includes the resolution, temporality and cost. The data should be detailed enough, but at the same time should be feasible in terms of processing. For general modeling, up-to-date data is always preferable.
This is also true for archaeological predictive modeling, but an old reference might be selected since there is massive destruction of archaeological data in the name of land construction, agriculture and various other activities.

Cost is one of the mainstream problems of data, where a predictive model should not go beyond the cost of an intensive archaeological survey, where a proper survey can produce much more accurate results.

Another problem is the selection of the scale of study area, and integration of data having different scales. At the end of a generalization (from big scale to small scale), a model can lose its precise character, and a move from small scale to big scale for the same study area might result in redundant data.
Those transitions are sometimes necessary in order to find a compromise between different scales of data. Selection of the scale is determinant on the analysis of the landscape. Wilkinson (2003; 9) states that viewing landscape according to broad zones, 'mountain landscapes, or landscapes of tells' makes it comprehensible. Each landscape has a 'sign' that shows particular property. For instance tell dominated landscapes can be claimed as static.

Compromise is a concern not only for the scale, but also for every component of the model. For that reason, integration should be obtained for date (currency) of the data, format of the data, and the integrity of different data formats, as well as the integration of non-spatial data to the model as attributes (Dalla Bona, 1997; 26-32).

Although it might not be a problem for some, subjectivity is another matter. Selection of variables, construction of the model, and final interpretation is based on the experience, knowledge, and even political stance of the researcher. From a westernized point of view, marginal areas should have less productivity resulting in less density of occupation, but people can develop certain technologies or strategies in order to overcome the negative effect of the environment. Marginality can be exemplified by poor thin soil, high altitude, low temperature, and various others, and it can be claimed to effect the settlement locations in a negative way.

Although it is true to some extent, there are some contradicting examples. For instance, a 'marginal' area, Lake Titicaca in South America above 3800m of sea level was and is densely populated (Erickson, 1999; 636)

Specific to the locational behavior of ancient humans, each archaeologist might have an idea about where the locations of settlements could be. Then the question raised is whether there is a need for a formulization of this subjectivity, which then can be grasped by others. By means of a well defined formal view, the implicit understandings of researchers, whether or not related with the subject, can be measured, tested and, more importantly, repeated.

If this is the case, multivariate techniques offer such tools, but it is still the archaeologist at the very end, who should reach a conclusion.

One important point, likely to be missed, is the character of the sites. Ebert (2000; 131) states that the time spent between sites is much more important than the time spent within the site. It is that travel time in which human interacts with the environment. Thus for the ancient human, sacrifice from environmental suitability might be the case for the sake of the stability of social network.

The problems with the interpretations of the results constitute a main part of modeling studies. One problem is about the presentation of the result of a predictive model. The implicit terms like 'high/medium/low potential areas' or 'areas of favorability /non-favorability' are not informative when unaccompanied by a proper explanation. Moreover, the breakpoints between those categories are rarely defined (Dalla Bona, 1997; 15).

The resulting map of a predictive model consists of pixels indicating probable site locations, but it should be read two times. High probability of a pixel does not necessarily mean that a site exists in this particular location.

Archaeological concentrations are very vulnerable to side-effects in the active landscapes, and they constitute a very small part of the total landscape. Thus the high probability does not show the existence of a site, but it shows the potential (Duncan and Beckman, 2000; 56).

Predictive modeling is a continuous process with endless iteration with the addition of each archaeological and environmental entry as well as other theoretical developments (Duncan and Beckman; 2000, 56). It is GIS that enables the researcher to perform iterations in a feasible manner, but although GIS is new, predictive modeling is not new. It is true that computer hardware and software brought efficiency to the analysis of data, but it is the critical thinking in archaeology that is creating models (Ebert, 2000; 129). The (inductive) model should be accompanied by an explanation rather than a simple and fancy colorful presentation of the data. It should differ "… whether it is done with GIS or a stack of semitransparent map overlays on a light table" (Ebert, 2000; 133).

The failure of early models then can be summed up as "failure address management needs, lack of specificity, poor use of existing data, ineffective or biased sampling designs, inappropriate statistical analysis techniques, failure to collect inventory data suitable for the development of a predictive model, development of models using non-replicable techniques, lack of comparability of and inappropriate use of environmental variables, lack of phasing to allow for model testing and refinement, and failure to use such technical aids as remote sensing and geographic information systems (GIS) to streamline model development" (Judge and Sebastian, 1988:10).

One very minor problem is omitted most of the time. The existing GIS software employs different algorithms for creating data sets, and even when they use the same algorithm the output can differ. Although it is not a big issue for larger scales, a focused study can suffer from these differentiations. Thus critical readings of the secondary variables are definitely needed (Altschul, 1990, 230)

2.8.3 Errors

Apart from problems just outlined, there are also errors. A model can predict that a particular location contains a site, when in fact it does not, and a model can also predict that a location does not contain a site, when in fact it does.

The former error is termed as wasteful error, and the latter is called gross error. The first type of error results in wasting time and money, whereas the gross error results in losing the archaeological property. Then the success of the model can be measured by the amount of error it has, especially with the amount of gross error (Moon, 1993; 2).

2.9 Future Studies

Traditional predictive modeling efforts are part of inductive ideas where known site locations are put into a model together with the environmental parameters. The other method (deductive methodology) investigates the decision making processes of humans. If ancient decision making mechanisms can be modeled, then the same tools can be used for archaeological predictive modeling (Whitley, 2001; 2).

As discussed earlier, human behavior is a set of extremely complex systems. Thus, in order to have a fully developed comprehensive model, a researcher should raise clever questions about the phenomenon. As expected, those questions can be handled with more elaborate thinking. In this respect deductive models seem to be a better choice to understand the decision making process of ancient human about location selections.

Bayesian inference allows expert knowledge and experience is to be included in the model, and fuzzy logic allows the manipulation of uncertain data. Such kind of procedures can cope with implicitly defined sets, but maybe the main advance can be obtained by coupling landscape reconstruction with given techniques in archaeological predictive modeling. It will be then possible to obtain a better model (vanLeusen, 2000; 5-14).

CHAPTER 3

STUDY AREA: ARCHAEOLOGY

3.1 General Locational Information

The area in concern comprises the Lake District, in the southwest of Anatolia as well as southern extensions of the Konya Plain bounded by the Taurus Mountains and western extensions of the Plain (Figure 3.1). There are significant differences as well as some similarities through the region. The rationale behind the selection of more than the core area is to include peripheries and beyond of peripheries to some extent with cultural similarities in order to search for a very general layout for a specific period of time, as well as to increase the limited number of samples obtained from the archaeological inventory.

Figure 3.1 A wide view of study area

The Lake District (Figure 3.2), as the name implies, full of lakes with various characteristics. The district is limited by Taurus Mountains to the South, whereas North North-West is bounded by the political border of Afyon and Denizli (Duru, 1999; 165). The East of Lake District is not so well examined as the west part, but for the moment there seems to be significant archaeological differences between east and west (Duru, 1999; 166).

Figure 3.2 A focused image on Lake District

The Konya Plain (Figure 3.3) is a part of the Southern Anatolian Plateau to the south of Tuz Lake. With a rainfall of 250mm/year, it is characterized as an arid region of Anatolia. On the other hand, southern parts of plain have wetter conditions (Yakar, 1994; 180).

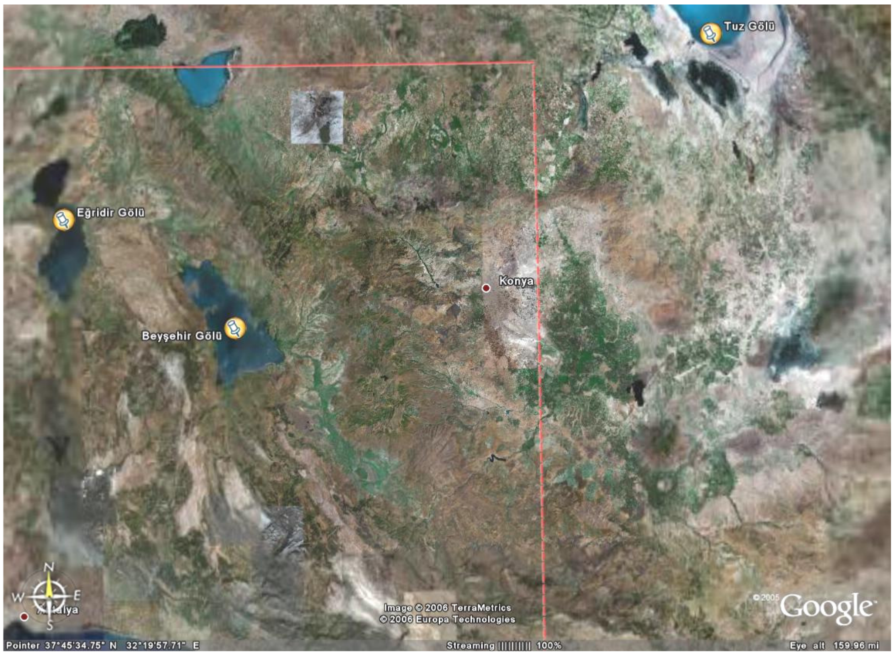


Figure 3.3 Western extensions of Konya Plain

Determination of exact boundaries of the study area is given after extensive discussion of the domains of area, namely spatial and temporal domains. Drawing the limits of the study area, both spatially and temporally, is of immense importance since determinants of settlement patterns can be heavily affected from the choice.

3.2 Archaeology of the Area

The Southwest of Anatolia, together with some parts of the south-central Anatolian Plateau is a well studied and documented area. Specific to the timeframe of this study, there are some major sites excavated and published (Bordaz, 1965, 1966, 1968; Mellart, 1970; Bordaz and Bordaz, 1976, 1982; Duru, 1994). Some surveys have produced valuable information about the distribution of sites (Kökten, 1952; Mellart, 1961; Solecki, 1964; Bordaz, 1965; French, 1970; Özsait, 1986; Baird, 2000) (Figure 3.4, and Appendix-B). It is also important to emphasize that more than 65% of the settlements have produced Neolithic and Chalcolithic material (Yakar, 1994; 144). Recent C^{14} dates for major sites in the study area are given in Appendix-C.

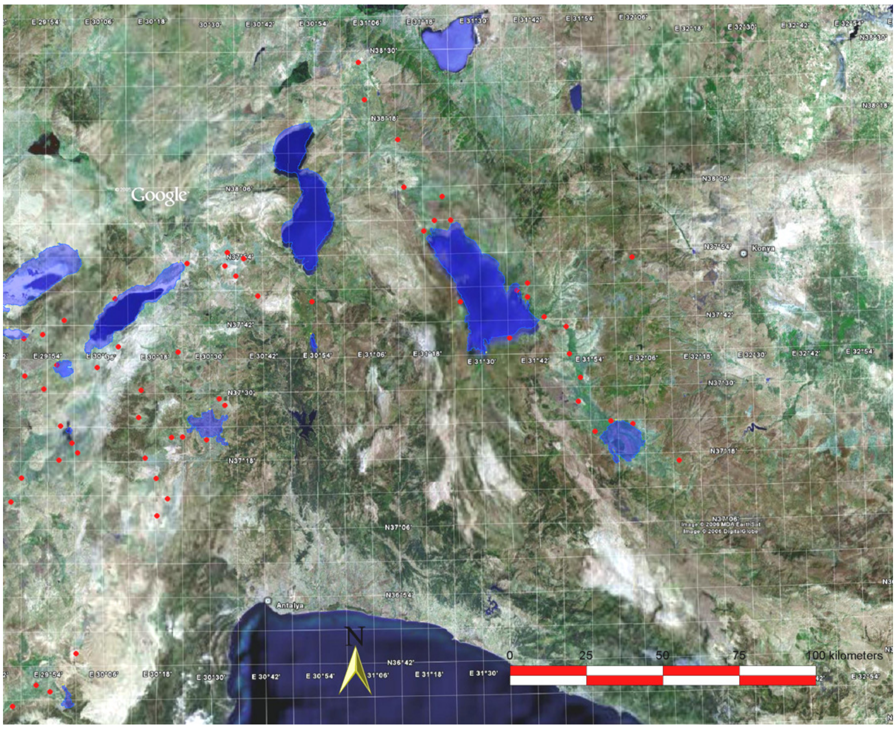


Figure 3.4 Distribution of sites over study area.

The neolithization of the study area is also a source of problems. As a general observation, Aceramic Hacılar is not really aceramic, and although pottery making was not known at Can Hasan III, from the very beginning in Bademağacı people were using pottery. This might be an indication for movements from the south of the Taurus where there are traces of pottery use even before Neolithic. It is also supported by the idea that the progress at south of the district is slower than north part (Duru, 1999; 166-168).

Then it is convenient to make an inference comparing north-south rather than east-west. Duru (1999; 168) states Bademağacı-Höyücek and Hacılar-Kuruçay produces grey-beige paste, red or brown washed, burnished and oven-dried pottery. Also the potteries are common in Bademağacı EN 5-3, Höyücek ESP, Kuruçay 13, and Hacılar EN. Moreover he explicitly states that the Lake District produces a community in terms of geography, but no such community can be suggested for the culture of the area, whereas it is clear that cultures have to be defined by more than regional and sequential pottery styles.

Duru (1999; 169), then puts the Beyşehir-Suğla basin (Figure 3.5) of the region as a transition region, but not a core part of the Lake District, and defines similarities between the regions on the evidence of stamp seals, arrowheads, and red paints on the floors of houses.

Figure 3.5 Basin boundary of Beyşehir-Suğla Region.

In fact, not only within the area concerned but also the overall Anatolian Neolithic is significant with its diversity in economy and in craftsmanship (Mellart, 1972; 283), and those can be responses to different life spaces, as well as the natural cultural diversities. A broad look at this diversity reveals a reliance on obsidian at Aşıklı Höyük, an extensive hunting at Suberde. At Aceramic Hacılar, there was an emphasis on construction, and in Aceramic Can Hasan it was on bone carving, and finally in Çatal Höyük specialization was on everything (Mellart, 1972; 283)

Such diversity is too complex to be put into a model, and any working effort will be too simplistic and general. A properly selected study area in well defined timeframe can reveal some, if not all.

There is also a need for an understanding of main characteristics of sites. By this way it is possible to determine the spatial and temporal domains of the model, and to discuss the result of the model from an archaeological perspective. Thus cognizance should be taken of individual sites without losing track of the community of sites, in order to understand the determinants of the site locations as well as their cultural associations. Because there are not many excavated sites in the area every piece of contributing information is important.

In the area, many settlements might have been established but a very small amount has survived as mounds since:
i) The construction materials used for settlements did not survive
ii) Short periods of occupation resulted in low mounds which were later on destroyed by nature or human affect (Yakar, 1994; 294).
iii) Burial beneath later occupation
iv) Failure to recognize the presence of early material on multi-period sites
v) Inadequate archaeological survey
vi) Inadequate reporting or publication of archaeological evidence

3.2.1 Suberde

Located 11km southeast of Seydişehir, Suberde is placed on a flat limestone ridge called Görüklük Tepe (Bordaz, 1965; 31) (Figure 3.6). The location and destiny of the site was closely related to the levels of Suğla Lake. A stream was bringing water from Beyşehir Lake, and when the Suğla was in its highest water level, Suberde/Görüklük Tepe must have been an island. Also it is highly possible that the abandonment of the site is due to this water activity in the Early Neolithic (Bordaz, 1968; 44).

Figure 3.6 Location of Suberde in relation to old Suğla Lake

There are 3 distinct strata reported at the site. The upper layer is disturbed by later burials. The bottom two strata are prehistoric. The second layer (or the Upper Prehistoric Layer) is 0.75m thick and mainly red-brown loam (Bordaz, 1966; 32). It is around 1300 square-meters (Bordaz, 1968; 45). There exist the poorly preserved remains of mudbrick walls and plaster floors. The bottom layer is around 2.00m thick with brown loam (Bordaz, 1966; 32). The total extent of the site is approximately 5000 square-meters.

There is no indication of sherds or ceramic vessels in the prehistoric strata. With the evidence of plaster floors the site then can be put earlier than Çatal Höyük VIa and contemporaneous with Aceramic Hacılar (Bordaz, 1966; 32). The first settlers lived in houses built of perishable material (Singh, 1974; 80).

The significant character of Suberde as a site is the existence of huge number of animal bone fragments (Bordaz, 1973; 283). Those include hunted animals as well as potentially domestic ones (Bordaz, 1966; 33), but the only domesticated animal is the dog (Bordaz, 1968; 57). Coming to the flora; cedar, pine, juniper, chestnut, birch, and poplar were abundant (Bordaz, 1968; 58).

If there is no significant loss of archaeological data, with its hunting and sedentary economy, Suberde had a group of conservative hunters who borrowed some architectural elements from farming villages (Bordaz, 1968; 60). It is possible if the environment is rich enough to sustain such a system, and if the abandonment is due to depletion of sources it is a well marked time that is extremely useful, or if it is the lake levels that caused the abandonment of the site, it is still valuable information in terms of reconstructing the lake levels which is quite important for the area.

3.2.2 Erbaba

It is located on a natural hill near Beyşehir Lake. It is approximately 80m in diameter and about 5000 square meters (Bordaz, 1969; 59) (Figure 3.7). It has no post Neolithic occupation (Bordaz, 1973; 283, Bordaz and Bordaz, 1976; 37). Location is suitable for prehistoric farmers. Open vegetation does not require extensive deforestation, which is also good for obtaining other wild sources (van Zeist and Buitenhuis; 1983; 48).

There are traces of poor quality grey plaster floor. Also there is evidence for entrances from the roof (Bordaz, 1969; 60, 1976; 37). It can be claimed that the site provides a tie between Hacılar and Çatal Höyük (Bordaz, 1973; 287, Bordaz and Bordaz 1976; 42). Also, the economy clearly indicates that it is later than Suberde (Bordaz and Bordaz, 1982).

There is an evidence of domestication and the archaeological inventory is similar to Çatal Höyük, except for lentils which are absent in Çatal Höyük. Sheep, goat and cattle seem to be domesticates as well, while 84% of faunal remains are mostly sheep and goat (Bordaz and Bordaz, 1976; 40).

Figure 3.7 Erbaba and Beyşehir Lake.

Chipped stone industries also reveal a significant difference between Erbaba and Suberde. Projectile points in Erbaba are 3% whereas they make up 20% in Suberde, which might be an indication of difference between hunter and agriculturalist economies (Bordaz and Bordaz, 1976; 41).

The locations of both Erbaba and Suberde are not in the core of Lake District, but at the periphery. According to Duru (1999; 166) this situation makes these sites extensions of Konya Plain.

3.2.3 Hacılar

It is located 26km southwest of Burdur, and close to Burdur Lake (Figure 3.8). A spring is observed in the limestone, where also the ancient site is located. The Koca Çay stream, passing from the west of the site and draining into Burdur Lake contains sulphur, arsenic and salt. This puts both the lake and the stream away from the use for vital purposes (Mellart; 1970; xii).

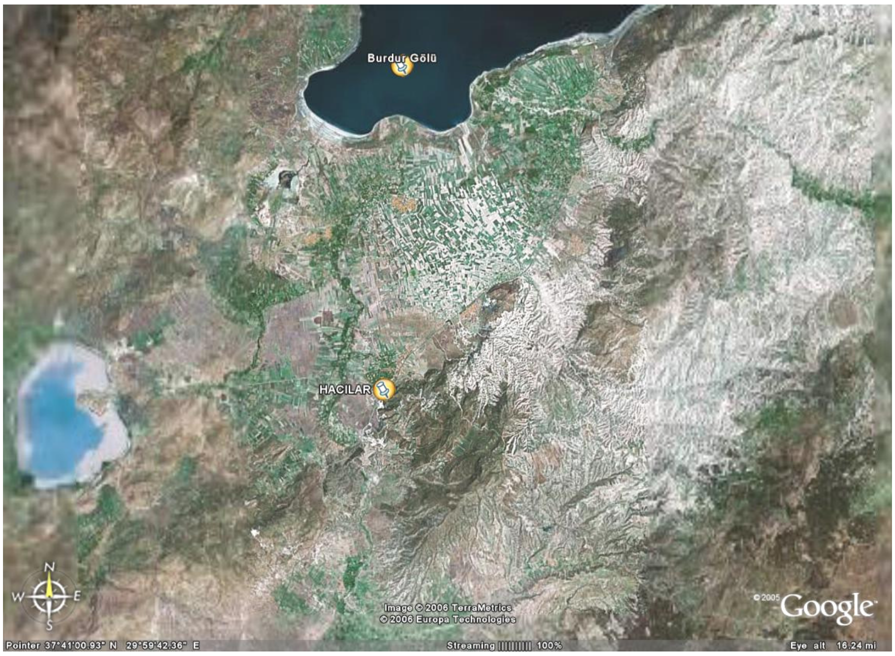


Figure 3.8 Location of Hacılar and other lakes.

In Aceramic Hacılar, there is an indication of agricultural practices (Singh, 1974; 66). On the other hand, there is not much information on the domestication of animals except dog (Mellart; 1970; 5).

The interesting thing about Hacılar is its abandonment for more than a millennium. The 'newcomers' brought a fully developed culture with them. First they built the new settlement on virgin soil, then two sites were merged destroying some elements of the aceramic phase (Mellart; 1970; 8). Such a re-visit raises a basic concern. How is it possible to locate the 'new' settlement on a very close position of the old one? The practical answer is the existing remains of the aceramic phase led the people to re-locate the site. On the other hand, the thinness of the walls of aceramic period suggests single storey, with no evidence of postholes even in the biggest rooms. Moreover stone foundations were found only under the heavier walls (Mellart; 1970; 4). It is then questionable to find the traces of such a basic architecture after more than a millennium. The other possible explanation is the protected link or memory with the ancestry, where the location of the 'old' site was never forgotten, and even there were visits to the site, but this also brings the isolated community preserving its oral tradition. This idea is open to any criticism, and one can hardly prove its truth. The extremist answer is the probabilistic one, but the most optimistic explanation in terms of human location behavior is re-locating the site due to the same attractiveness was valid, with the implication that no environmental change occurred to shift this attractiveness.

Another concern with Late Neolithic Hacılar is the apparent effort given to defense (Mellaart; 1970; 10). Defense can be against human or animal, or both. Other needs, such as creating demarcations might very probably result in simple structures rather than fortified enclosures. On the other hand, in any case such a risen need is a response to a happening. It might be a response to human if the social system is degraded so that there is disorder, or it might be a safety measure against attack by big wild animals. Such attacks do not start suddenly, but might be due to lack of food or other resources for those animals. The depletion can be a result of environmental change either naturally happening or happening due to human activity such as heavy deforestation for fuel and construction, where Mellart (1970; xii) states that remains of deer, wild cattle, and representations of leopard, and bear shows the forest was not so far away in the Neolithic and Chalcolithic.

3.2.4 Kuruçay

It is located one of the hills around Burdur Lake, where the topography is much undulated (Harmankaya, Tanındı, Özbaşaran, 1997) (Figure 3.9). Kuruçay 11 is characterized by fortification walls with towers. Duru (1994; 12) states that a well established fortification reflecting a strong architectural tradition protected the inhabitants of the site from external dangers. If this reconstruction is true then the fortification must be a response a threat more than an animal attack. Hacılar VI, was contemporary with Kuruçay 11, and is thus expected to have an independent fortification, since these sites are so close to each other (Duru, 1999; 168).

Figure 3.9 Location of Kuruçay and Hacılar, and relations to Burdur Lake.

Kuruçay 11 is dated to the Late Neolithic, and after level 11, the site is reported to have shifted to the south. Level 10 of Kuruçay belonged to the Early Chalcolithic. Since a fortification wall is not reported with the beginning of the Early Chalcolithic no more speculation can be made about a possible struggle within the social system in the area, but it should be kept in mind that the transition to independent fortification is sudden (Duru, 1994; 15).

Architectural tradition from Early Neolithic to the end of the Early Chalcolithic (12-8) is stated to be continuous and minor differences can be acceptable (Duru, 1994; 15). Moreover, in terms of technique there were still minor differences (Duru, 1994; 14).

3.2.5 Höyücek

It is not an ordinary mound with regular urban elements, but comprised special buildings (Figure 3.10). Thus the strata are named accordingly (Duru, 1995; 449). There are significant similarities of pottery of Early Neolithic of Höyücek, Kuruçay Level 13, and Aceramic Hacılar. There are also architectural parallels between the Shrine Phase of Höyücek and Ceramic Hacılar (Duru, 1995; 467). There are clues of a temporary settlement in Early Settlements Phase for Höyücek (Duru, 1995; 470). The same discussion can be made for the Sanctuaries Phase, at least for a period of time (Duru, 1995; 471). Thus Höyücek turns out to have been a cult centre rather then an ordinary settlement as indicated at the very beginning.

Figure 3.10 Höyücek and other major sites in area.

CHAPTER 4

STUDY AREA: GEOGRAPHY and ECOLOGY

4.1 Geography

The study area is characterized by lakes, and various basins with several types of hydrological feedings as well as valleys of differing sizes. Overall, the area is limited by the Taurus Mountains at the South, whereas North and West are bounded by a line following the outer depressions of geological features as well as the ancient sites thought to be located at the periphery forming cultural limits. The east end of the area, on the other hand, is hard to determine and extremely speculative. Passing from the Beyşehir-Suğla Region to the east, there starts the great plain of Konya, where it is stated that there was a pluvial lake covering the Konya Basin at a depth of 20-25m (Erol, 1971; 13).

If the (ancient) Konya Basin is excluded from the area in concern, then the remaining part forms what is today called the Lake District. This term is not cultural but geographical. Thus the boundaries of this study have been selected using both barriers (Taurus Mountains) as well as cultural affiliation.

The south-west end of the area, the western part of the Western Taurus, is characterized by the existence of individual chains that are perpendicular to the range. The Western Taurus can be divided into two limbs forming an angle near Isparta. The western limb is the Lycian Taurus, and the eastern range was formerly called the 'Taurus Occidental' At the East of the Antalya-Isparta line are NW-SE strikes, parallel to the range prevail, whereas at the Lycian Taurus, the strikes are oriented NE-SW, which is also related to the existence of nappe systems (van Zeist, Woldring and Stapert, 1975; 57). Due to post-alpine epeirogenetic movements, rift valleys occurred, where lakes possibly filled those valleys, such as Eğirdir Lake located in one of the N-S oriented rift valleys (vanZeist, Woldring and Stapert, 1975; 58).

Including the Beyşehir-Suğla depression in the Lake District of Anatolia, Eğirdir-Kovada Basin, Acıgöl Basin, and Burdur Basin with similar traits in terms of physical geography to some extent constitutes the broad study area.

From a wider perspective, a broader area than Lake District also can be divided ecologically for the provinces of Antalya, Burdur, and Isparta:

1) The Western Plateau: It is an area with closed drainage basins and large depression plains. Plains and basins are suitable for wheat and barley cropping due to extensive alluvial and hydromorphic soils.
2) Lakes Zone: It is comprised of the drainages of three main lakes. Cultivation is performed in alluvial plains mainly radiating from the lakes; the rest is mostly rock outcrops.
3) The Coastal Zone: It is basically a strip infiltrating to the foothills of the Taurus. Cereal cultivation is affected by annual flooding.
4) The Mountain Zone: It is a part of the Taurus system where the terrain is extremely undulated. Arable land is limited to lower slopes and mountain shelves (Allan, 1972; 214). Karstification is a characteristic of the Taurus. Although it is a glacial process, it is also active in post-glacial epoch. Karstification creates a dynamic landscape, especially for hydrological processes (Erinç, 1978; 104).

According to this schema, the study area mainly focuses on the Lakes Zone as well as some marginality of other zones, with a continuation to the east covering the east of Konya Plain.

4.2 Ecology

Apart from the archaeological studies, geological, geomorphological and palynological studies can help to identify significant shifts that happened in history, where those shifts might be the reasons for changes in settlement patterns. On the other hand, such kind of studies which are applicable to archaeology are few, but still helpful while doing regional research. A study performed in the east-central area of Anatolia (Woldring and Bottema, 2002) reveals temporal zones of homogenous environment, and sheds light on possible conditions of the environment. There exist five zones according to pollen studies: 13000-10860 B.P., 10860-7900 B.P., 7900-3870 B.P., 3870-1650 B.P., and 1650-0 B.P. It is also given that around 17000-12000 B.P. arid conditions of a Late Glacial Period are detected, and with the end of this phase, lakes and marshes had been developed in basins at various times (Wilkinson, 2002; 22). Another set of zones based on pollen stratigraphy is given by Eastwood et al. (1999) for Gölhisar Lake (37 8'N, 29 36'E) in the Burdur province as ~9500-8600B.P., ~8600-3500B.P., ~3500-3000B.P., ~3000-1300B.P., ~1300-200BP. Also it is claimed that landscape around Gölhisar indicates a major pine dominated forest, and important local and regional variations exist in forest composition from Early to Mid Holocene. Similar results were obtained by vanZeist et al. (1975) for south of Lake Beyşehir. Those results are used for constructing the temporal domain of the study area. Although, such reconstructions are local, results are still helpful for validating the results.

In fact, a general reconstruction of forests showed that at roughly 8500BP, the interior Anatolia was covered with Pinus silvestris and Betula, which are the species in North and North-Eastern Turkey (Erinç, 1978; 97).

4.3 Lakes

The study area is significant with its various sized lakes, each with different characteristics (Figure 2.3). The main lakes of the area are Beyşehir, Eğirdir, Burdur, and Acıgöl.

Their immediate surroundings both favored and repelled settlements due to their dynamic character. According to Lahn (1948), Suğla was disappearing once every 12 years in the early 50s. Although modern lake levels are heavily affected by agricultural activities and artificial channels by-passing rivers, the geological history of the study area is worth looking at. Again Lahn (1948) gives a brief history of the development of the Lake District. Accordingly, extensive shallow lakes were covering the south-central Anatolia, connected to each other in the Pliocene. By the early Pleistocene basins were closed, and lakes were isolated. The isolated lakes then either developed a subterranean karstic outlet due to soluble limestone (fresh-water) or they were not able to develop an outlet in which case their water became saline. Some lakes disappeared altogether, as did broad Pleistocene lake in the Konya basin.

Evidences are clear that lakes were there in the Pleistocene, where no weathering was observed in the samples taken from the centers of the basins (Cohen, 1970; 120). Farrand (1964; 150) asserts that Beyşehir-Suğla depression fault was "occupied by an extensive shallow lake" which is at1180m above sea level.

It was observed that the basins of Western Taurus contained more lakes than today. Also they had higher lake levels, disregarding human impact such as artificial channels or conscious draining. Both indicate pluvial conditions, and are important in terms of settlement locations since the recessions produce fertile soil behind as well as being physical obstacles to locate settlements (vanZeist, Woldring and Stapert, 1975; 61, Erinç, 1978; 99).

The lakes are both fresh and salt water; Beyşehir, Eğirdir, Söğüt, Avlan, Kestel being fresh-water lakes, while Eber, Akşehir, Acı, Salda, Burdur are salt-water lakes.

According to Saraçoğlu (1962; 308) lakes close the Mediterranean are less prone to water level fluctuations. Those with external connections, either outlets or fresh-water inlets, have fish, while those that are closed have brackish-salty water and there is no fish inside.

The fresh-water lakes are assumed to be having subterranean drainage, as well as karst processes that are observed in limestone mountains of Lycia (van Zeist, Woldring and Stapert, 1975; 57). Thus, overall fluctuation of lake levels can be due to karstic terrain (Eastwood, Roberts, Lamb, and Tibby, 1999; 690). Apart from creating a dynamic framework, it is claimed that limestone also distorts the C14 dates obtained from the area (Eastwood, Roberts, and Lamb, 1998; 69).

4.3.1 Evidences of Lake Levels

Cohen (1970; 120) states that the decrease in the lake levels are a result of evaporation, or lower surface drainage, rather than an increase in the underground drainage. This statement covers the basins of Burdur and Konya-Ereğli. The studies examine footprints of lake levels above modern levels, where they are evident with shorelines, raised beaches and terraces (Cohen and Erol, 1969; 393).

Although changes are evident, the fluctuations of wet areas are not as significant as the ones which are desert-like margins (Butzer, 1964, 25). Thus comparative examination of the levels is a must, and it is logical to compare and contrast the behavior of lakes in the same climatic zone, such as Burdur and Beyşehir (Cohen, 1970; 121).

Burdur Lake level reached about 860m, where there seems to be a recession until the Quaternary (Kökten, 1952; 187).

There is a continuous recession of lake levels through time, though not all of the movements of theses lakes are visible. Thus during the study only available data is used in terms of reconstructing the ancient lake levels.

4.4 Archaeological Inferences

The dynamic environment of the study area puts the archaeological evidence in an interesting framework. The area does not have the typical characterizes of

Mediterranean climate, but it also does not have the arid conditions of Central Anatolia (Altınkale, 2001; 19). Due to karstic landscape, fluctuations of lake levels can be examined physically, but they should also be examined culturally in order to reveal the role of lake level changes for the ancient people of the area as well as to make use of the environmental reconstructions.

Although recessed land is fertile in the immediate vicinity of a lake, there are some problems in terms of locating the settlement. In the area, soil is so wet that it can not easily carry the weight of a house. Thus, a minimum distance exists to the vicinity of the lake. Also, insects can create immense problems such as malaria. For Swiss lake dwellings the ultimate action was to drawback the settlement to the dry land. Each layer was an indication of such a withdrawal (Muller-Beck, 1979; 250).

The non-existence of the pre-Neolithic material can be explained by the lake levels of basins, which were not permitting access for settlements. Paleolithic artifacts are found in uplands, around the Burdur Basin, and to the south of Lake District the zone where the Taurus stands (Cohen; 1970, 120). With the start of recessions in the Konya basin and in the Burdur Basin the earliest evidences of Neolithic shows up, and it is only the Dervişin Hanı tools that give evidences of pre-Neolithic culture (Cohen, 1970; 131).

Hacılar was also affected by this dynamism. There are three raised beaches existing between the present lake level and Hacılar, and there is 3,5m of accumulation since the first aceramic establishment of the site. Two recessions might have occurred before the Late Neolithic of Hacılar. Although it is stated as the second one is not earlier than 6750 BC, or the abandonment of the site, the new cal B.C. is 7550. In any case, the abandonment of Hacılar can be explained with the intrusion of lake water to the site, or due to flooding of the area made the site inhabitable (Cohen, 1970; 130).

Suberde was occupied when it was an island. Alan Höyük and Ortakaraviran Höyük were clipped by waves during the occupation. The lake dominated landscape was surely important for the ancient inhabitants of the area (Farrand, 1964; 153).

CHAPTER 5

DOMAINS

Although some basic archaeological and geological information is given, the main compilation is made for defining the study area both in spatial and temporal context. The boundaries of areas that are prone to predictive modeling are arbitrarily selected (in perfect rectangles or sudden cuts of a study area). The study area is also determined arbitrarily for the ease of use of projection systems, where the boundaries are the lines of projections which are geological, but not archaeological, or very little or, indeed, no discussion is made about the peripheries.

On the other hand, there is no agreement on the boundaries of cultures. Indeed, one can hardly define the exact area of an ancient settlement, if not fortified. Even it is the case; inner core of the settlement cannot fully represent the interaction of people with the environment or with other human systems. Moreover, if a predictive model is environmentally derived, which is the case most of the time, there is problem of ignoring areas where ancient people were interacting with their immediate environment. This raises the issue of transhumance, which was usefully be defined as "… a passive reaction to environment" by Raikes (1967; 13). Although a settlement is said to be permanent, its population can be dynamic. According to Allan (1972; 221) movement was alive between the villages of south-western Anatolia, and the population was doubling during the specific periods of the year.

Another fallacy of a predictive model concerns time. Like space, time is hardly discussed in predictive modeling, or else a very wide timeframe is used to produce a precise model, disregarding the environmental and cultural changes, otherwise such a frame is divided into smaller components. A better approach might be to make use of time as a fourth component in a GIS.

Despite the fact that revealing the boundaries of space and time is not easy and beyond the scope of this thesis, there should be at least an attempt to put the predictive modeling into a better context.

5.1 Spatial Domain

Endless discussion can be made for the spatial domain of the area, where not a single researcher can draw exact boundaries of the cultures. Even it is drawn, the periphery will contain areas affected by the neighboring culture zones. The choice of study area can reduce the effects of some variables while enlarging others (Van Leusen, 1993; 114).

The modeling will produce better results if it is applied to a core area with minimal disturbance of other cultures. While it is impossible to detect such core areas, some general terms as ‘Fikirtepe Culture’ or ‘Urfa Culture’ are suggested in archaeological literature. Also there is a tendency to put Çatal Höyük, Can Hasan and Tepecik Çiftlik to Central Anatolia, Suberde, Erbaba, Bademağacı, Höyücek, Kuruçay and Hacılar to Lake District of Anatolia (Özdoğan, 2002; 94). Neolithization of the Lake District has a different style from Konya Plain or Eastern counterparts where most probably the origins are also different (Duru, 1996; 59).

Although such divisions are useful in terms of research and understanding, not much discussion is made concerning the transition zones. This can be also observed by viewing the distribution map of sites for the Neolithic and Chalcolithic Periods of Anatolia (Figure 5.1). Thus it should be approached critically that if such distinct zones exist or if there is not enough information to see a complete transition. For this reason it is essential to perform researches at those void areas with specific questions and with prior information of possible site locations.

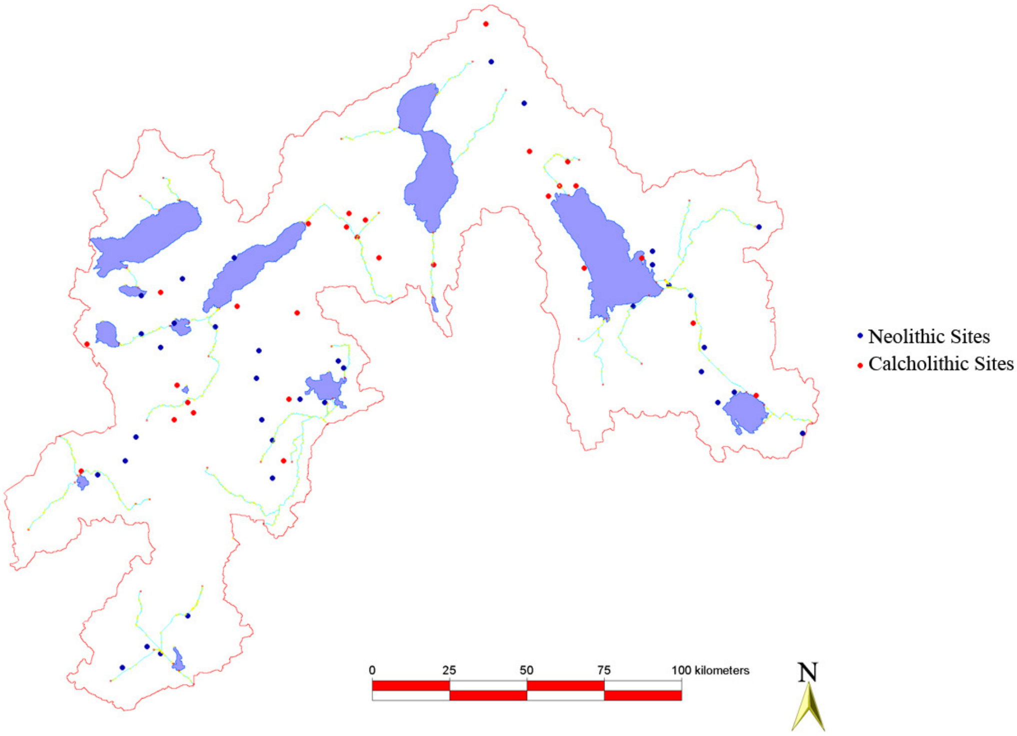


Figure 5.1 Distribution of Neolithic and Chalcolithic Sites in the study area

The study area that the predictive modeling is applied on consists of geographical limits of basins of lakes and discussions of previous researchers. The discussion below aims to reveal the archaeological standpoint of the differentiation in the study area with spatial considerations.

According to Mellart (1954; 187) there is an evidence of contact between the Beyşehir-Seydişehir Region and Cilicia, where the important point is given as the role of Taurus in the spread of painted pottery cultures form Cilicia. Not the Beyşehir-Seydişehir Region, but Beyşehir-Suğla Depression can give clues on transformation as well as the contact information between Hacılar and Çatal Höyük (Bordaz, 1973; 282), although the areas are the same, but named differently by different researchers. The selection of Beyşehir-Suğla Depression most probably was based on geological reasons rather than archaeological, and again according to Bordaz (1973; 283) there exist more than 60 sites, which are mostly later than Suberde. Those sites are most probably due to the activities between Hacılar and Çatal Höyük, while ancient people were passing the 'fertile basin area'. Moreover the area is suggested to be attractive to Paleolithic people, and it is also suggested that the area might be link between Antalya and Ankara (Solecki, 1964; 130), where no indication of an Upper Paleolithic exists.

5.2 Temporal Domain

For the specified study area, the main archaeological period of concern is the Neolithic, but the process of Neolithization is beyond the scope of this study. The Neolithic period is characterized by significantly increased human cognition over the environment. Domestication is the most evident result of it. Although there are many theories for domestication (Hodder, 1990; McCorriston and Hole, 1991; Thorpe, 1996; Richerson, Boyd, and Bettinger, 2001; Hayden, 2003; Veen, 2003; Hole, 2004) it is clear that there is a strong two-way influence between domestication and human cognition of environment. This period is also characterized by permanent settlements which do reflect a final, but critical decision about locating a settlement to a proper place with some criteria. Although these criteria are assumed to be energy minimizing by the beginning of the Holocene, cultural systems turned to be much more complex, but at the same time leaving much more material remains which enables mapping, testing, and predicting. To conclude, the specified temporal domain is the latest period that permits construction of the model, and the earliest period of which human-environment relation peaked but not in a destructionist way. The beginning of Bronze Age, by contrast, shows a significant increase in the size and area of the settlements indicating a level of urbanism (Baird, 2000; 15). Also for the eastern Mediterranean vegetation disturbance is clear by palynology after the emergence of complex societies; i.e. Bronze Age societies (Roberts, Meadows, Dodson, 2001; 634). This is also supported by the heavy use of the wheel and ploughing from the beginning of the 3rd millennium. This is accompanied with the use of ox-power with the increased level of technology, resulting in an extreme increase in the level of the complexity of the society (Allan, 1972; 224), and manipulation of environment in behalf of humans, but independent of environmental constraints.

This is not an isolated fact, but also observed in Jordan where it is indicated as heavy soil erosion, and environmental degradation between the Early Bronze Age and Classical times (Barker et al., 1998; 278) which might be due to the heavy manipulation of environment with the increased level of technology, and unlimited demand from social complexity been reached. Similar discussion is made by Wilkinson (2003; 20) where the correlation between the decline of species with the appearance Early Bronze Age mounds might have been caused by mainly human effect rather than environmental change. This simply means that although there was no detected significant change in the environment, ancient humans were able to alter it according to their needs and the environment itself was no longer the determinant over humans.

Although the Neolithic period of the region is the main concern, there is no way to define exactly the beginning and ending of this period, but each archaeological period is diffused to its temporal neighbors to some extent. Thus if there is no clear evidence of settlement pattern change, introduction of a new technological innovation, or a clear environmental event then the temporal limits of the study will be arbitrary rather than caused.

Mellart (1972; 279) added the Early Chalcolithic to his Anatolian Neolithic settlement patterns study because he thought that the economy of the Early Chalcolithic was still purely Neolithic, but he also suggests (1963; 199) a shift of settlement patterning after the end of Early Chalcolithic period. On the other hand Duru (1996; 55) states that in the Early Chalcolithic of Hacılar a deep change is observed in both the plan and the techniques of construction, concluding that a new group came to the area with new traditions.

According to the discussion given above the study focuses on mainly the Neolithic period while the Early Chalcolithic is also considered as well as the Epipaleolithic.

5.3 Drawing Exact Boundaries in Space

Although the results of discussion given above cannot produce an exact polygon limiting the area of concern, it helps to make a rough estimate both in space and time.

Spatially, limits of study area basically follow the ridges of drainage basins of lakes. On the other hand, since GIS analysis produces many watersheds from SRTM data in the broader context, the selection of basins was based on the archaeological excavations and surveys. Northern and western fringes were automatically selected with the watershed algorithm, but little manual correction was made with the help of 1:100000 topography maps as well as considering archaeological studies. Selection of southern borders was straightforward due to the existence of the Taurus Mountains, acting as a barrier. On the eastern side, it was hard to draw a proper and satisfying boundary due to various reasons (Figure 5.2).

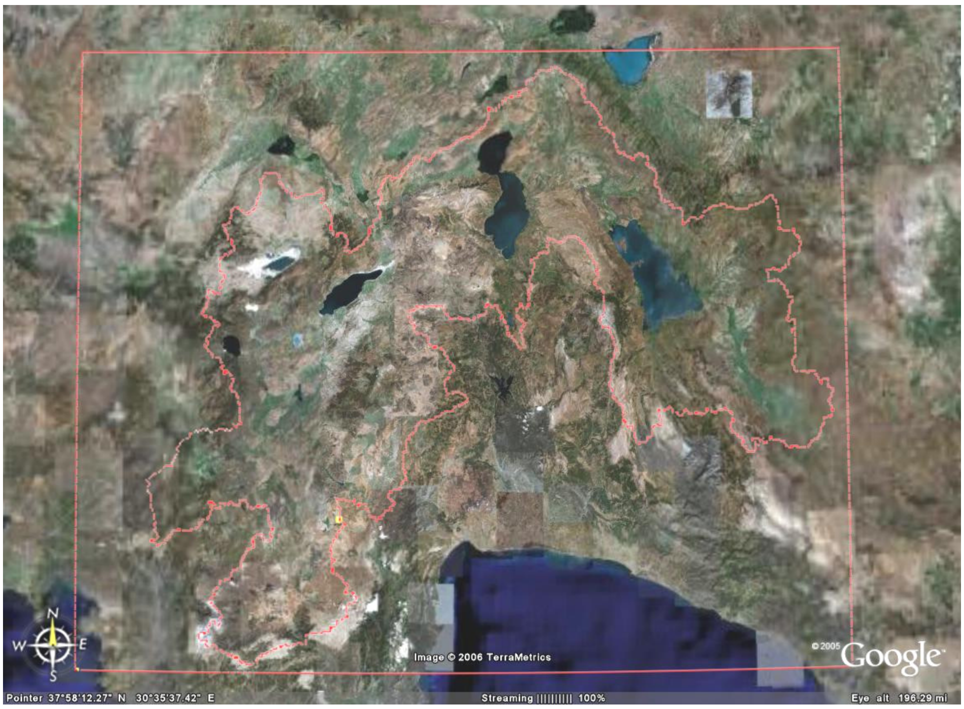


Figure 5.2 Exact boundaries of study area (inner polygon), and initial area of concern (outer rectangle)

A westernmost fringe of the Konya Plain, which is given as Beyşehir-Suğla area, is included into the study region. The area is claimed to be a geographical transition area to the Lake District since the Beyşehir Plain discharges into the Konya Plain (Kuzucuoğlu, 2001).

The Beyşehir-Suğla area is an also archaeological transition area although some basic similarities exist with core of Lake District Region. Then, although it will not be totally correct to add the sub-region to the study area, it is also not a great mistake while the area is not totally different from the rest.

The initial idea about the study area was also to add Konya Pluvial Lake. Although main pluvial lake and its neighboring wetlands do not exist today, they might have played a similar role as the lakes to the west of the south of plateau. Moreover, the validity of arid conditions of Konya Plain cannot be projected to earlier than fourth millennium BC (Erol; 1978). On the other hand it is claimed that Konya Pluvial Lake dried suddenly around 17000 BP, and never reached its original scale again (Fontugne et al., 1999; 579). Thus contemporaneity is highly questionable, and since lakes are the main focuses of the study, Konya Plain was withheld from the study area.

Apart from hydrological problems, area in question is also problematic in terms of alluvial deposition. It is stated that at Can Hasan III there is at least 1.5m of accumulation above the seventh millennium deposit (French, 1970; 139). If this is the case, then the sample from the area will not be representative, and future possible environmental constructions will not be liable.

The Lake District is promising in terms of obtaining a general reconstruction where reconstructions obtained from individual sites will reveal information about the immediate environment. Thus several samples from several lakes are needed for a regional reconstruction (Eastwood, Roberts, and Lamb, 1998; 69). Moreover, climatic changes in Anatolia are not easily followed since there is not enough material to be dated by C14 and there is no continuous sequence in shallow lakes. On the other hand, some lakes (Van or Burdur) permit such procedures (Fontugne et al., 1999; 586).

CHAPTER 6

DATA

In the analysis, logistic regression is employed to construct a predictive model. Data is comprised of a dependent variable, or location of sites, and independent variables with various sources and characteristics. This chapter is devoted to an introduction of data and their sources as well as to reveal some basic statistics for a better understanding.

Study area is divided into 8 sub-zones based on watersheds. Determination of sub-zone boundaries were automated, but some minor manual corrections were made as in determining the study area.

6.1 Dependent Variable

Locations of sites are all obtained from the TAY database (Harmankaya, Tanındı, Özbaşaran, 1997), in which locations were supplied with exact geographical locations in the Lat-Lon coordinate system. On the other hand, accuracy is a matter since data does not contain 'seconds' part of that projection. The archaeological inventory also contains physical dimensions for most of the sites as well as some orientations. It is discussed that the determination of site and study boundaries are of immense importance and it is known that current measurements hardly reflect ancient settlements. The case is more problematic for multi-period sites.

Minimum and maximum values from the set of site areas are removed from the inventory in order to get rid of the extreme values. Then, average site size is obtained by creating 133m x 133m pixels, where the necessary resampling is applied both to satellite imagery and soil and lithology data. Then total number of 64 sites ended up with 95 pixels. The number of sites is very few in consideration of the immense study area. Previous predictive modeling studies focuses on very small areas with large sample sizes. Then it is a must to obtain a large scale site database as well as making an effort to merge the existing knowledge in a useful manner. Randomly selected 85 pixels are used as training sample to construct the model, whereas 10 of them are left for testing the model.

An initial look at the distribution of sites reveals that there was a focus on lakes and streams and in flat areas. In contrast, there is an irregular distribution of sites when sub-zones are examined (Figure 6.1).

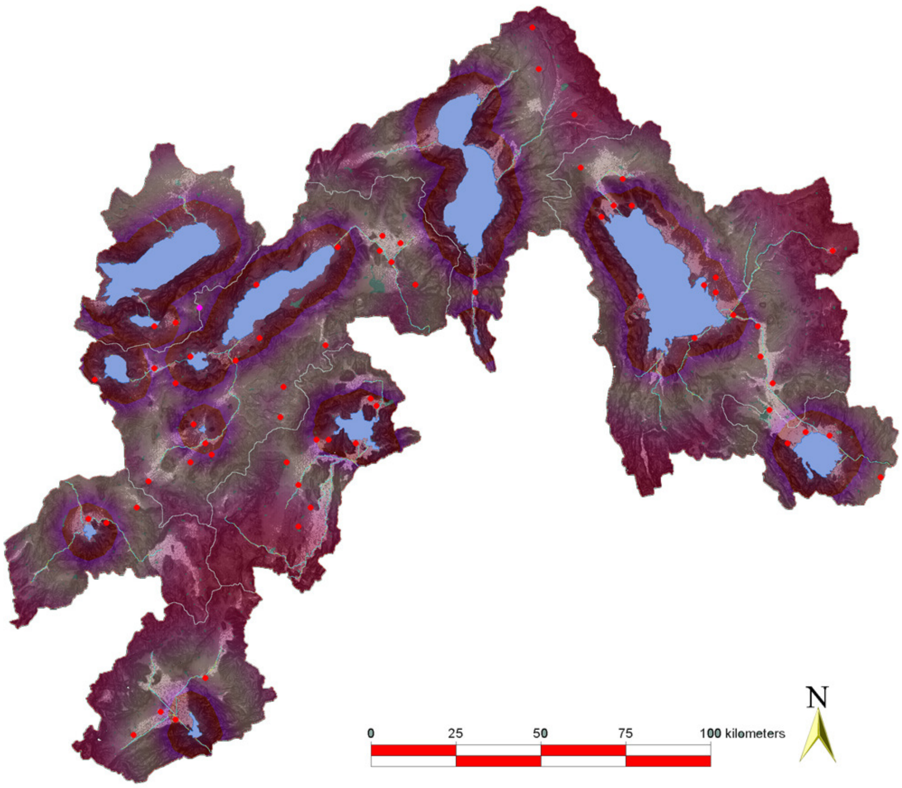


Figure 6.1 Slope map overlay on Distance to Lake Map. Site distributions are also put in order to show the frequency of matching of the sub-areas concerned.

6.2 Independent Variables

Independent variables of the model come from various sources. Some of them were rasterised before being put into the analysis such as lithology and soil data, while some were in raster format already, such as topographical data. In any case, all of the 'end' raster were resampled to a 133m pixel resolution. Since they were produced by different institutes in different time periods they have different scales, though the problems of integrating different scales were discussed in previous pages.

Major sets of the model are topographic data, rock data, soil data, and hydrological data. Each data set is processed in order to be used in the model, where some are used as they are, such as rock data, and some are used to produce some minor variables such as 'Adjusted Distance to Basin Ridges'.

Topographical data is obtained from SRTM (Shuttle Radar Topography Mission) satellite imagery. Data have a 90m pixel resolution and a16m vertical accuracy (Japan Aerospace Exploration Agency, 2006) (Figure 6.2). There are some other ways to obtain general topographic characteristics of any area, such as GPS mapping, creating (or digitizing) contours and various others. The preference for SRTM data comes from various aspects. First, and maybe the most important, is its public accessibility. One of the mainstreams of any research is cost, and such 'free' data source is a gateway for researchers. Second, by using SRTM data, immense time can be saved. Although some automated or semi-automated contour identification is available, accuracy is questionable and manual correction is always needed. Although 90m pixel resolution might not be totally desirable for a study with particular research questions, for this case, when the dimensions of study area are considered this resolution is more than enough.

According to the report published by (Rodriguez at al., 2005) absolute error for study area is between 5 and 10m around ridges of the Taurus Mountains, and less than 5m for core of study area.

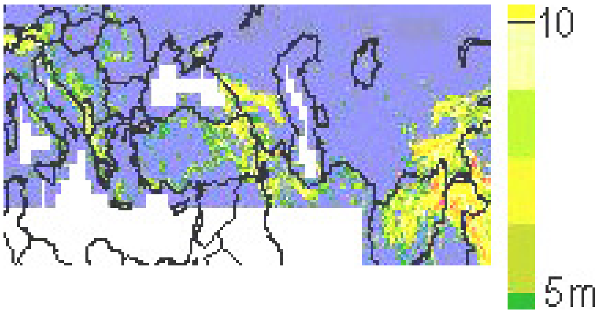


Figure 6.2 Estimated 90% vertical absolute error

Rock data were already in digitized form, where it comprises of four sheets; Ankara, Konya, İzmir and Denizli. The main source of data is MTA (General Directorate of Mineral Research and Exploration). Data is available at 1:500,000 scale.

Soil data is available in vector format and can be obtained from General Directorate of Rural Services. Data is accompanied by an attribute table so that some variables can be queried and used in a model. Although it needs extensive extra work, final information is valuable.

Hydrological data (rivers and lakes) is extracted from topographical data and checked against existing 1:100:000 topographical maps. The dynamic character of this data layer required some special attention because of altering lake levels and changing river courses both due to natural and artificial processes.

6.2.1 Topographical Data

Topographical data is composed of some basic elements as well as other components which are extracted by processing in GIS. Basic elements are Elevation, Slope and Aspect, whereas others are Adjusted Elevation, Degree of Surface Roughness, Distance to Roughness Joints, and Distance to Basin Ridges.

Basic elements are produced with a straightforward analysis in TNTMips 6.9 platform. Thus, any error introduced into data is due to the algorithm, but minor topographical variables are already produced, but with some assumptions so that use of them can be criticized.

6.2.1.1 Elevation

Elevation is in fact is source of each variable used as topographical component, and imported into GIS from 'hgt' data format (Figure 6.3). Thus any problematic part in data is inherited to other variables, such as voids and vertical inaccuracies in SRTM data collection.

Although elevation can be a direct determinant, it can be also considered as a proxy variable since it is a measure of snowline or it determines the faunal and floral abundance or scarcity.

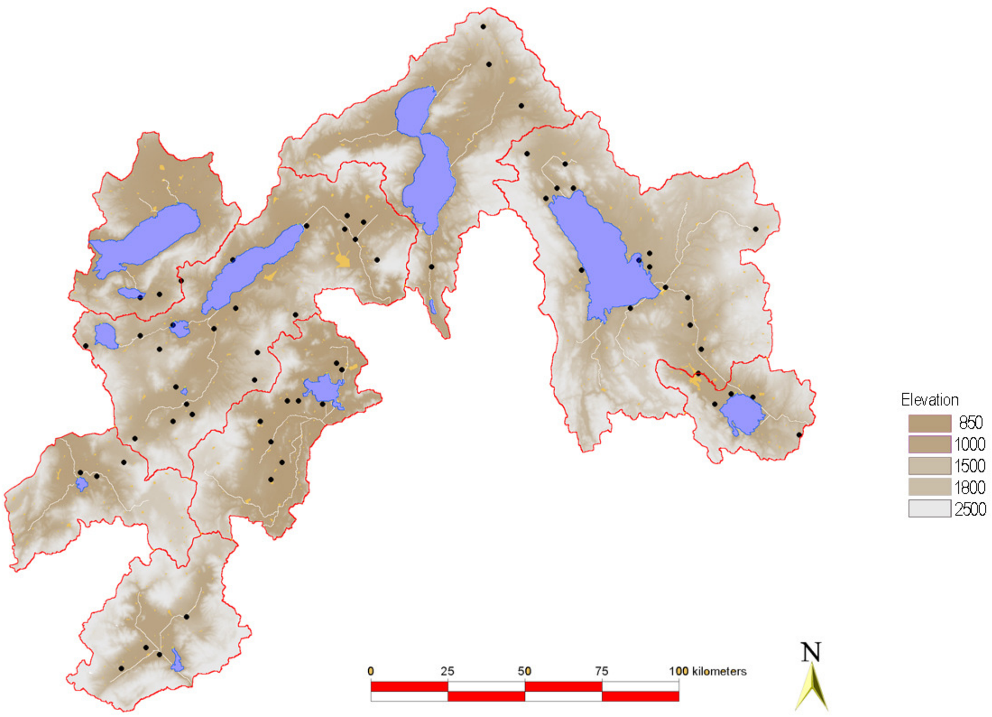


Figure 6.3 Elevation map of study area**.**

The study area has a mean elevation of 1590 meters. The value around 1500m dominates data, forming 60%. The rest is distributed somewhat equally at either side (Figure 6.4). Elevation is generally lower around lakes and it increases as one gets further from lakes. The effect of the Taurus Mountains on topography is visible at the South side of study area.

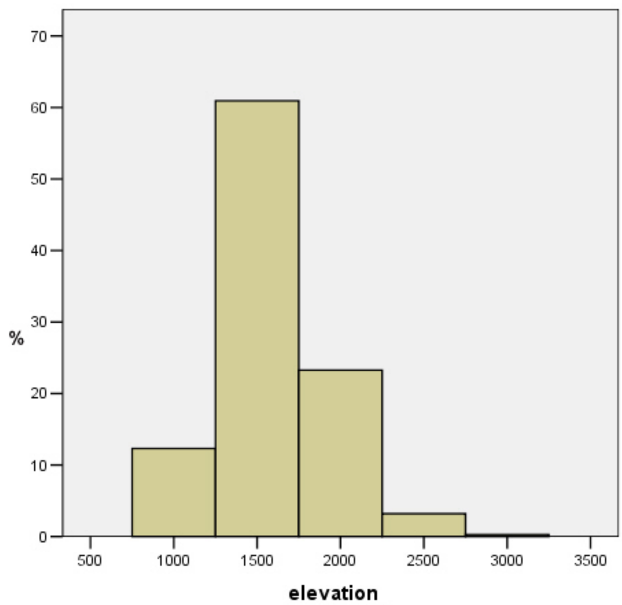


Figure 6.4 Histogram of study area elevation values.

The average elevation of pixels having site values is around 1360 meters, which is 250m lower than average elevation of whole study area (Figure 6.5). On the other hand, site pixels are also characterized by values around 1500m being more than 70%, which decreases the significance of this data layer, but it should be also said that the maximum site pixel of the inventory is less than 1500m. Thus although there is a tendency for site pixel values to follow general topography, elevation higher than 1500m is avoided.

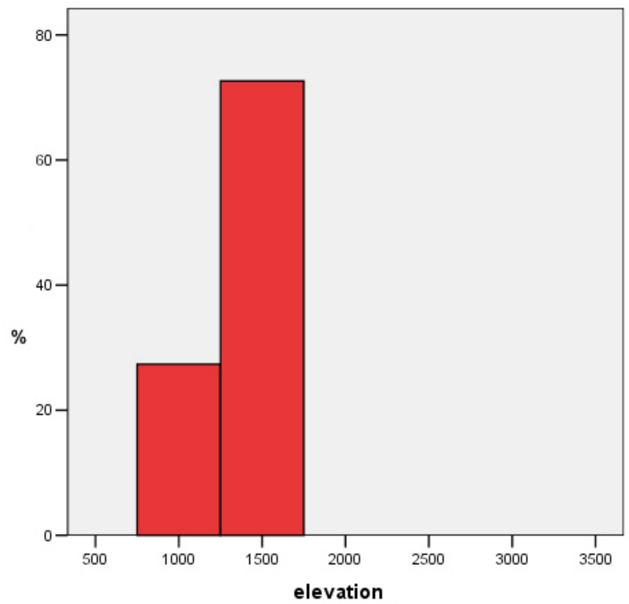


Figure 6.5 Histogram of site pixel elevation values.

6.2.1.2 Adjusted Elevation

Adjusted Elevation is constructed by subtracting base lake levels of each basin from each pixel in that basin. If an ancient lake level is known then it is used instead of the current level. The aim is to eliminate relative height differences in each basin and to consider sites relative positions to their focal lake levels (Figure 6.6)

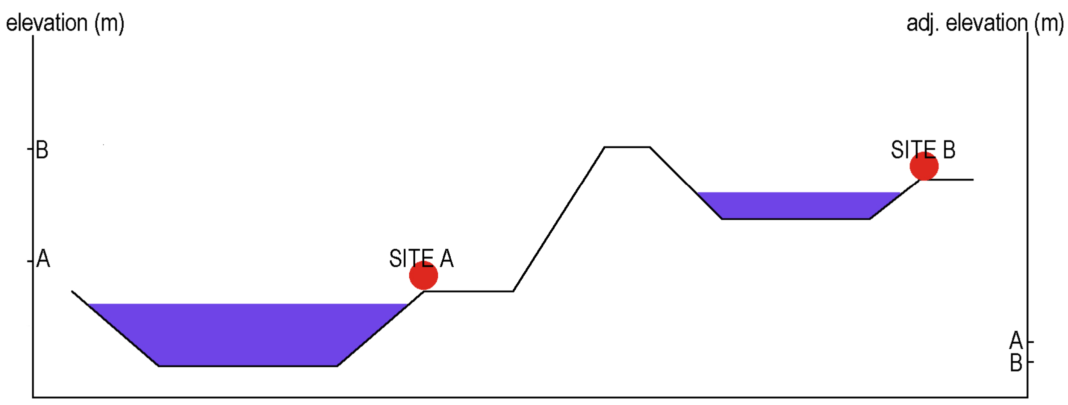


Figure 6.6 Hypothetical relative and absolute site elevations according to lake levels.

Definitely, reconstruction of ancient lake levels should be approached critically. Moreover, not all of the ancient lake levels are available. Thus, there is no such final decision on lake levels, but in any case operation was subtraction since it is known that lake levels were higher before, where by this operation some amount of bias is introduced to the model. After subtracting lake levels from each basin, an adjusted elevation map of the study area is obtained (Figure 6.7)

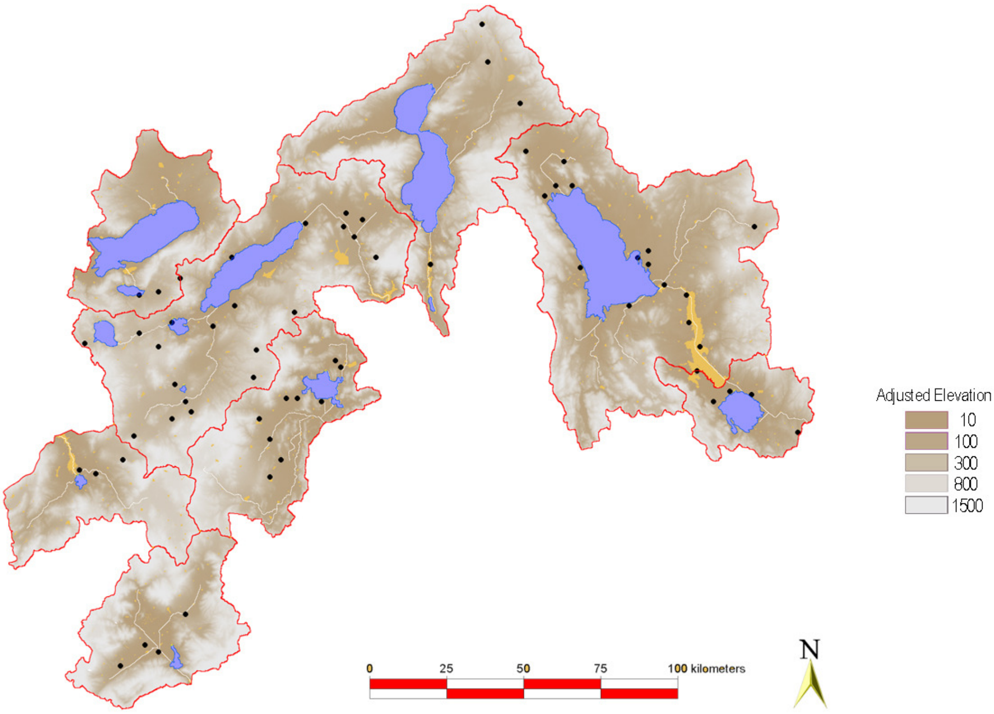


Figure 6.7 Adjusted elevation map of study area (in meters)

After making adjustments in elevation values, the overall study area elevation drops to 670 meters (Figure 6.8), where each basin has its own average elevation. On the other hand, it is helpful to examine to what extent elevation drops by such an adjustment. Adjustment also changes the distribution of data to a skewed one.

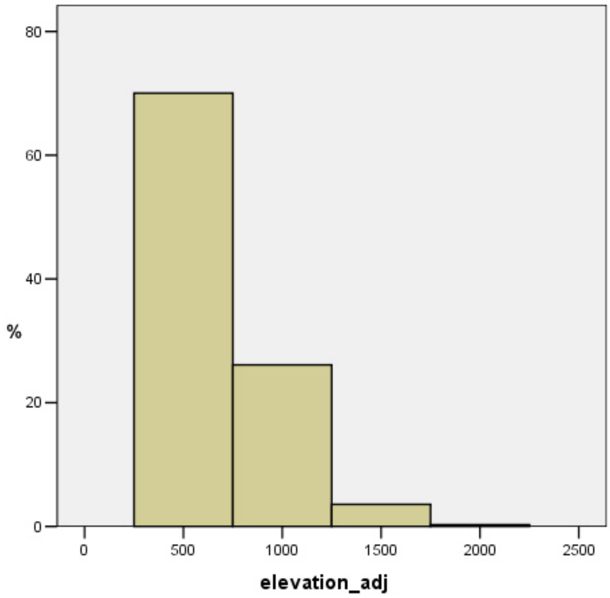


Figure 6.8 Histogram for Adjusted Elevation values of study area

Adjustment applied to pixel site values is significant, since this operation gathers those values around 500m with 95%. Then it can be stated that 500m is the critical limit as a distance to lake level in the vertical axis (Figure 6.9). This situation makes this layer a possible contributor to the model.

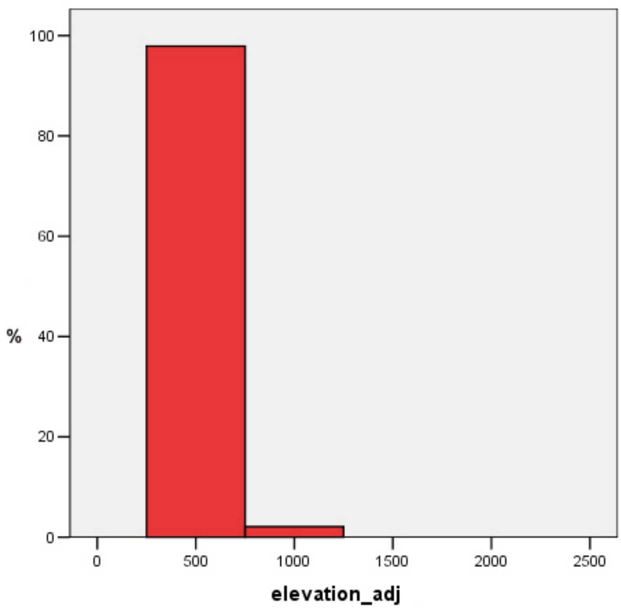


Figure 6.9 Histogram of site pixel adjusted elevation values.

6.2.1.3 Slope

Slope is calculated in degrees from the horizontal axis. Hypothetically, sites are expected to be located in flat or relatively flat areas. On the other hand, since the study area is geographically in contact with flat lands (Konya Plain), and since the area is characterized by recession of lakes which produce smooth areas, this site-slope relation (if exists) can be artificial due to the abundance of *flat* pixels.

Also, as started the immediate vicinity of the lakes is relatively flat, and might have been used accordingly by the ancient settlers (Figure 6.10). The basin ridges are expected to have higher degree of slopes.

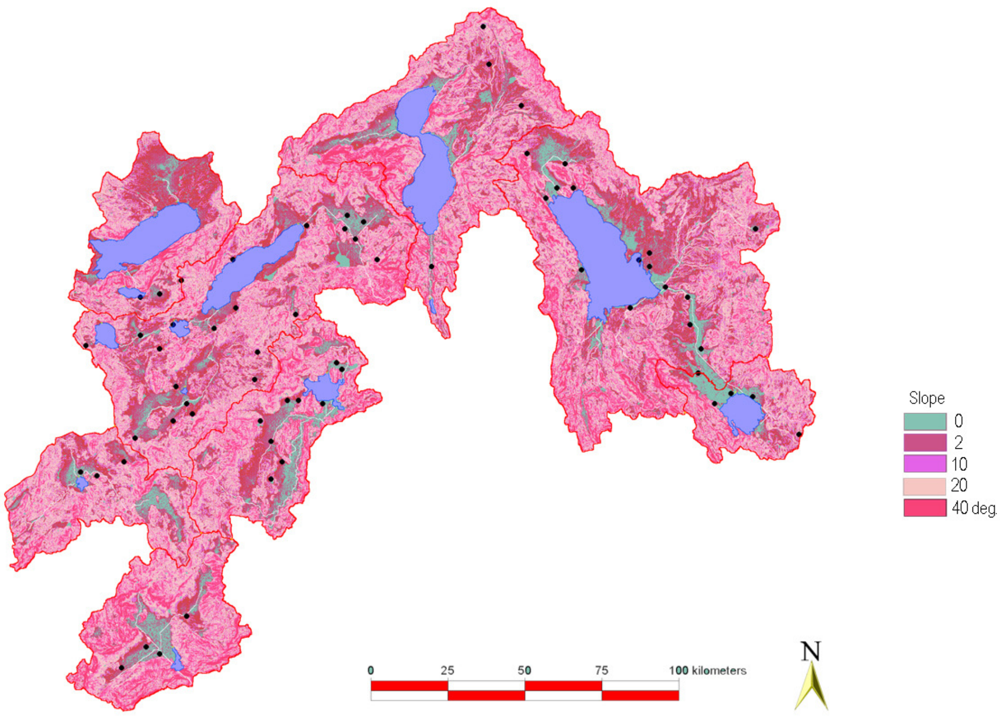


Figure 6.10 Slope map of study area in degrees**.**

The study area can be claimed to be relatively flat when its average slope value 10 is considered, and emphasis is on small values of slope with a skewed distribution (Figure 6.11). Since this is the case, there is a categorical look of data rather than a continuous one, so no reclassification is performed. Moreover, such reclassifications are biased in a way that it should be avoided if possible. There is no exact threshold slope value that decides whether a pixel is flat or not. In other words, it would be highly questionable to reclassify 2 degrees of slope as 'flat' areas and 3 degrees of slope as 'not flat' areas.

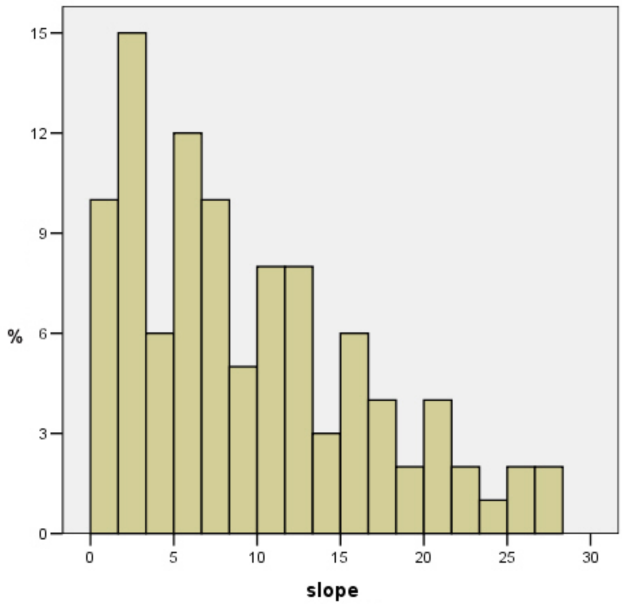


Figure 6.11 Histogram of slope values of pixels for study area.

Site pixel values are distributed unevenly. Nearly 70% of data is less than 3 degrees (Figure 6.12). Then as a general observation and with a logical guess, it can be stated that a tendency exists for 'flat' areas. Then again, it should be re-emphasized again that humans are very ambitious and successful in changing their immediate environment so that any possible relatively flat area can be used to locate a settlement. Then it is a question of deciding what is flat, what is potentially flat, and what is not flat.

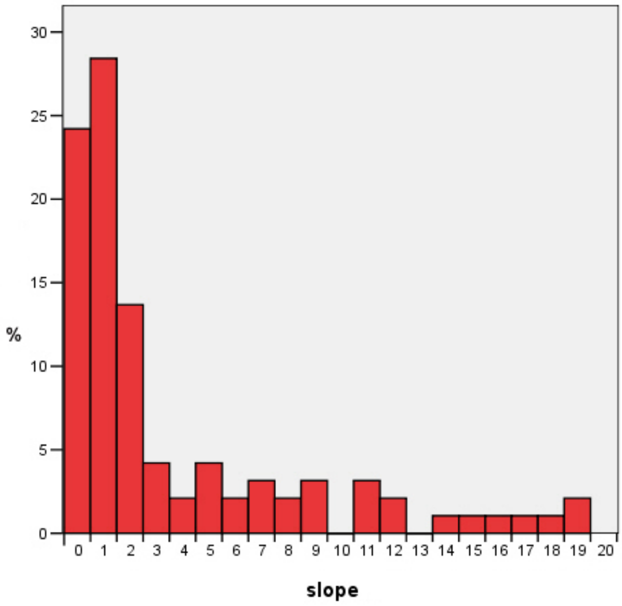


Figure 6.12 Site pixel histogram.

6.2.1.4 Aspect

Aspect data is reclassified according to the principle axis. This approach aims to diminish the complexity of situations as well as to produce a better interpretable result. Aspect values are classified into 9 classes where 8 of them contain directional information, and the last one comprises pixels which do not have orientations. The categorization is summarized in Table 6.1, and representing Aspect map is given in Figure 6.13.

Table 6.1 Reclassification of Aspect values

DIRECTION	INTERVAL of DEGREES
FLAT	Slope=0
NORTH	337.5-22.5
NORTHEAST	22.5-67.5
EAST	67.5-112.5
SOUTHEAST	112.5-157.5
SOUTH	157.5-202.5
SOUTHWEST	202.5-247.5
WEST	247.5-292.5
NORTHWEST	292.5-337.5

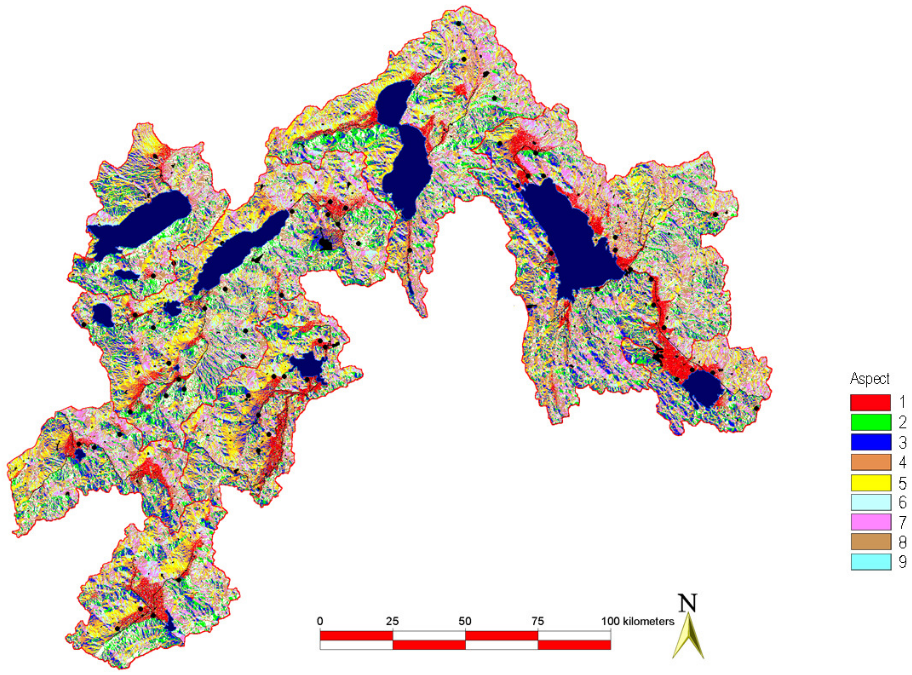


Figure 6.13 Aspect map of study area after reclassification.

Each principal axis contributes equally to the model. It is helpful in a way that any minor significance in response data can contribute to the model. Choice for 'flat areas' affected this situation, and resulted in such distribution of aspect values (Figure 6.14).

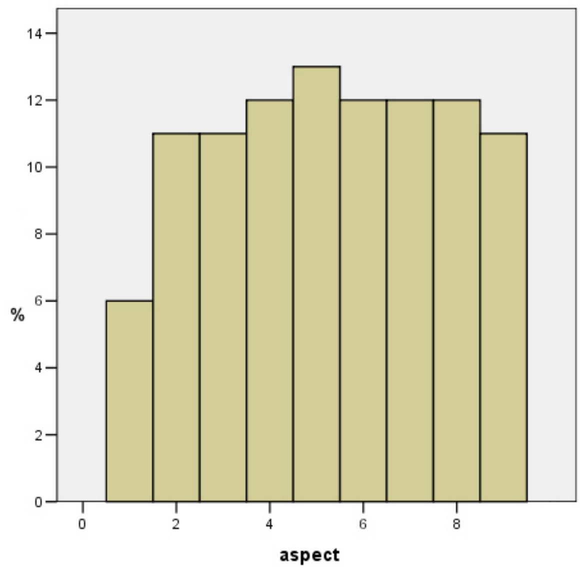


Figure 6.14 Study area Aspect values in percentages.

Aspect values of site pixels (Figure 6.15) differ from study area aspect values, most significantly for South values. This is much more visible if ‘flat pixels’ with coding ‘1’ are omitted. This is, in fact, an unexpected situation when previous researches are considered.

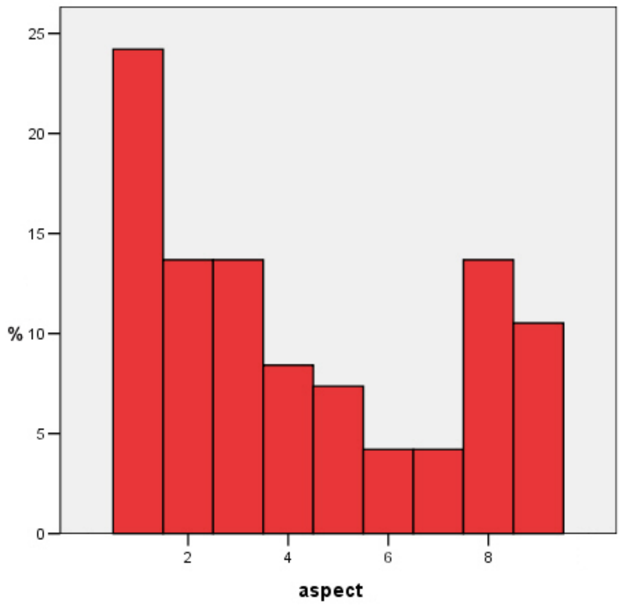


Figure 6.15 Aspect values of site pixels in percentages.

6.2.1.5 Adjusted Distance to Basin Ridges

Distance to Basin Ridges can be an important determinant. It can be stated that although each basin has similar general characteristics, they also have specified differences. This is both due to the different water qualities of lakes and different average elevations. While Distance to Basin Ridges is an important variable, it cannot be used directly. Each basin has different amount of land cover. Thus there is no standardized distance between any point in any basin and ridge line of that particular basin. In other words, even if a passenger goes the same amount of distance in different basins, he or she ends up with different niches since relative location to any particular distance, thus an adjustment is needed.

Assuming the areas of basins are the same, shapes of basins produce another problem, and thus standardization is needed. In order to do this, a shape parameter, *compaction*, is used. Simply, a compaction is a ratio between the area of any shape and the area of a circle. By doing this each basin is standardized according to their perfect circles. To obtain an overall adjustment, first Distance to Basin Ridges is found for each pixel and they are multiplied with reciprocals of area ratios and compaction values. The result of each step is given in Table 6.2, and after multiplication of Distance to Basin Ridges with Scale values Figure 6.16 is obtained.

Table 6.2 Compactness Values and Area Ratios for Each Basin

BASINS	BASIN BOUN.	R	COMPACTNESS	1/COMPACTNESS
Suğla	2,E+13	4,E+12	0,098	10,255
Acıgol	3,E+13	5,E+12	0,172	5,825
Avlan	3,E+13	5,E+12	0,189	5,302
Beysehir	6,E+13	9,E+12	0,521	1,919
Burdur	5,E+13	9,E+12	0,434	2,303
Gölhisar	3,E+13	5,E+12	0,158	6,311
Eğirdir	5,E+13	8,E+12	0,326	3,067
Kestel	4,E+13	6,E+12	0,204	4,910

BASINS	AREA RATIO	1/AREA RATIO
Suğla	0,046	21,827
Acıgol	0,083	12,062
Avlan	0,086	11,679
Beysehir	0,250	4,000
Burdur	0,207	4,825
Gölhisar	0,073	13,622
Eğirdir	0,160	6,263
Kestel	0,095	10,488

BASINS	1/AREA RATIO	1/COMPACTNESS	SCALE
Suğla	21,827	10,255	223,83913
Acıgol	12,062	5,825	70,26080
Avlan	11,679	5,302	61,92471
Beysehir	4,000	1,919	7,67810
Burdur	4,825	2,303	11,11067
Gölhisar	13,622	6,311	85,97074
Eğirdir	6,263	3,067	19,20620
Kestel	10,488	4,910	51,49373

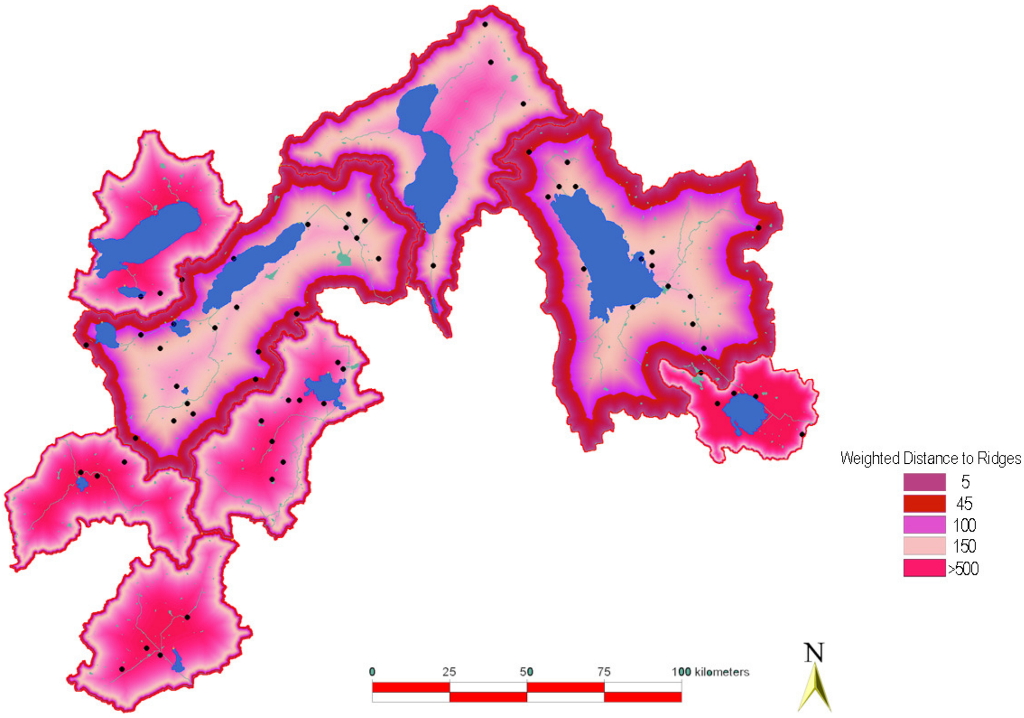


Figure 6.16 Map of Weighted Distance to Basin Ridges in unit distance.

Average adjusted Distance to Basin Ridges in the study area is 40 units. Distribution is highly skewed favoring a mean value (Figure 6.17).

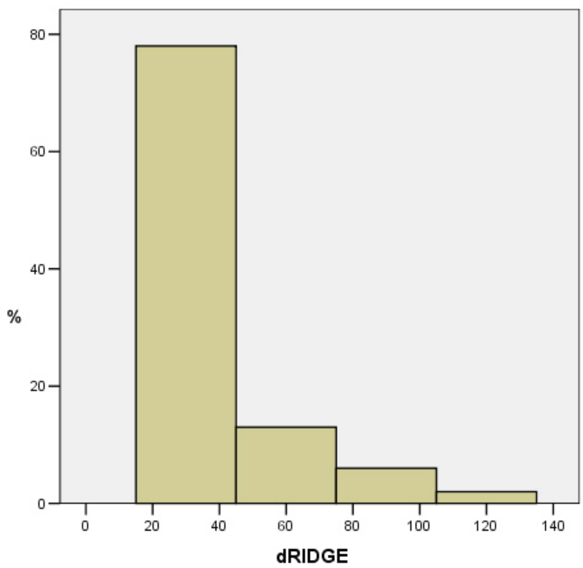


Figure 6.17 Histogram of Distance to Basin Ridges for study area.

The average distance for site pixel values is not very different from the overall value (Figure 6.18). It has a unit distance of 50. When exaggerating the effect of the scale is considered, this difference is larger than observed.

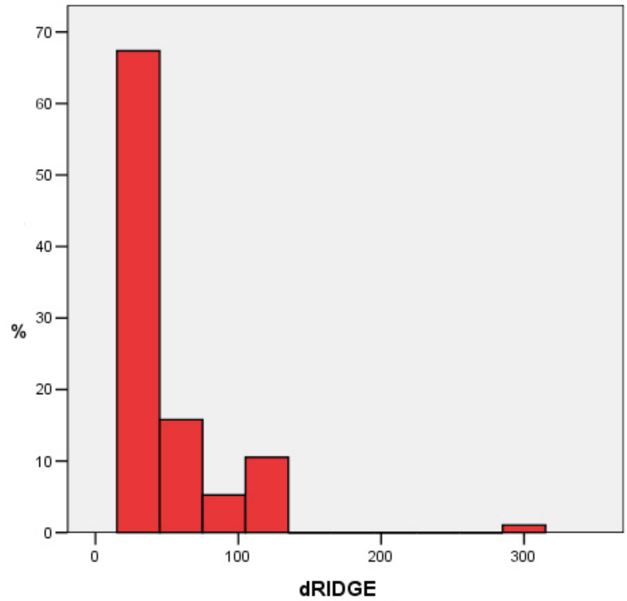


Figure 6.18 Histogram for site pixel values for Distance to Basin Ridges

6.2.1.6 Surface Roughness

Different topographical characteristics lead to different land uses and different perceptions of landscape. Thus, in this study topographical variety is represented by Surface Roughness.

The variable aims to measure complexity of terrain. The amount of change of elevation in any unit area can be used to create such a variable. This approach can be formulized by taking a second derivative of the elevation map, or slope of the slope map, which is easy to produce with GIS tools. In this study, a new approach will be followed.

If topography is represented as a mathematical function, then observation frequencies of local and global minimums and maximums as well as saddle points are indications of complexity, or roughness. To obtain a measure, the densities of those points are produced and classified and used as Surface Roughness (Figure 6.19).

On the other hand, such a measure is not totally true, and there should be a difference between smooth undulations and big topographical changes, where both happenstances are represented by similar configurations of points in space. For this study, no adjustment is made, and it is assumed that there are no such coincidences of happenstances.

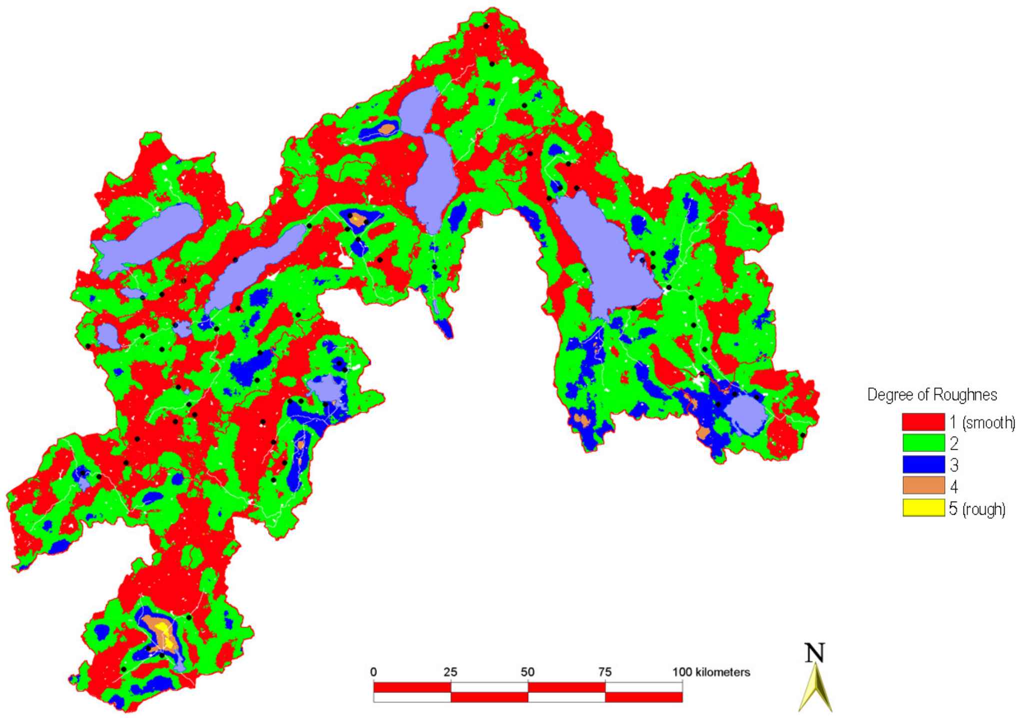


Figure 6.19 Surface Roughness degrees for study area.

Total study area can be claimed as homogenous in relief since first and second smooth categories constitute 90% of all data (Figure 6.20). This can be explained with the abundance of lakes, thus lake shores. The situation might be different when lake levels were higher.

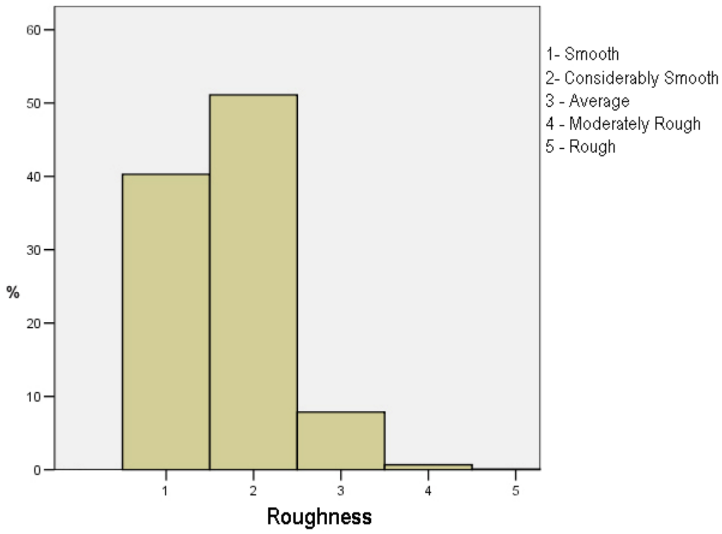


Figure 6.20 Surface Roughness values of overall study area.

The character in distribution of Surface Roughness degrees for sites is very similar to the overall pixels' case (Figure 6.21). Sites close to lakes are contributing to this event so that frequency of sites located on smooth terrain is increased.

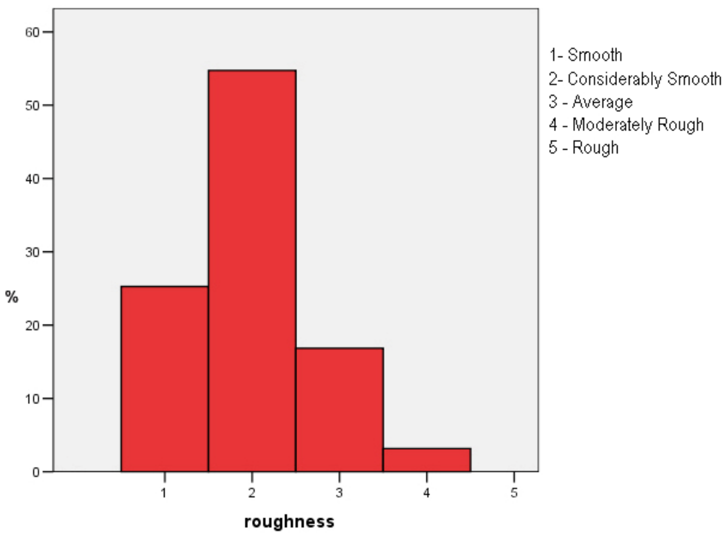


Figure 6.21 Site pixels Surface Roughness values.

6.2.1.7 Distance to Surface Roughness Junctions

Different landscape characteristics result in different flora and fauna. For instance, goats prefer to live in undulated topography whereas sheep prefer flat areas. If it is assumed that human aims so maximize its exploitation, then it is logical to look for areas of junctions where in that case, a farmer or a hunter can exploit different niches.

Then Distance to Surface Roughness Junctions (Figure 6.22) can reveal environmentally rich areas. In this case, a pixel close to two or more surface roughness degrees should be preferable, whereas a homogeneous environment is avoided.

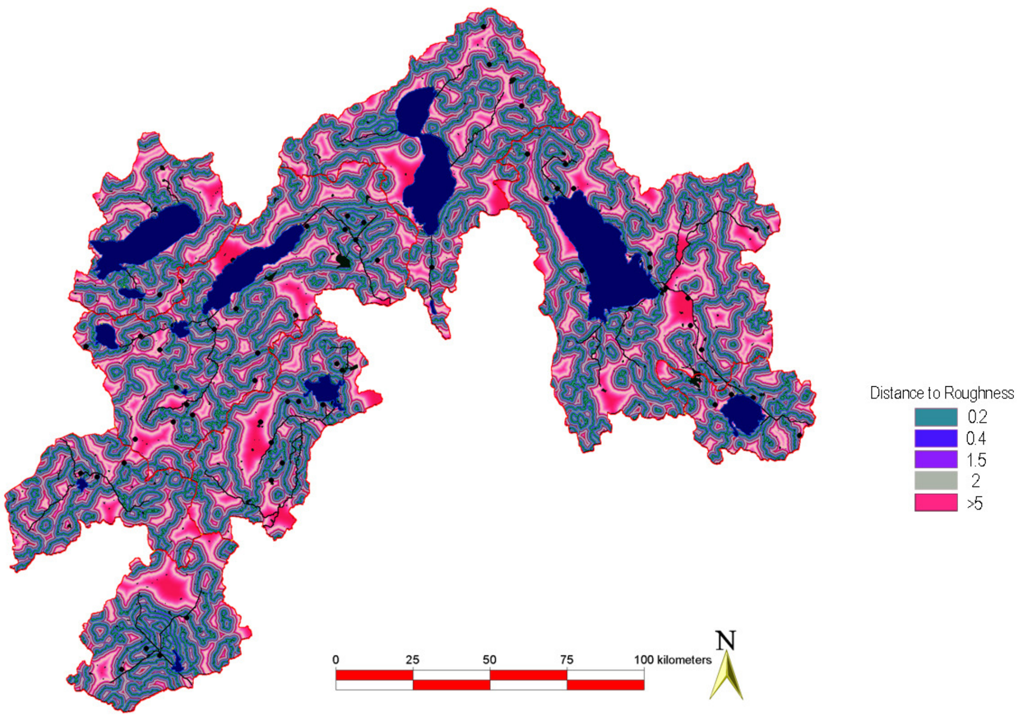


Figure 6.22 Distance to Roughness Junctions in kilometers.

Any pixel in the study area is close to a junction of different topographies with an average of 2 kilometers (Figure 6.23). Other question to be asked then, for this variable, if there is a difference in the same amount of distance in different topographies. In other words, to what extent it is feasible to distinguish 2 km in a mountainous area and in a flood plain. Although there is no clear answer for it, such kind of superficial variables can shed light on the phenomenon.

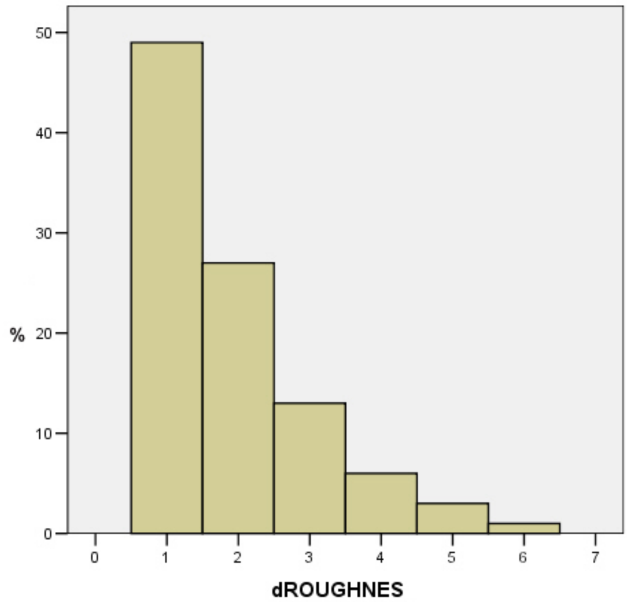


Figure 6.23 Histogram of Distance to Roughness Junctions values for study area.

Average distance to roughness junction is 2 km. There is no substantial diminishing of distance, but after 2km it is suddenly cut off. One pixel significantly differs from the general with 10 km (Figure 6.24).

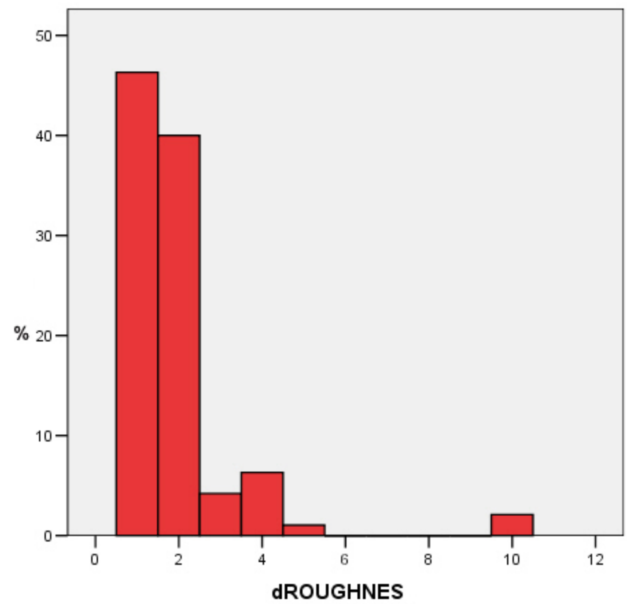


Figure 6.24 Site pixel distance histogram.

6.2.2 Rock Data

Rock data is maybe the most stable agent among all of other variables, where topography is manipulated by people, soil is heavily affected by deforestation and other natural processes and water is continuously active. Thus this data type turns out to be less biased than others, but only one layer of data is used, and no other minor variables are produced.

6.2.2.1 Lithology

The attribute table of this layer contains many variables so that it is reclassified for a better regression model. This categorization is geological rather than archaeological. Existing inventory for rock data includes alluvial fan, andesite, basalt, carbonates and clastic rocks, clastic rocks, dunite, gabro, marble, neritic limestone, melange and various others. Reclassified data is given in a simplified Table 6.3. Due to subjective classification and deep geological emphasis, there is little expectation for a contribution the model, but definitely, this data layer must have played an important role, at least in an indirect way (Figure 6.25).

Table 6.3 Reclassification table for rock data. Left hand side denotes super classes and right hand side contains original attribute data.

Quaternary	Undifferentiated Quaternary
Volcanics	Andesite, Trachite, Pyroclastics
Melange	Peridotite, Ophiolithic Melange, *Schist*
Neritic	Neritic Limestone,
Clastic	Clastic Rocks, Continental Clastic Rocks,
Clastic and Carbonates	Clastic and Carbonate Rocks,

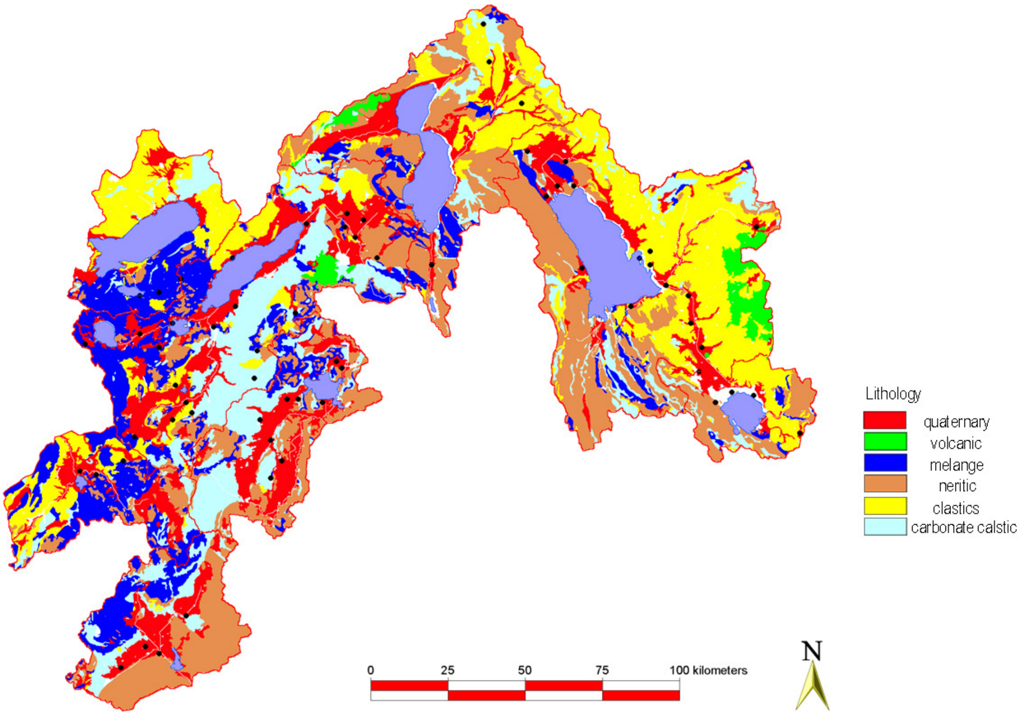


Figure 6.25 Rock data distribution of the study area.

Rock types are distributed evenly except the volcanic class, which is not surprising. Thus any patterning in site data will contribute directly to the model (Figure 6.26).

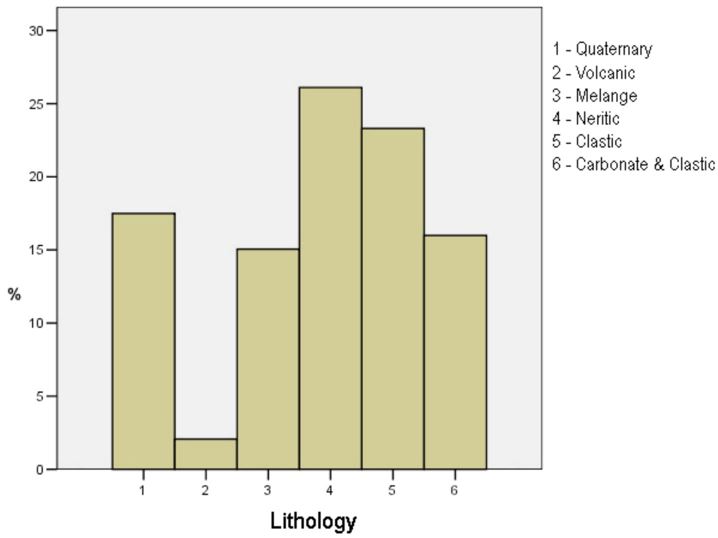


Figure 6.26 Percentages of rock types in study area.

The Quaternary class significantly differs from remaining classes. No site pixel is observed with volcanic class. Then it is the discussion of finding out the reasons for the correlation between quaternary geology and archaeology, or the effect of Early Holocene landscape and settlement strategies.

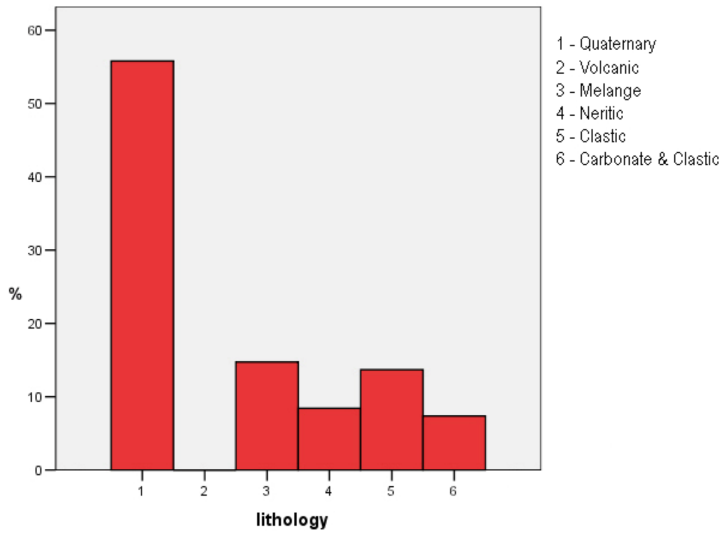


Figure 6.27 Percentages of rock types for site pixels

6.2.3 Soil Data

Soil data is composed of many sub datasets. Although not all of them are used in constructing the model, they still produce invaluable information on the study area. As stated, this data represents the current layout of soil, not that of the time in concern. Thus this datasets should be approached critically.

6.2.3.1 Land Use Potential

Land Use Potential is a degree of value of soil, assigned by the General Directorate of Rural Services. Data can be used as a decision making tool for modern investments, but as discussed earlier on; to what extent such degrees reflect the past is highly questionable. An interesting issue about this set is the question of to what extent ancient humans were aware of the soil quality without employing chemical techniques. Some observations should have been made on soil. Definitely, the difference between clay soil and silty soil was understood by ancient people in their qualities and compositions but it must be the technology that has changed the manipulation.

Existing data does not contain Land Use Potential information for every part of the study area, as it can be seen in Figure 6.28. Thus it is not used in the analysis.

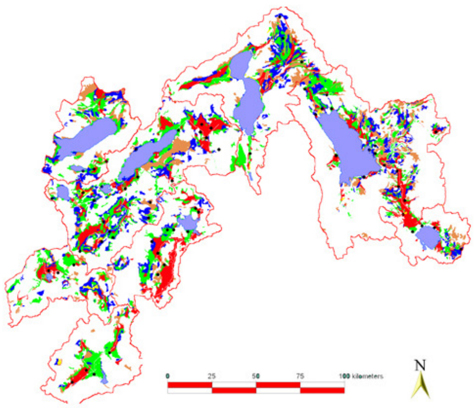


Figure 6.28 Land Use Potential Map of the study area.

6.2.3.2 Current Land Use

As the name of the layer implies, data reflects 'current' land use rather than the ancient land use. If there was a choice among ancient people, then it should be reflected into the data, at least indirectly.

As seen in Figure 6.29, there are also some void values (white areas) in north part of study area, but not as wide as 'Land Use Potential'. The situation is overcome by assigning random values to null pixels. Moreover, after an initial look it can be stated that most of the null values belong to agricultural fields and shrubbery and forests, and with small probability of grasslands.

The study area is dominated by shrubbery and forests. Second, agricultural fields comprise an important proportion. Immediate observation on distribution of sites reflects a potential of selection of (modern) agricultural fields, which is also evident in the histogram. On the other hand, it is the logistic regression which will decide on the importance.

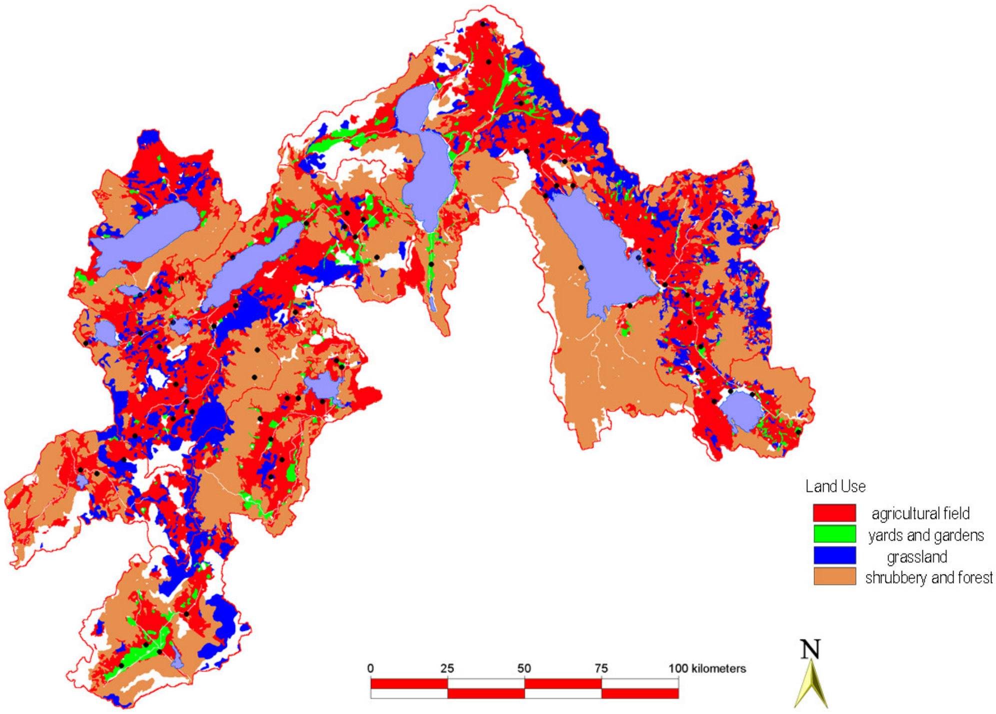


Figure 6.29 Land use classes of the study area.

The study area is dominated by agricultural fields and forests and shrubbery. The remaining 20% is shared between grasslands and gardens (Figure 6.30). This kind of distribution is expected when the effect of the Konya Plain and Taurus Mountains is considered.

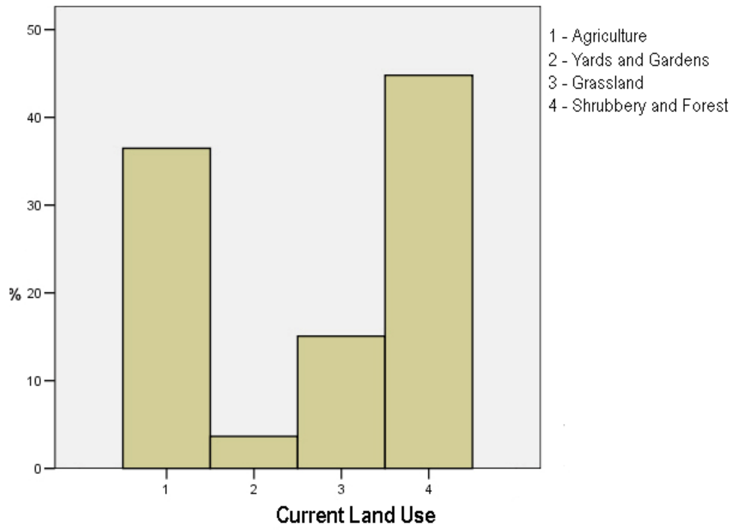


Figure 6.30 Land use categories of the study area, and their overall percentages.

Site pixel distribution resembles the distribution above except for the forests and shrubbery category. Modern agricultural fields are the best choice for the concerned ancient settlements. Remaining 30% categories are shared equally by other sites (Figure 6.31).

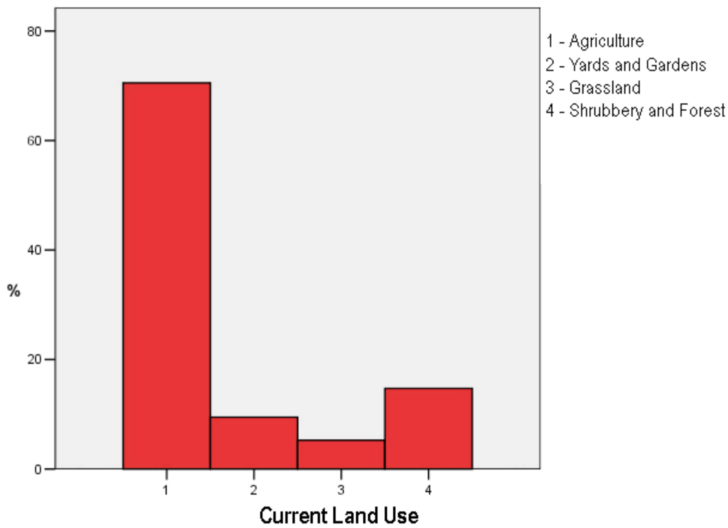


Figure 6.31 Site pixel land use classes.

6.2.3.3 Soil Inclusion Size

Soil Inclusion Size can be a proxy for the potential land use. A finer (thinner) soil might be a choice for ancient people depending on the level of technology, whereas coarse soil was avoided. If this is the case, it should be reflected in one of variables, and Soil Inclusion Size is a good candidate for it (Figure 6.32).

Also, the layer can be used in combination with 'Current Land Use', to create another data set which can be used as a substitute for 'Land Use Potential'. On the other hand, the rules of combination are not clear, thus it is avoided in the study.

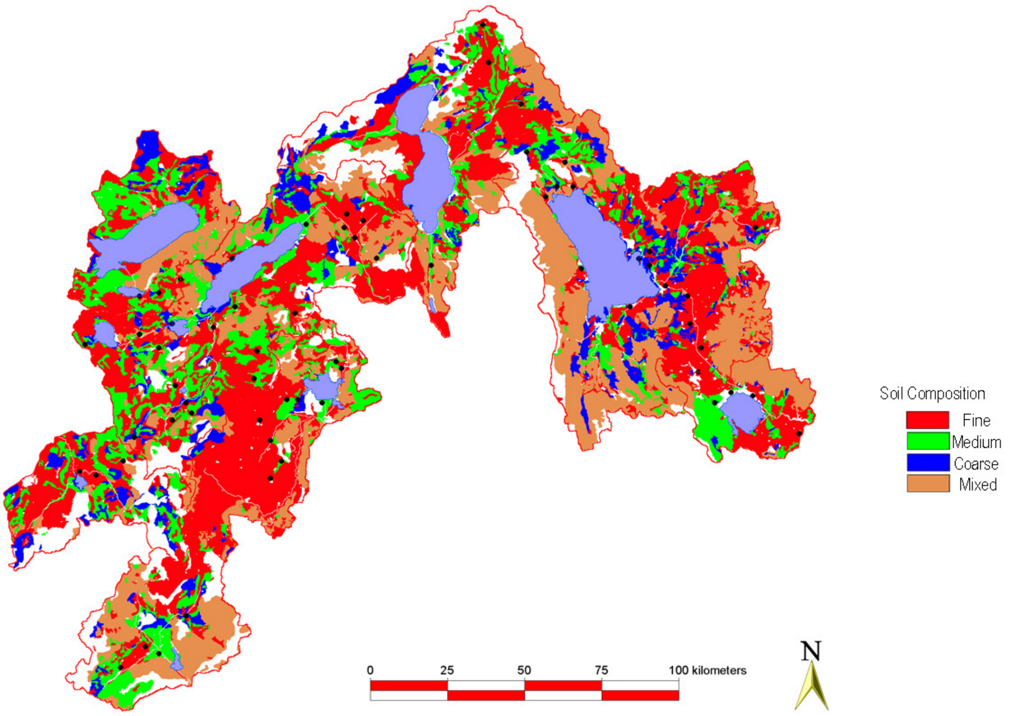


Figure 6.32 Soil Inclusion Texture Distributions

Fine and mixed soil (1,4) dominates the data set (Figure 6.33). Fine soils might be due to the recessions of lakes.

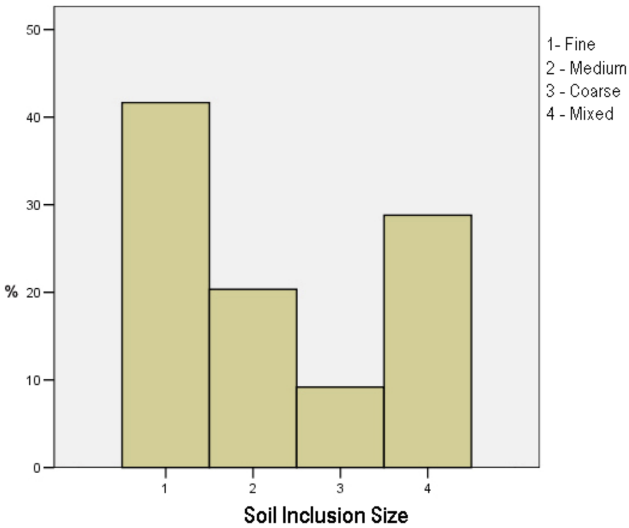


Figure 6.33 Inclusion classes of study area.

60% of site pixels represent fine soil type. Then medium coarse soil follows the leading category at 25% (Figure 6.34). There is no visual significant difference in the domain and site pixel distributions, so that any comment on determinants of patterning will be speculative.

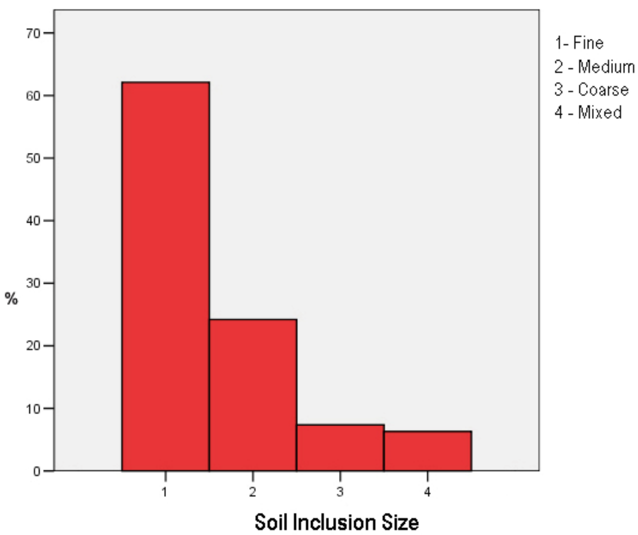


Figure 6.34 Inclusions for site pixels.

6.2.3.4 Erosion

The use of this layer (Figure 6.35) can also be considered as an example of the cultural management vs. academic approach, as in Soil Depth. Again the survival rates of sites (mounds, in this case) can be explored via this layer, and can be used as a filter for future archaeological surveys.

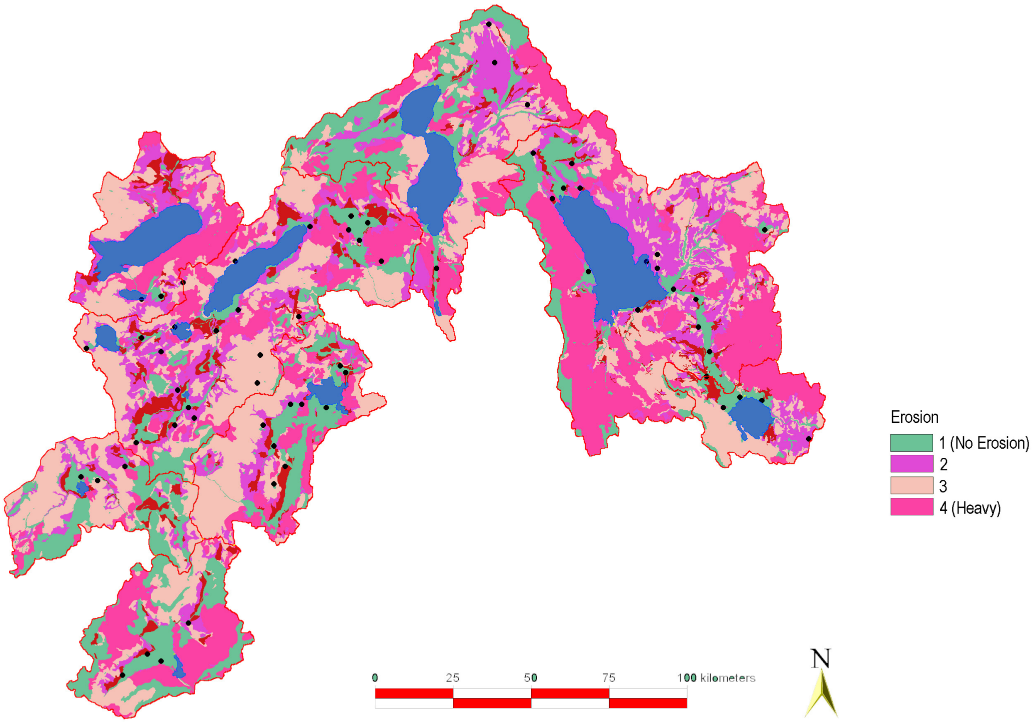


Figure 6.35 Erosion levels of study area.

The overall study area is exposed to heavy erosion (Figure 6.36). Then, it is lucid not to expect flat settlements and sites built up by perishable material in the area. This situation can be a cause for not finding Upper Paleolithic evidence. On the other hand it is not true to charge only one variable of all non-existence. The basic aim of this study is not to solve a hugely complex problem, but to screen possible variables in order to construct better models in future.

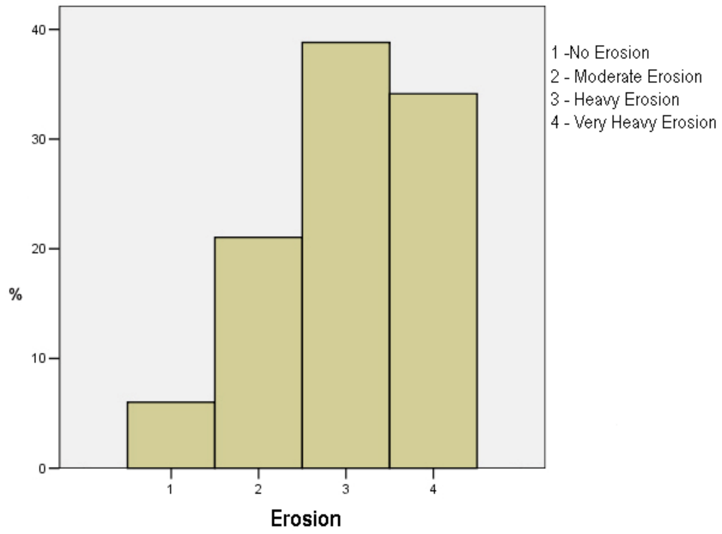


Figure 6.36 Erosion level percentages of the study area.

More than half of the site pixels are observed where there is no erosion or only limited amount of erosion. The existence of 10% with heavy erosion, in contrast, is a guide that even under heavy erosion some sites can survive (Figure 6.37).

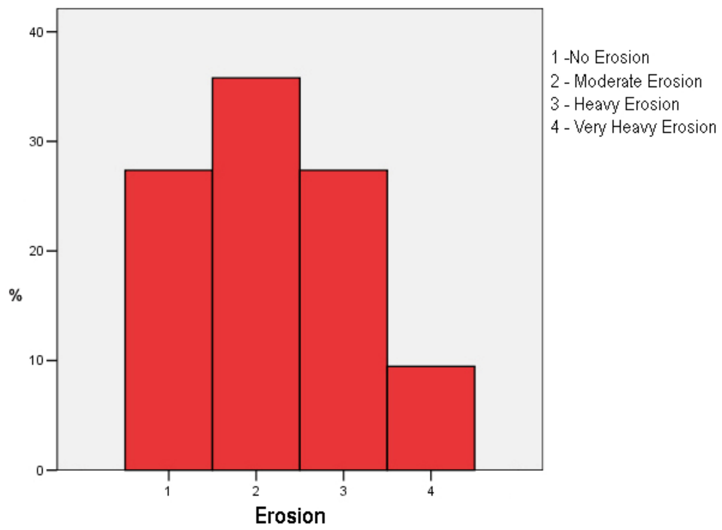


Figure 6.37 Erosion levels for site pixels.

6.2.3.5 Soil Depth

Although this variable is questionable in terms of being a criterion for a site selection procedure, it is important to show the geological processed happened before. Then it is more helpful for a management aspect than an academic aspect. Dependent data is comprised of sites which are surviving. Thus, if geological processes affecting the survival, then this layer can be seen as a filter for the constructed model, in a manner that different processes results in different soil depth, and different survival rates. To examine this, Figure 6.38 is produced.

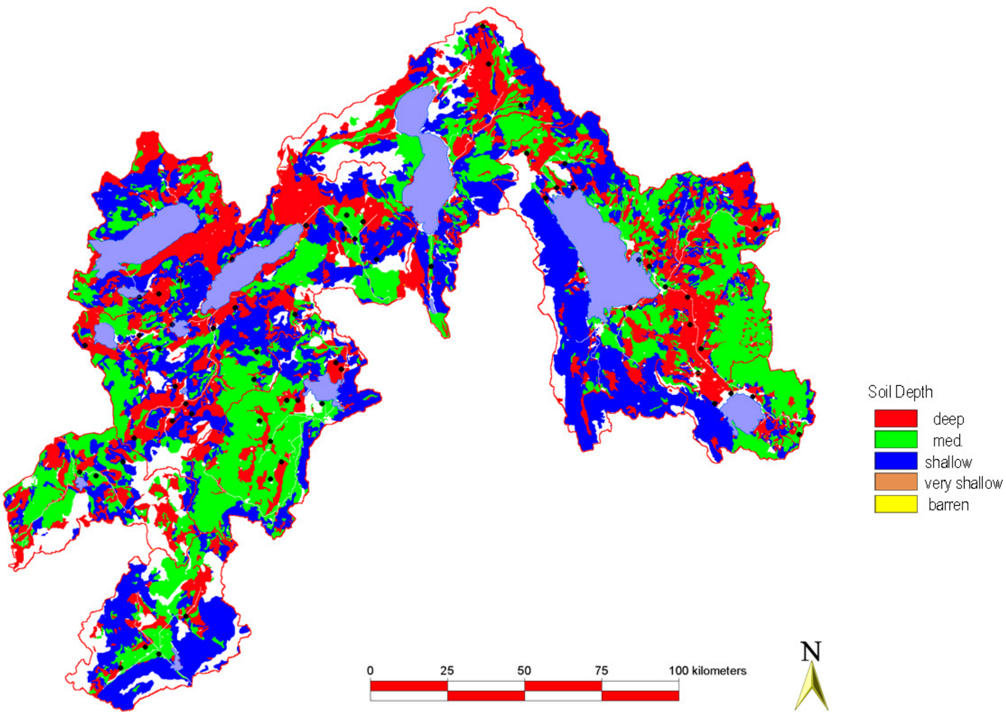


Figure 6.38 Soil Depth distribution of study area

Soil depth of the study area varies from deep to shallow, but the amount of very shallow soil and barren rock is so small that they are not visible as percentages (Figure 6.39).

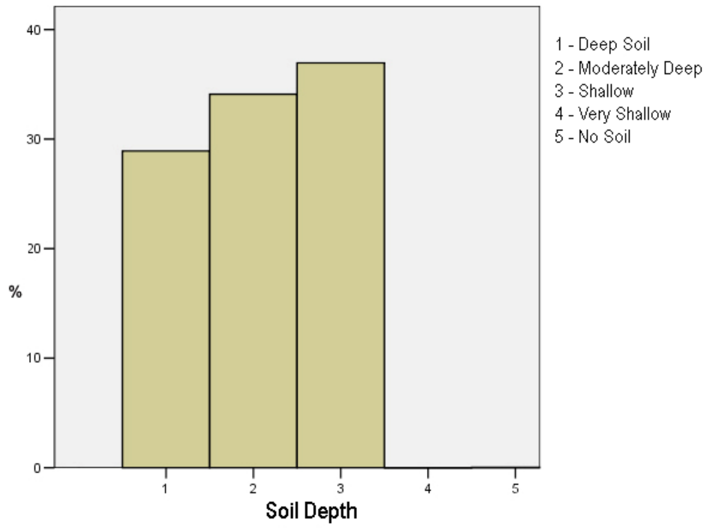


Figure 6.39 Soil Depth Classes and percentages of the study area

Site pixel values are very similar in overall study area values except for shallow soil. This category value is shared between deep and half deep categories (Figure 6.40).

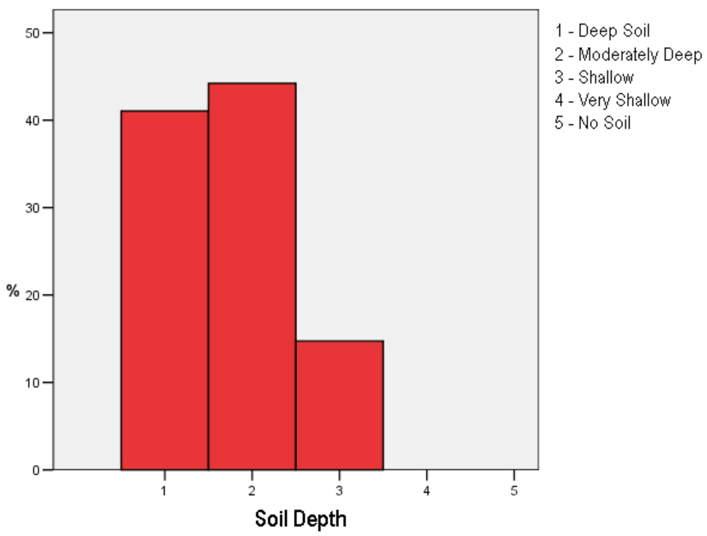


Figure 6.40 Soil Depth Classes of site pixel values.

6.2.3.6 Major Soil Classes

Apart from other measures for soil, type itself can be a basic determinant for site selection. Different cultures might be sought after different soil types (hunter-gatherers vs. agriculturalists), but with the beginning of deforestation, started in very early times, changes in soil classes should have happened. Then it is very critical to use soil classes as a layer in the analysis (Figure 6.41).

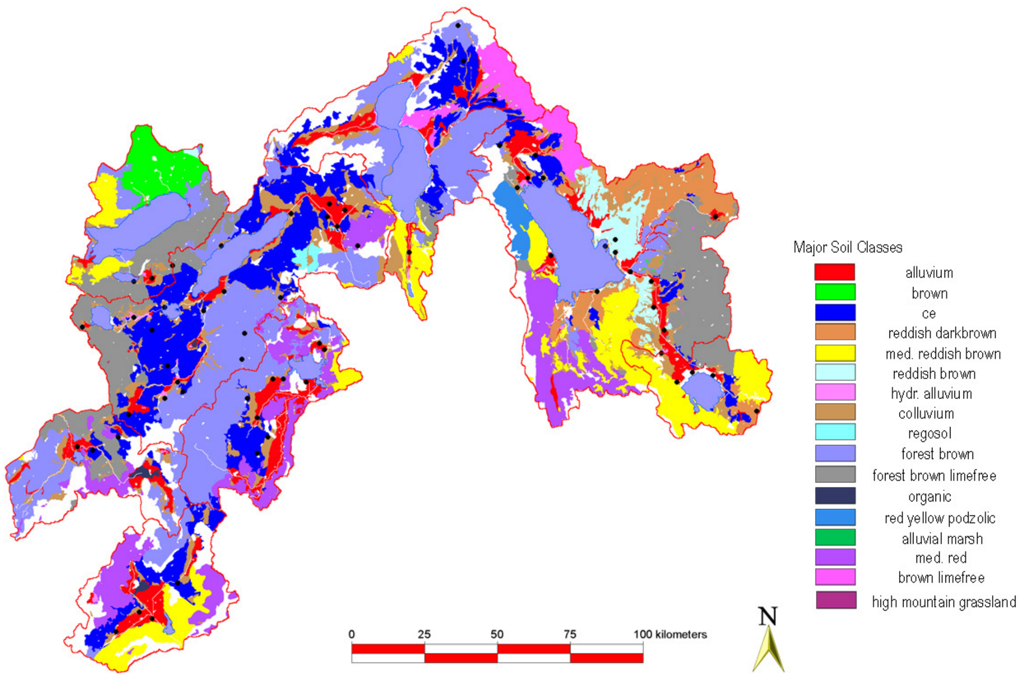


Figure 6.41 Major Soil Class distribution map.

Since there are many soil categories, it is hard to visualize and interpret the graph (Figure 6.42). Moreover, percentages are so small that even an impression of a significance of one particular major soil class should be examined critically.

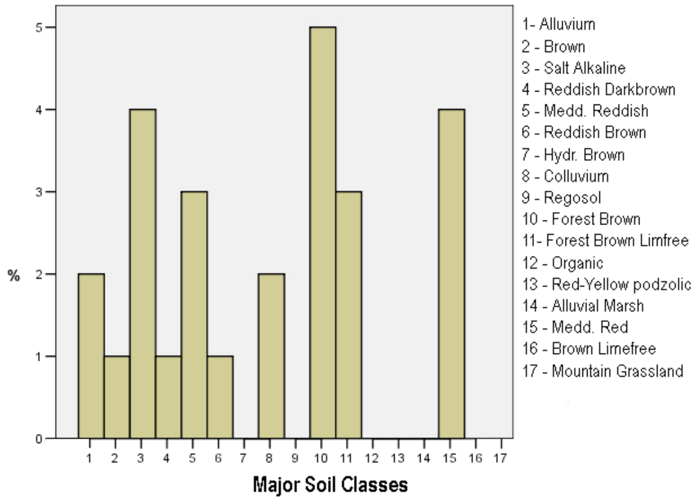


Figure 6.42 Major Soil Classes and their percentages in the study area.

For site values Alluvium gains an important percentage when abundance of soil classes is considered (Figure 6.43). It is also approved by Current Land Use in the area. Assuming that, this category is significant for ancient people in terms of locating their settlements, then it might be considered that environmental shifts were not able to utterly change the landscape and land use, but still this variable is not used due to its unpredictable dynamic character.

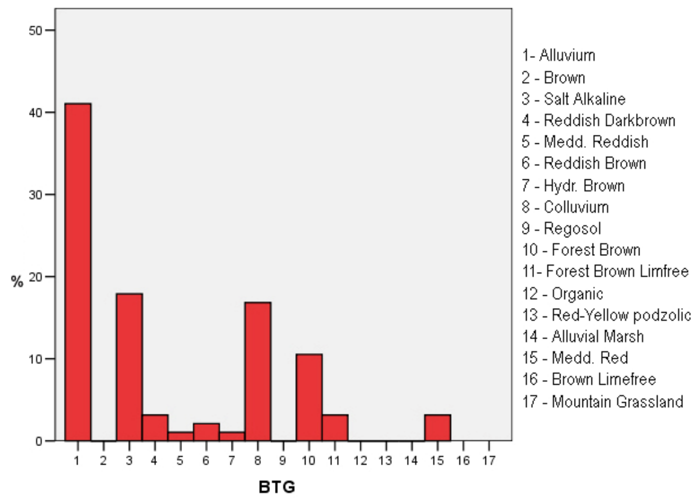


Figure 6.43 Major Soil Classes for site pixels.

6.2.3.7 Distance to Major Soil Classes

If Major Soil Classes is not used as a layer, a layer with a less strict assumption can be used, such as Distance to Major Soil Classes (Figure 6.44). In this case, continuous data measure (distance) is less affected by geological changes when compared with categorical data (classes).

Moreover, it can be easily assumed that there was a tendency to exploit as many ecological niches as possible. Then a site is expected to be located at the periphery of soil classes rather than residing in the core area. Then Distance to Major Soil Classes gains more importance, and surpasses Major Soil Classes.

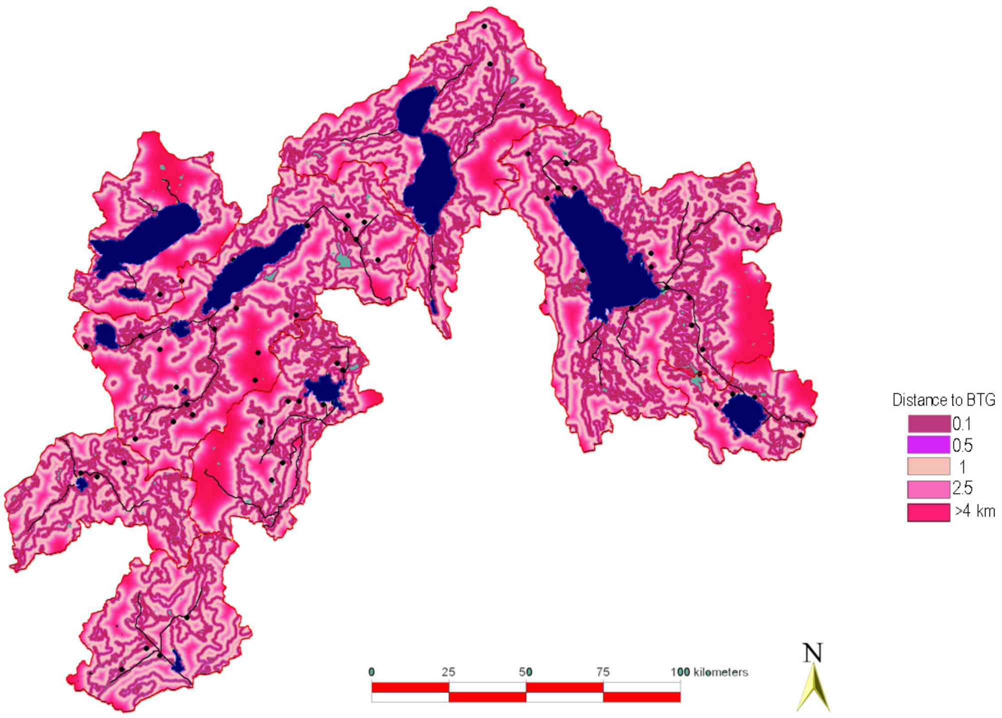


Figure 6.44 Distance to Major Soil Classes in kilometers.

The average distance to any soil category is so low, 2km, that it does not falsify but weakens the idea of exploiting different niches. This can be explained by the abundance of different soil categories.

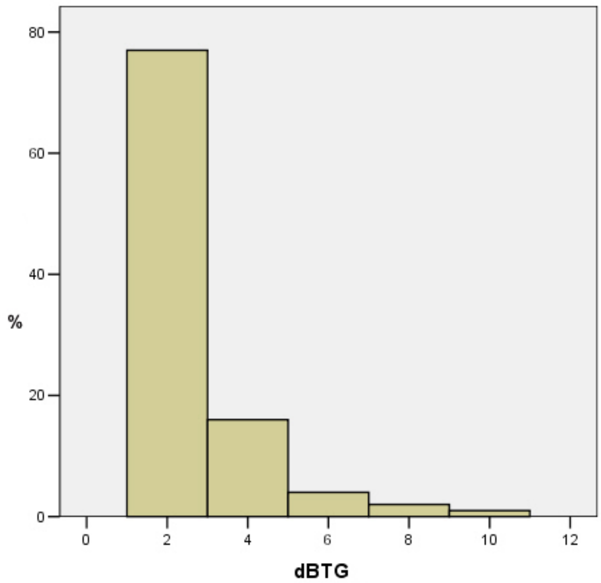


Figure 6.45 Histogram of Distance to Major Soil Classes.

The average distance as well as the distribution is like the overall study area values, so that little can be extracted from this data layer. On the other hand, within site pixel values there is a remarkable decline after 3km, so that layer still deserves attention.

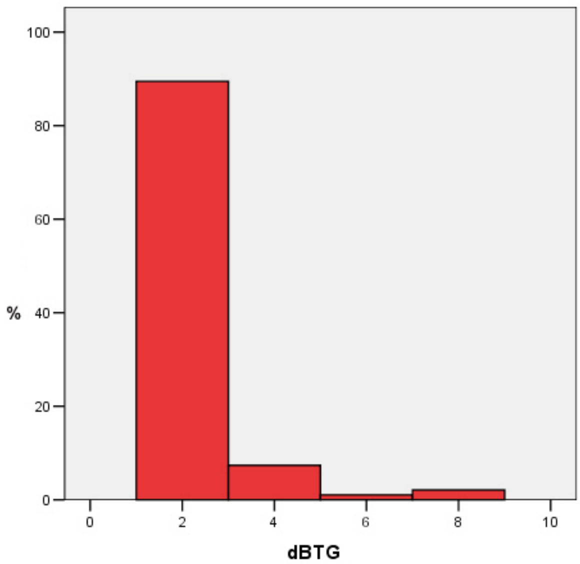


Figure 6.46 Histogram of Distance to Soil Classes for site pixels.

6.2.4 Hydrological Data

6.2.4.1 Distance to Water

Distance to Water layer (Figure 6.47) is the union of two datasets, namely Distance to Water and Distance to River. Thus, it is artificially created in order to observe the behavior of location of sites with respect to the term 'water', and it is not used as a variable in the analysis.

Definitely, such a generalization is not fully correct. In the study area, lakes have different characteristics in terms of water abundance and quality, and rivers have dynamic characters so that any projection is speculative.

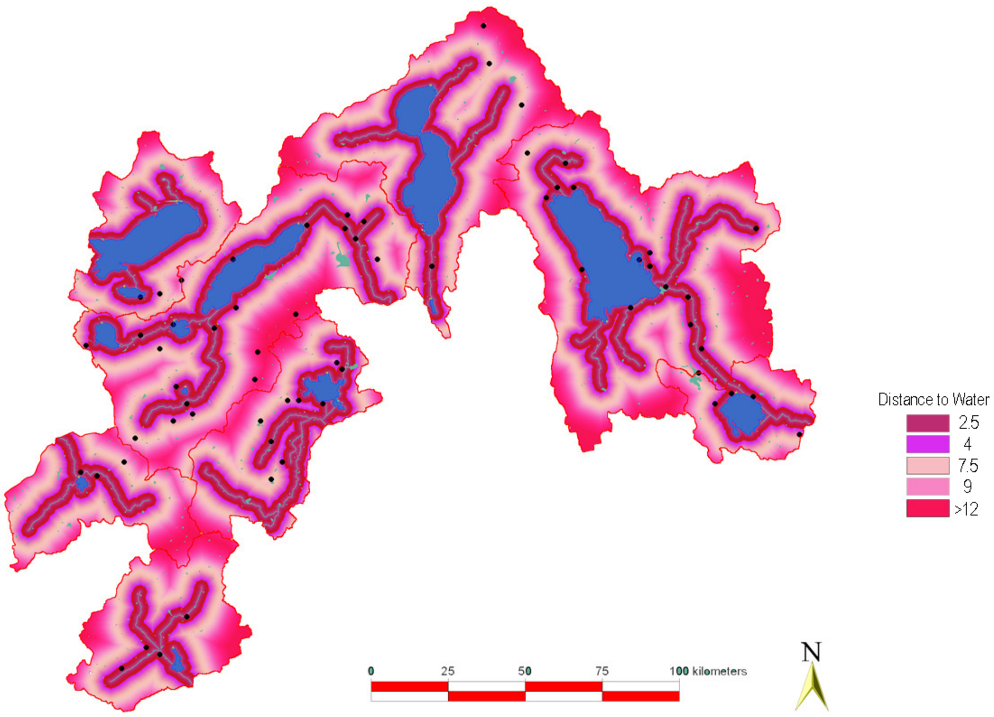


Figure 6.47 Distance to Water (Rivers and Lakes).

Any pixel in study area is having an average of 9km to water source, namely rivers or lakes. The skewed distribution implies that there is no big water shortage issue in the study area, assuming those sources can be used for domestic purposes as well as for social considerations (Figure 6.48).

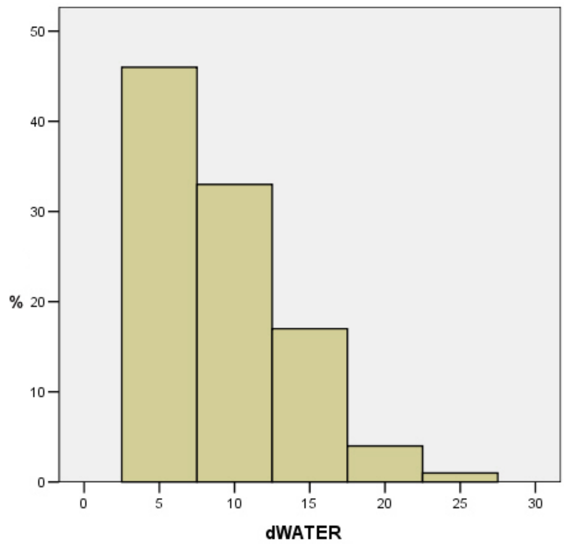


Figure 6.48 Histogram for Distance to Water.

Site pixel values have a slightly lower average than overall pixels, but there is no change in the shape of distribution. The histogram is helpful in terms of showing a general tendency, and being close to water is a general fact for sites.

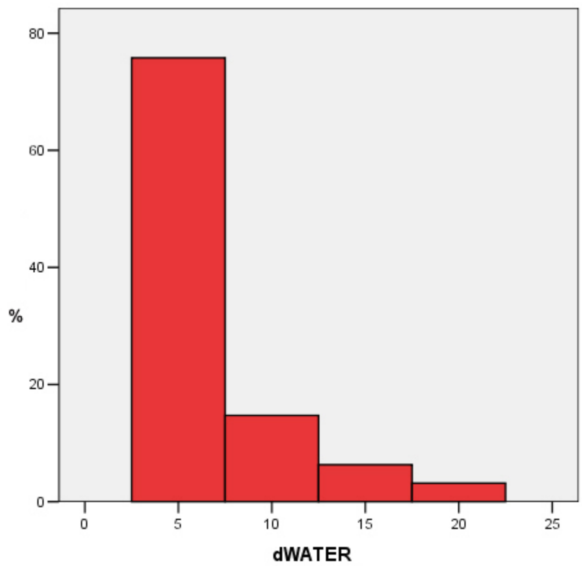


Figure 6.49 Histogram of Distance to Water for site pixels.

6.2.4.2 Distance to River

River courses are obtained by employing GIS watershed analysis. Distances are obtained in kilometers. Since seasonality, soil type and various others are important factors in the existence of a river, it is of immense importance to create a layer that is not affected by such parameters. The easiest solution, which is also used here, is to include major branches of rivers, and omit less loaded ones.

A general measure to classify rivers is the Strahler Order. In this study, river courses having Strahler Order 4 and more is used, since the lower class ones produce redundant data in terms of distance. Since only major courses were considered the domestic use of water is overlooked. Then the use is not only domestic anymore, but some social considerations are also included, such as transportation. Moreover, there are some site locations which cannot make use of branches of rivers, but were still able to survive, which can be explained with the existence of springs, which most of the modern villages are based on.

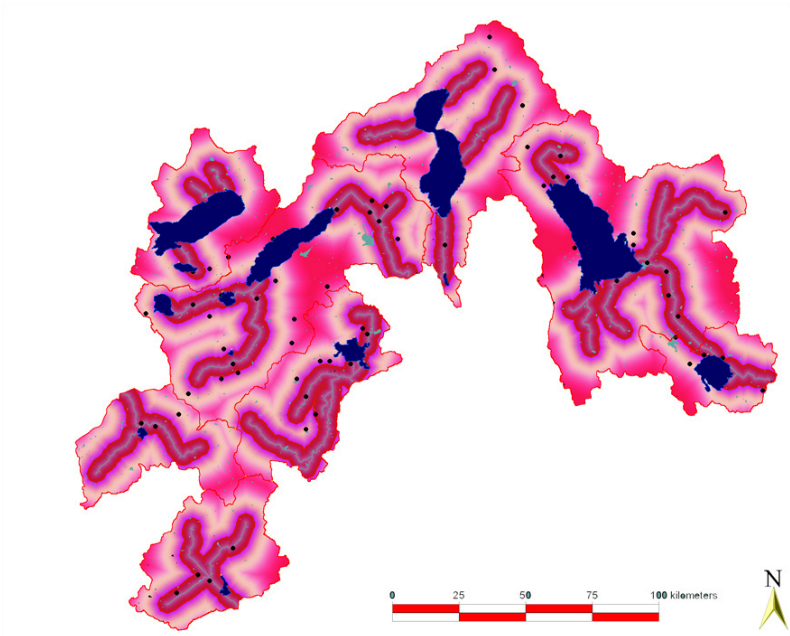


Figure 6.50 Map of Distance to Rivers.

Although river network is constructed according to Strahler Order 4, the river system covers the study area so that the average distance of a pixel to any river is 9km (Figure 6.51).

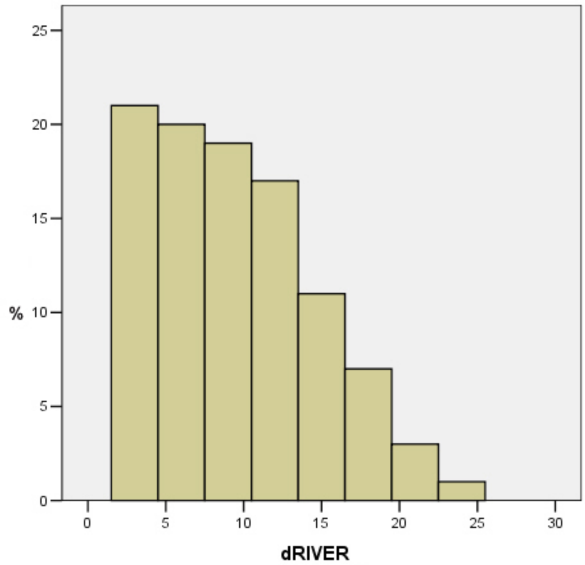


Figure 6.51 Histogram for values of Distance to River.

Average Distance to River for site pixels is slightly less than general average. Moreover, sites and non-sites have similar distributions, but there is a cutoff value after 12km (Figure 6.52).

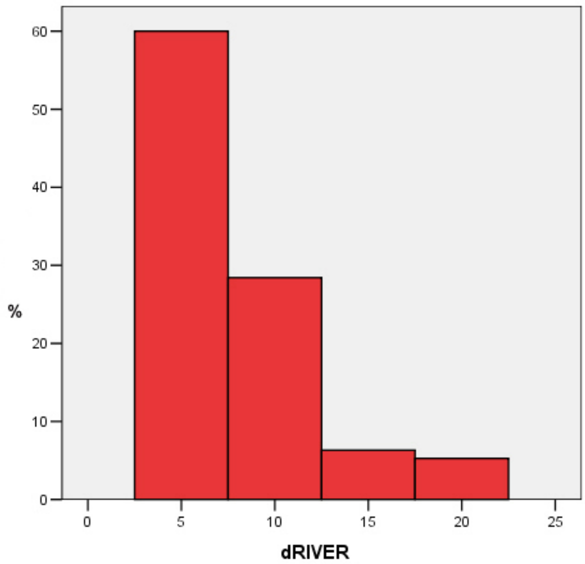


Figure 6.52 Histogram for site pixel values of Distance to River.

6.2.4.3 Distance to Lake

Lakes are major players in the study area. Each basin is selected with a focal lake. As expected, current lake levels cannot be consistent with ancient lake levels. Even, some lakes do not exist anymore, such as Suğla Lake. Then it is essential to construct ancient lake levels. Lake levels used in the analysis are given in Table 6.4. Marked levels are hypothetical ancient lake levels, remaining is used as it is since to information is available. For the given lake levels, Distance to Lake map is produced (Figure 6.53).

Table 6.4 Lake Levels used in the study.

Avlan	Acıgöl	Beyşehir	Burdur	Eğirdir	Gölhisar	Kestel	Suğla
1028m	860m	1122m	870m	916m	950m	784m	1093m

But definitely it can be stated that the levels were higher than today. This is due to both climate change and by-passing sources of lakes.

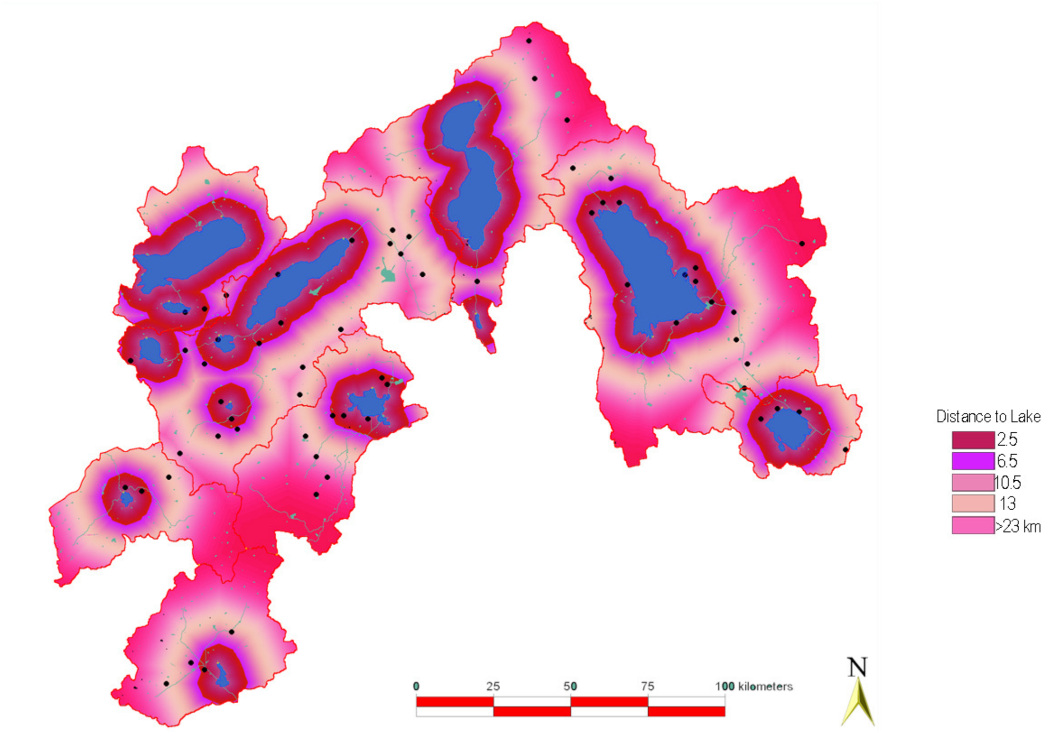


Figure 6.53 Map of Distance to Lake**.**

Pixel values for Distance to Lake have an average of 17km in the study area. This is mainly due to the abundance of lakes and lake systems (Figure 6.54). This value does not take into account different characteristics of lakes.

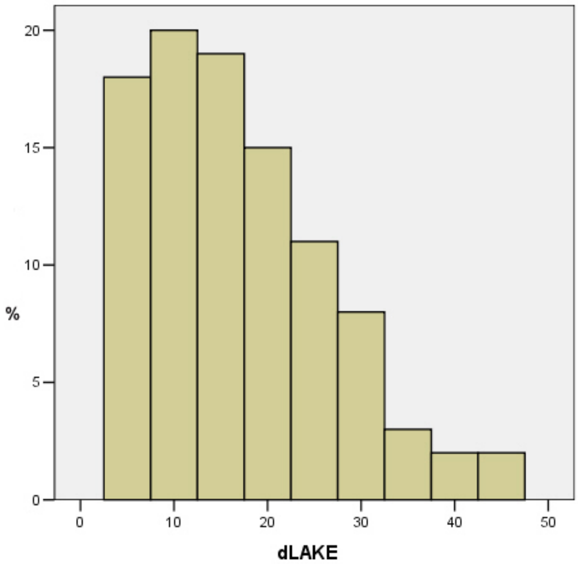


Figure 6.54 Histogram of Distance to Lake Values of the study area.

Values for site pixels are slightly different than study area values. It has an average distance of 15km, but most importantly there is a clear choice for being close to lake, and there is a sharp decline after 20km of distance (Figure 6.55).

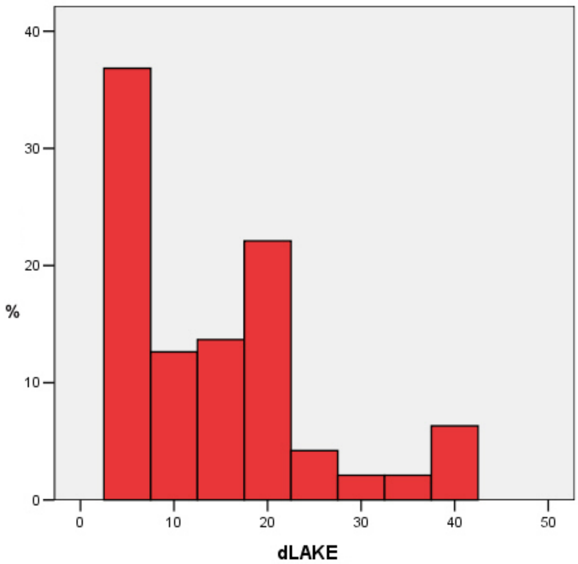


Figure 6.55 Histogram of Distance to Lake Values of site pixels in the study area.

CHAPTER 7

MODEL

After evaluating the possible choices for the construction of a predictive model it was decided to use logistic regression. This has been selected as a tool since it provides a comprehensive statistical evaluation of the subject. Moreover, it is much more open for future studies. Another interesting tool which stems from Dempster-Schafer Theory was not chosen, since it is still in its infancy for archaeological applications.

Although Logistic Regression considers less strict assumptions, it is also problematic for this particular study. This is because, there is a huge imbalance between the dichotomous dependent variable. In more than 1 million pixels, there are only 95 site pixels, and 10 of them are withheld from the total for testing purposes. For such an extreme case, there is no clear answer for constructing a model. To solve this, non-site pixels selected randomly from whole study area. In this selection, there is a possibility of using sites as non-sites, which will introduce and error to model, but considering the total amount of pixels this is unlikely to happen.

The other problem is the determination of sample size for possible non-site values. As a criterion, behavior of standard error of coefficients is used. Double size of site-pixel values, which is around 200, produces inflated standard errors, whereas a sample around 120 produces considerable standard errors

But then another question is raised about the coverage of sample of size 205 (120+85) for such a huge study area. To over come this up to a point, 20 randomly selected samples are used to construct 20 different predictive models. Then a resulting model is obtained from those different predictive models by taking average.

7.1 Predictive Models

Each predictive model constructed is given with the resulting probability surface and its logistic regression formula. For continuous data coefficients are supplied in the equation, whereas categorical data coefficients are given as a list. For any pixel, the regression equation is run to obtain the resulting image.

Sites and non-site values are put into a statistical program (SPSS 13.0 Evaluation Version). Response variable is encoded as 1 for pixels having sites and 0 for pixels having no sites. Cut value is selected as 0.5 for each model and according to this value beginning block predicted sites around 55% when intercept values were included in the models.

For categorical variables, Indicator contrast is used and as reference, the last category of each categorical variable is used.

As the method, Forward Stepwise (Likelihood Ratio) is selected, and variables entering to the model are selected accordingly. '-2 Log likelihood' is used as a criterion and after deciding on the final variables with a cut value 0.5 prediction values are observed. An average of 80% total true prediction is obtained. Necessary model summary is given in Appendix - D

As model diagnostics, VIF (Variance of Inflation Value) and tolerance levels are used to detect multicollinearity. SPSS 13.0 does not provide this tool for Logistic Regression, so model variables are treated as ordinary linear regression components. Although this is not the best way to test multicollinearity, it is given as a way in (Field, 2000). Results are given in Section 7.5 in a combined table for each predictive model, and categorical variable codes are given in Appendix-A.

7.1.1 Predictive Model 1 (PM1)

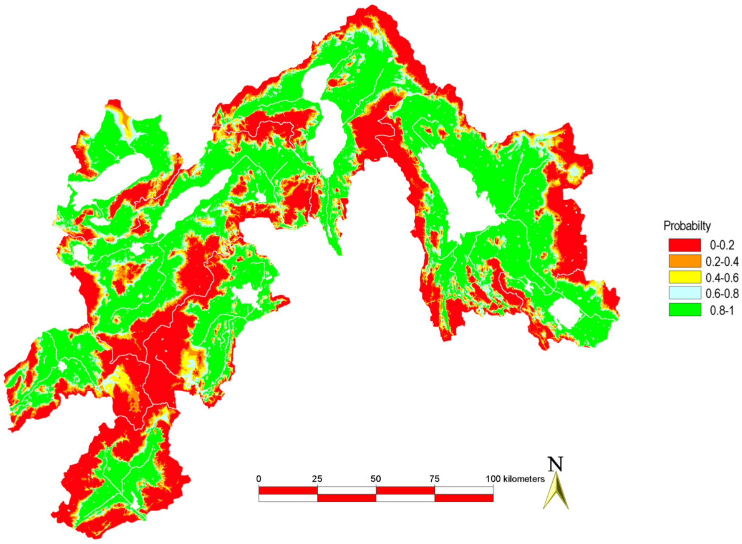


Figure 7.1 Probability surface of PM1.

$\hat{y}$=7.194+(-0.097xS)+(-9.63xEA)+(-0.096xDtL)+(-0.657xDtMSC).

S: Slope – EA: Adjusted Elevation – DtL: Distance to Lake – DtMSC: Distance to Soil Classes

Randomly selected 'non-site' values with 85 'site' values have produced a highly optimistic predictive model. The study area is basically divided into two: Areas with very high probability, that is sites very likely to occur, and areas with the smallest probability, in other words areas in which no sites are expected.

There are very few pixels assigned to transition probabilities. Two continuous variables Adjusted Elevation and Distance to Lake put emphasis on vicinity of lakes. The other two variables as well as the two discussed have negative values. That is, a higher degree of slopes and elevation, and increased distance from lakes and major soil classes does not favor the occurrence of sites. There are no categorical variables in the equation, which is due to random selection of the training sample and study area pixels.

7.1.2 Predictive Model 2 (PM2)

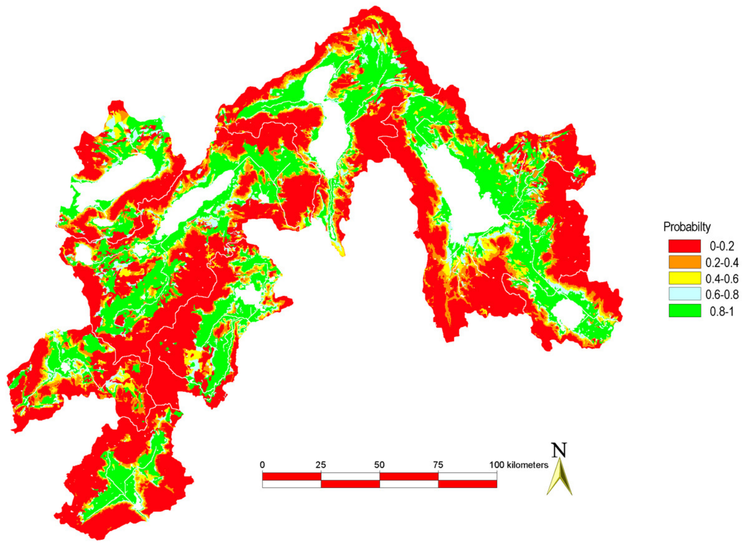


Figure 7.2 Probability surface of PM2.

$\hat{y}$=3.674+(-0.097xS)+(-6.82xEA)+(-0.072xDtL)+(-0.416xDtMSC)+A+CLU

S: Slope – EA: Adjusted Elevation – DtL: Distance to Lake – DtMSC: Distance to Soil Classes – A: Aspect – CLU: Current Land Use

Categorical Variable Coefficients for their particular values are given as:
A(1): 5.930, (2):-0.445, (3):-0.518, (4):-1.823,
(5):-2.648, (6):-2.909, (7):-1.866, (8):-0.865
CLU(1): 2.131, (2): 1.434, (3):-0.092

A new set of samples produced a similar predictive surface with a shift in very high and very low probabilities. Emphasis is still on lake vicinities and at adjusted low elevations.

Two categorical variables with given coefficients are introduced to the model. Continuous variables still have negative coefficients, and model has a smaller intercept value. Then again, there are very few pixels with transitional probabilities.

7.1.3 Predictive Model 3 (PM3)

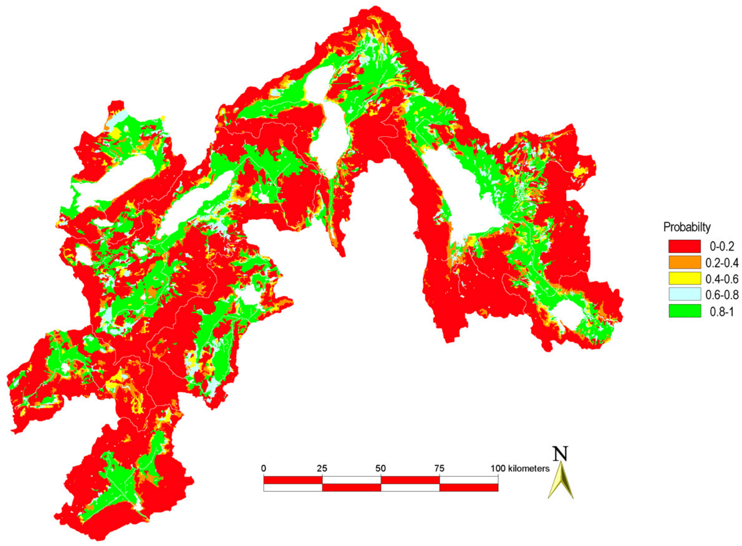


Figure 7.3 Probability surface of PM3.

$\hat{y}$=1.503+(-6.869xEA)+(-0.050xDtL)+A+CLU

A: Aspect – EA: Adjusted Elevation – DtL: Distance to Lake – CLU: Current Land Use

A (1): 0.193, (2): 1.401, (3): 0.285, (4):-0.859, (5):-1.905, (6):-2.411, (7):-1.924, (8): 0.042
CLU(1): 2.095, (2): 8.025, (3):-0.368

Result of PM3 is not so different from result of PM2. The same zones are marked as high probability areas with nearly the same coverage. Transition probabilities also located at roughly the same places as in the previous model. This is basically due to employing same variables except Slope, and Distance to MSC. Variable coefficients are also similar to other model coefficients. The probability of finding a site decreases as one gets higher in elevation and gets further away from a lake. As a land use coefficient, if Shrubbery and Forests (4) are fixed then Grasslands (3) are avoided and Agricultural Fields (1) and Yards and Gardens (2) are preferred. The same discussion can be made for Aspect values according to their signs of coefficients.

7.1.4 Predictive Model 4 (PM4)

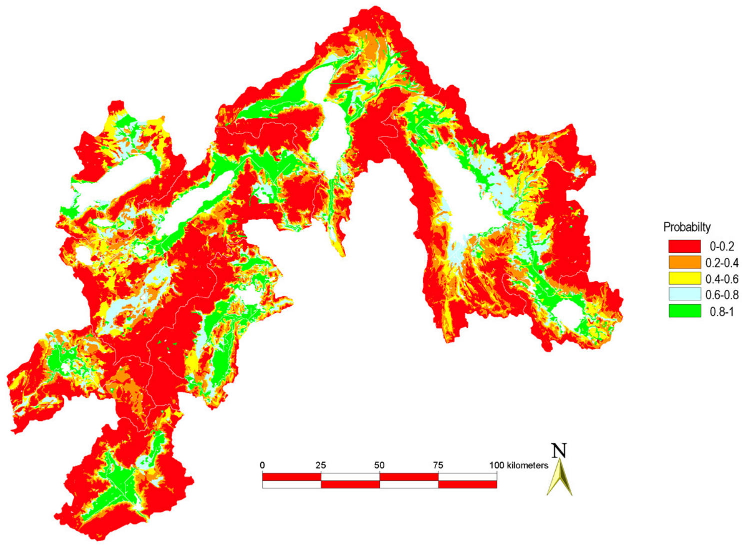


Figure 7.4 Probability surface of PM4.

$\hat{y}$=-0.388+(-5.963xEA)+L+CLU

L: Lithology – CLU: Current Land Use – EA: Adjusted Elevation

CLU(1): 1.197, (2): 8.0, (3):-0.503
L (1): 1.814, (2):-3.007, (3): 1.810, (4): 1.107, (5): 0.044

PM4 produces fewer maximum probability areas, but transitional probabilities are much clearer in this model, and they are basically located at the East and West sides of study area. Interestingly, the intercept coefficient of the predictive model is below zero in this case. Moreover, the coefficient of the Adjusted Elevation is also significantly different from previous models. Current Land Use and Lithology are also in the model as categorical variables. Current Land Use has similar emphasis on labels of data. Again, Agricultural Fields (1) and Yards and Gardens (2) are preferred and Grasslands (3) are avoided when they are compared with Shrubbery and Forests (4). Similarly, Lithology coefficients can be interpreted as when Clastics and Carbonates (6) are fixed then Volcanics (2) are avoided and rest is significantly preferred except for Clastics (5) itself.

7.1.5 Predictive Model 5 (PM5)

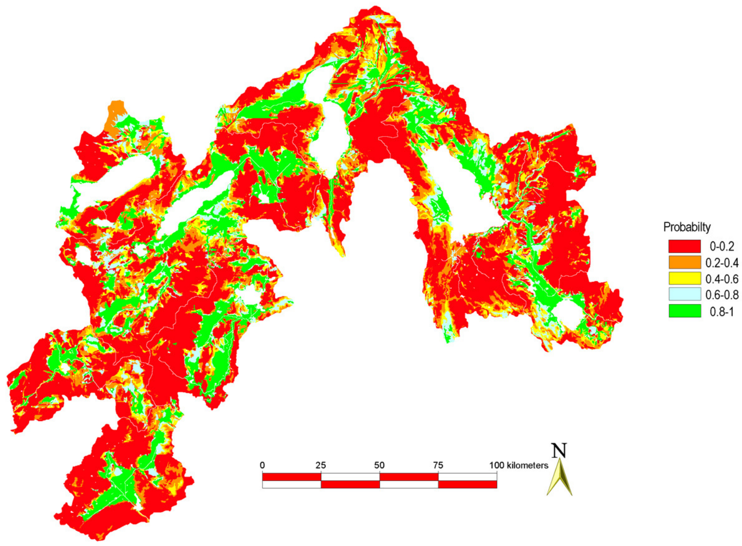


Figure 7.5 Probability surface of PM5.

$\hat{y}$=-2.289+(-4.885xEA)+(-0.188xDtoR)+A+CLU+L

EA: Adjusted Elevation - A: Aspect - L: Lithology - CLU: Current Land Use DtR: Distance to River

A (1): 1.617, (2): 1.380, (3): 0.622, (4):-0.943, (5): 0.117 , (6):-0.682, (7):-0.047, (8): 0.270
CLU(1): 1.763, (2): 9.650, (3): 0.907
L (1): 1.331, (2):-8.011, (3): 0.844, (4): 1.228, (5):-0.643

This model also has negative intercept value. The coefficient of Adjusted Elevation is also decreased. This might be due to a new variable introduced, Distance to River.

Although balance between low and high probabilities is unchanged, the area for transitional probabilities has decreased. This should be basically a result of employing different combinations of categorical and continuous variables.

7.1.6 Predictive Model 6 (PM6)

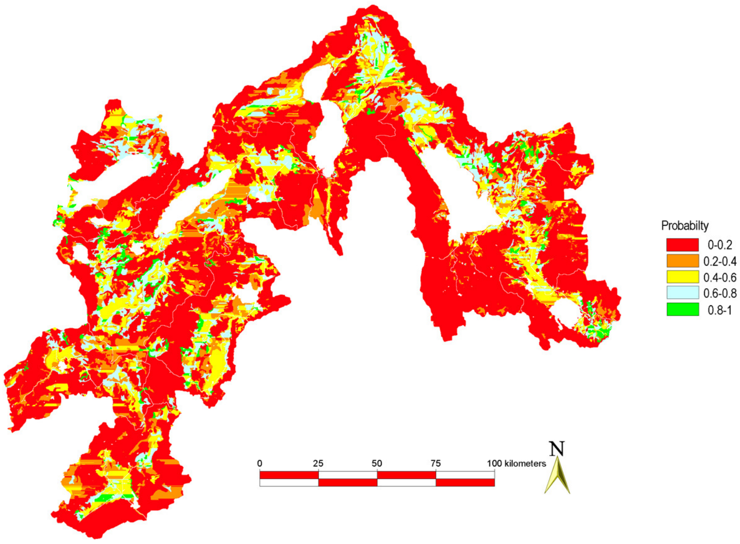


Figure 7.6 Probability surface of PM6.

$\hat{y}$=-3.412+A+CLU+L+TOK_D

A: Aspect - L: Lithology - CLU: Current Land Use – TOK_D: Soil Depth

A(1): 0.901, (2): 0.532, (3): 1.462, (4): 0.479, (5):-0.567, (6):-1.146, (7):-0.757, (8):-0.068
CLU(1): 1.768, (2): 0.787, (3): 0.684
L(1): 1.597, (2): 1.397, (3): 1.439, (4): 0.253
TOK_D(1): 0.386, (2): 1.723

This model has produced a very pessimistic result on existence of sites. Only few areas have high probabilities, whereas the whole study area is dominated by low probability pixels. A new categorical variable, Soil Depth, is introduced and model does not make use of continuous variables.

7.1.7 Predictive Model 7 (PM7)

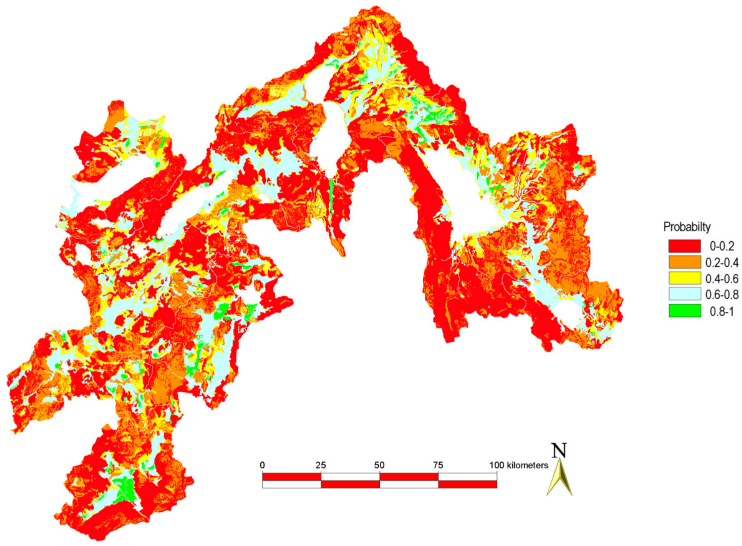


Figure 7.7 Probability surface of PM7.

$\hat{y}$=-0.155+(-0.075xS)+A+TOK_D

A: Aspect – S: Slope – TOK_D: Soil Depth

A (1): 0.478, (2): 0.689, (3): 0.296, (4):-0.671, (5):-1.124, (6):-1.824, (7):-1.241, (8):-0.194
TOK_D(1): 0.375, (2): 1.007

This model resembles to PM6 in terms of pixel probability combinations and distributions. On the other hand, a continuous variable, Slope, is in the model this time. A negative coefficient of Slope states that as the degree increases, the probability of finding a site is decreased. This is an expected case but, as stated earlier, the human effect on topography should be considered as well.

7.1.8 Predictive Model 8 (PM8)

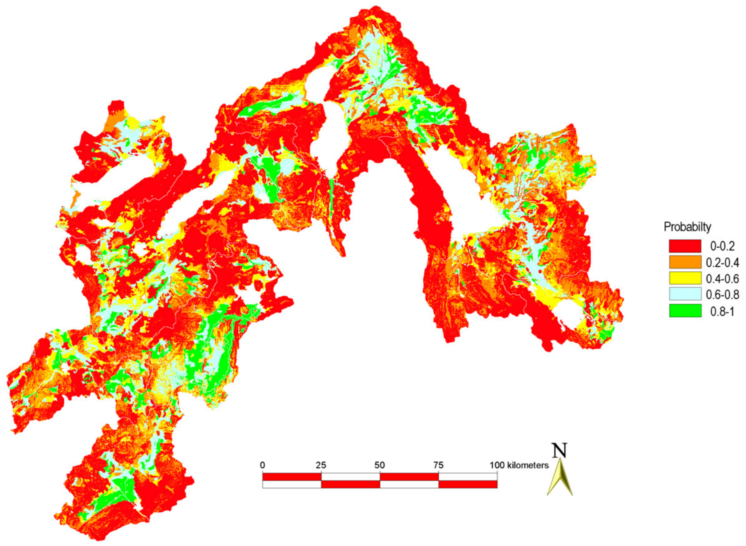


Figure 7.8 Probability surface of PM8.

$\hat{y}$=0.415+(-0.099xS)+(-0.107xDtR)+(0.047xDtL)+A+TOK_D

A: Aspect – S: Slope – DtL: Distance to Lake – TOK_D: Soil Depth
DtR: Distance to River

A (1): 0.276, (2): 0.530, (3): 1.082, (4):-0.633, (5):-0.468, (6):-1.622, (7):-1.226, (8):-0.551
TOK_D(1):-0.405, (2): 0.823

This model has increased probability pixels in southwest of the study area. The significance of the model is the positive coefficient of the Distance to Lake, that is, the probability of site existence increases as one gets further away from a lake, which is contradicting with the previous models. This situation replaces high probability pixels with low probability pixels around lakes.

7.1.9 Predictive Model 9 (PM9)

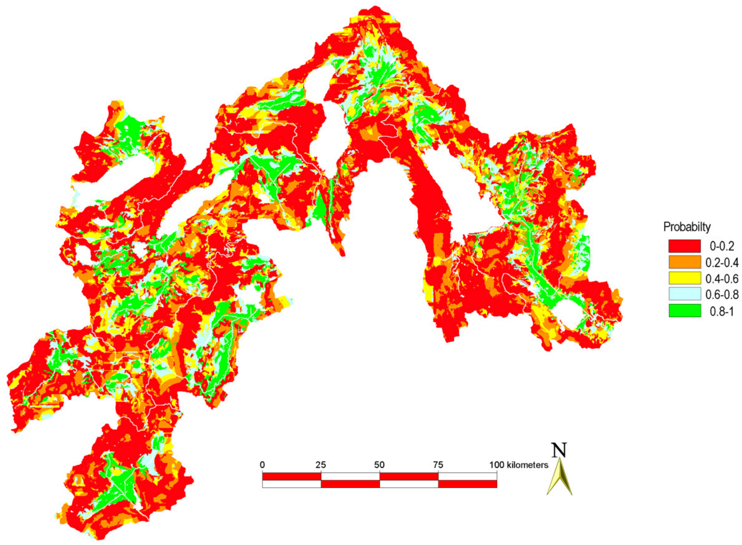


Figure 7.9 Probability surface of PM9.

$\hat{y}$=-0.389+(-0.093xDtR)+(0.323xDtMSC)+A+CLU+RGH

DtR: Distance to River - DtMSC: Distance to Soil Classes – A: Aspect - CLU: Current Land Use
RGH: Surface Roughness

A (1):-0.197, (2): 0.175, (3):-0.249, (4):-1.507, (5):-1.414, (6):-2.378, (7):-1.483, (8):-1.041
CLU(1): 1.831, (2): 1.818, (3): 0.573
RGH(1):-0.573, (2): 0.294, (3): 1.270

This is an ideal result for a predictive model with distinct high and low probabilities as well as some transitional probabilities in between those high and low values, but since Distance to Lake is not in the model, an expected emphasis to those areas is missing. A categorical variable, Roughness is included in the model, favoring moderately rough areas, where no extremely rough data (4 and 5) is included via random selection.

7.1.10 Predictive Model 10 (PM10)

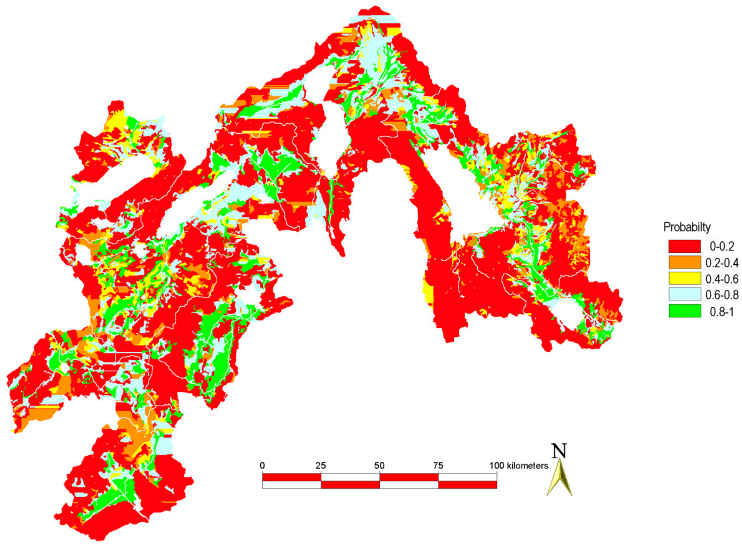


Figure 7.10 Probability surface of PM10.

$\hat{y}$=-2.254+A+CLU+TOK_D

A: Aspect – CLU: Current Land Use – TOK_D: Soil Depth

A (1): 1.423, (2): 0.648, (3): 0.679, (4):-0.380, (5):-0.982, (6):-1.690, (7):-0.671, (8): 0.258
CLU(1): 1.767, (2): 1.745, (3): 0.090
TOK_D(1): 0.819, (2): 1.515

No continuous variable is used to construct the model, but a logical model was still produced, with a fairly good distribution of probabilities. This implies that any combination of independent variables can produce any model so that no perfect combination is possible with a particular sampling strategy. Soil Depth favors deep soil, and Land Use is in favor of agricultural fields and yards and gardens. Aspect values are in consistency with other models, where southern directions are avoided while locating settlements.

7.1.11 Predictive Model 11 (PM11)

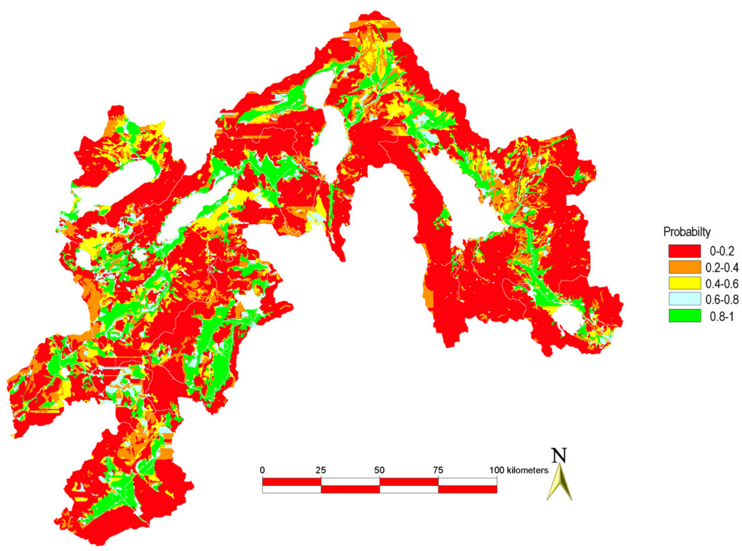


Figure 7.11 Probability surface of PM11.

$\hat{y}$=-2.186+A+CLU+L+TOK_D

A: Aspect – CLU: Current Land Use – TOK_D: Soil Depth – L: Lithology

A (1): 8.341, (2): 0.887, (3): 0.204, (4):-0.598, (5):-0.555, (6):-1.452, (7):-1.042, (8):-0.037
CLU(1): 1.649, (2): 0.471, (3): 0.303
L (1): 1.390, (2):-7.137, (3): 0.508, (4):-0.125, (5):-0.183
TOK_D(1): 0.007, (2): 1.342

This is another model with no continuous variable, and with a negative intercept value. Aspect values favor non-south directions, and in terms of lithology, volcanics, clastics, and clastics and carbonates are avoided. Agricultural areas are also contributing to the model.

7.1.12 Predictive Model 12 (PM12)

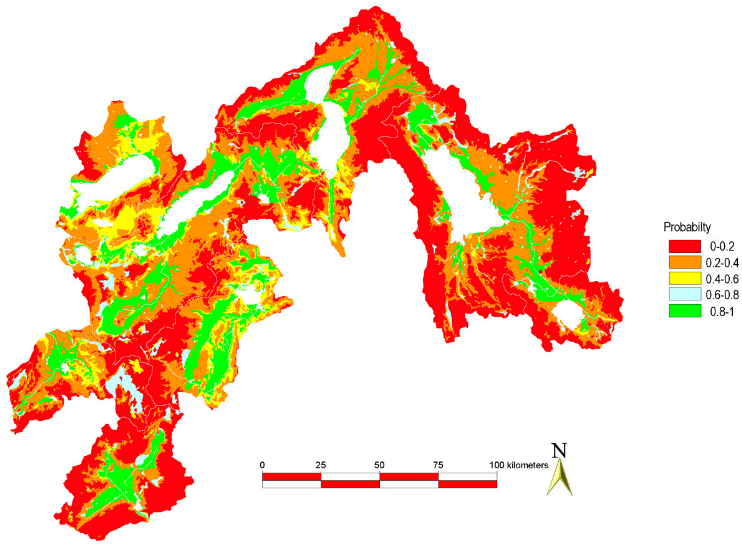


Figure 7.12 Probability surface of PM12.

$\hat{y}$=2.462+(-2.879xE)+L

L: Lithology – E: Elevation

L (1): 2.305, (2):-4.352, (3): 0.726, (4):-0.025, (5):-0.228

This model makes use of limited number of variables, namely Elevation and Lithology. With a positive intercept, low areas and quaternary lithology is producing high probability.

The resulting map is very similar to the distribution of lithology data. Then elevation data is used as an adjustment for the model.

7.1.13 Predictive Model 13 (PM13)

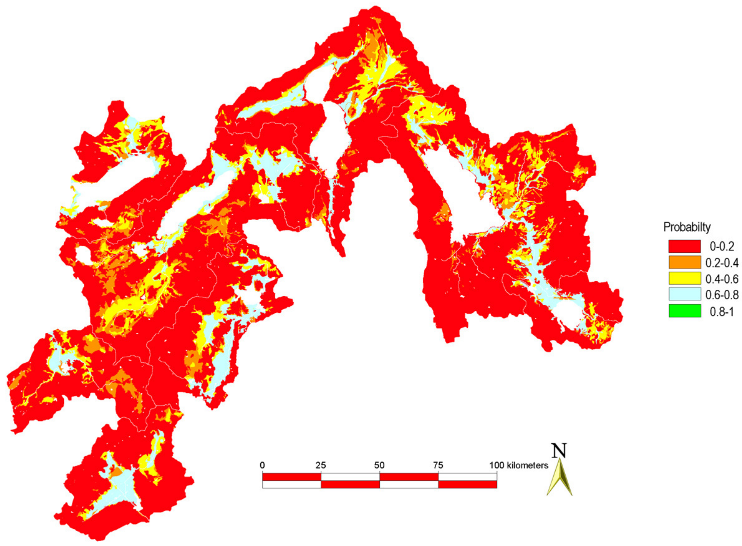


Figure 7.13 Probability surface of PM13.

$\hat{y}$=-0.431+(-3.882xEA)+A+CLU

A: Aspect - CLU: Current Land Use – EA: Adjusted Elevation

A (1): 1.753, (2): 1.402, (3): 1.317, (4):-0.216, (5):-0.646, (6):-1.469, (7):-0.928, (8): 0.091
CLU(1): 1.356, (2): 1.783, (3): 0.438

The variables produce probabilities with very small probabilities, and even with no maximum probability. Transition probabilities have occurred with patches of areas indicating that one or a simple combination of variables were influential on the model.

7.1.14 Predictive Model 14 (PM14)

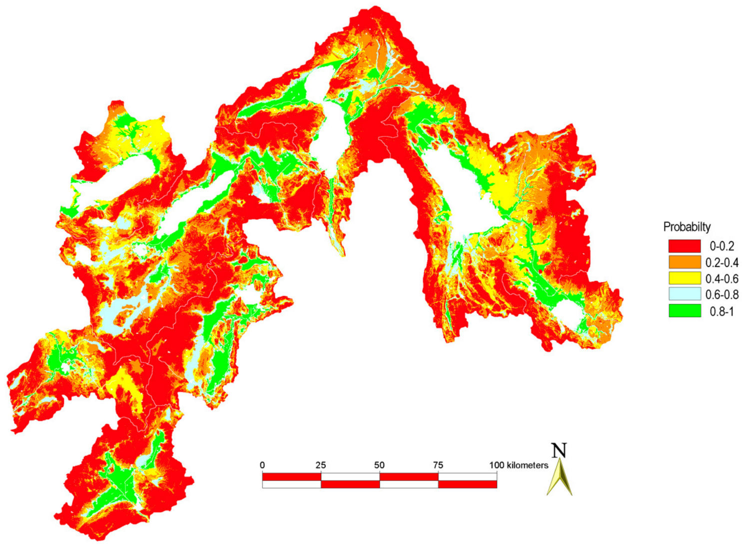


Figure 7.14 Probability surface of PM14.

$\hat{y}$=0.751+(-0.092xS)+(-3.99xEA)+L

S: Slope – EA: Adjusted Elevation – L: Lithology

L (1): 1.287, (2):-5.436, (3): 0.876, (4): 0.704, (5):-0.354

This is again a simplistic model with limited number of variables. As in most other models, there is a focus on lake vicinities, but the variable causing this is not Distance to Lake, but Adjusted Elevation. Slope variable is also influential for producing higher level probabilities. In terms of lithology, the Quaternary category is the contributor.

7.1.15 Predictive Model 15 (PM15)

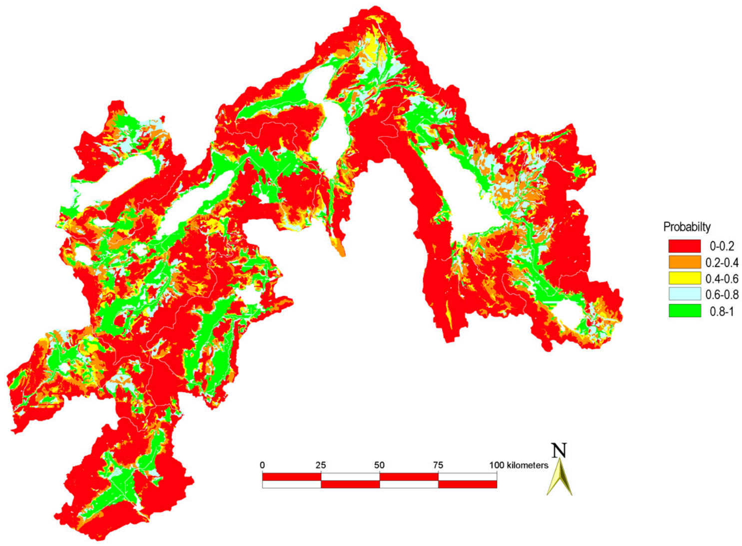


Figure 7.15 Probability surface of PM15.

$\hat{y}$=-1.499+(-4.011xEA)+A+L+ ERZ

A: Aspect – EA: Adjusted Elevation – L: Lithology – ERZ: Erosion

A(1): 2.393, (2): 1.160, (3): 0.970, (4):-0.488, (5):-0.705, (6):-0.581, (7):-0.288, (8):-0.001
L(1): 1.733, (2):-3.509, (3): 0.917, (4): 0.112, (5):-0.581
ERZ(1): 1.601, (2): 2.288, (3): 1.635

The model produces small number of transition probabilities. Favorable areas do not contradict other predictive models. Southern directions, and volcanics, carbonates and clastics are avoided. Erosion, surprisingly, made no significant differentiation between the erosion categories when entered to the model.

7.1.16 Predictive Model 16 (PM16)

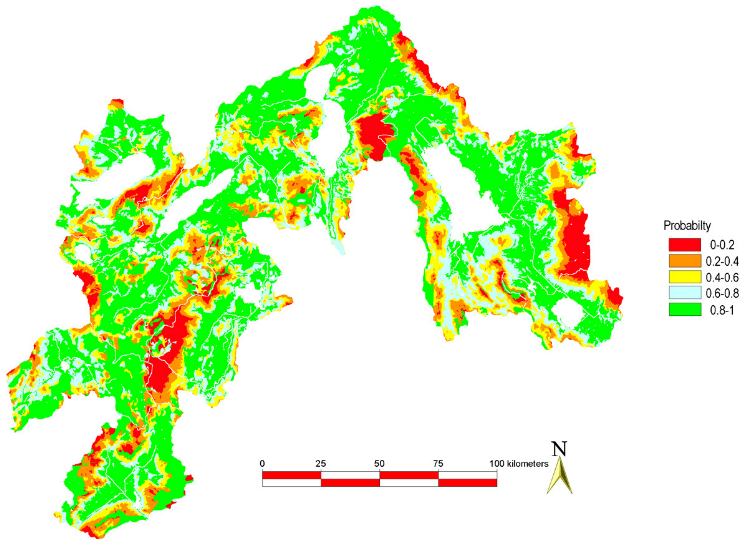


Figure 7.16 Probability surface of PM16.

$\hat{y}$=6.984+(-3.073xE)+(-0.072xDtMSC)+A

DtMSC: Distance to Soil Classes – A: Aspect – E: Elevation

A(1):-1.931, (2):-3.198, (3):-1.723, (4):-3.150, (5): 4.621, (6):-4.610, (7):-5.435, (8):-0.323

PM16 is another optimistic model produced through the continuous and categorical variables. Maximum probabilities surpass any lower class probabilities. Continuous variables, Elevation and Distance to MSC, produce negative coefficients as expected. Aspect, on the other hand, contains positive coefficient only for the Southeast direction, where as in other models, this variable takes negative values in most cases.

7.1.17 Predictive Model 17 (PM17)

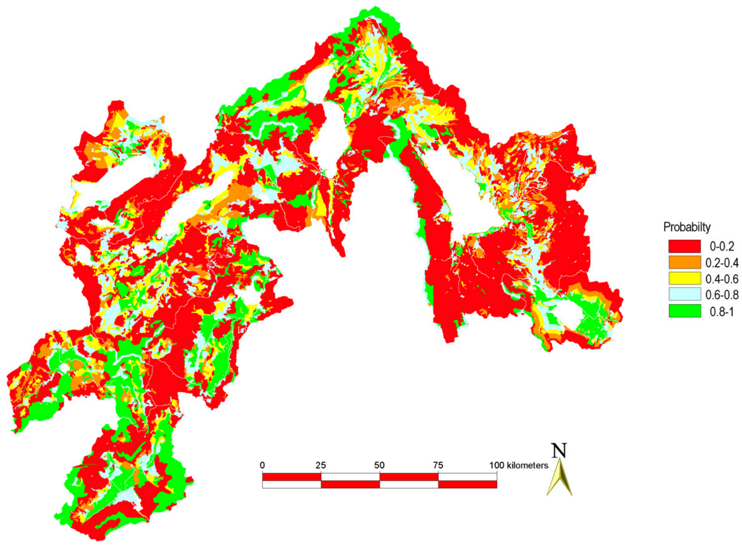


Figure 7.17 Probability surface of PM17.

$\hat{y}$=7.87+(0.002xDtRd)+A

A: Aspect – DtRDG: Distance to Basin Ridge

A(1):-6.778, (2):-8.331, (3):-7.415, (4):-8.745, (5): 0.013, (6):-9.687, (7):-11.104, (8):-6.500

Variables of the model are Distance to Ridges and Aspect. Although this is the case rare maximum variables are aligned at the ridge of the southernmost basin. This can be due to the influence of Aspect values or intercept value or both. The same anomaly also happens at the Northernmost of the study area.

7.1.18 Predictive Model 18 (PM18)

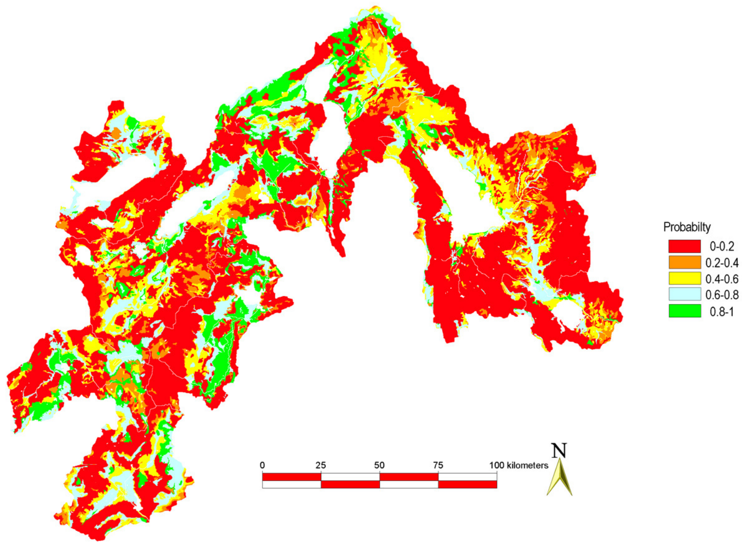


Figure 7.18 Probability surface of PM18.

$\hat{y}$=11.468+(-2.730xE) +A

E : Elevation – A : Aspect

A(1):-7.387, (2):-8.236, (3):-8.661, (4):-8.327, (5):-0.438, (6):-10.210, (7):-10.436, (8):-0.574

Comprised of one continuous and one categorical variable, PM18 produces an easily readable but interesting model. Intercept value is high, and all of the coefficients are negative. It is expected that as elevation increases, the probability of expecting a site decreases. On the other hand, all of the Aspect categories are likely to contribute to 'Non-Site' dependent variable, in relation to Northwest areas, having Aspect coding 9.

7.1.19 Predictive Model 19 (PM19)

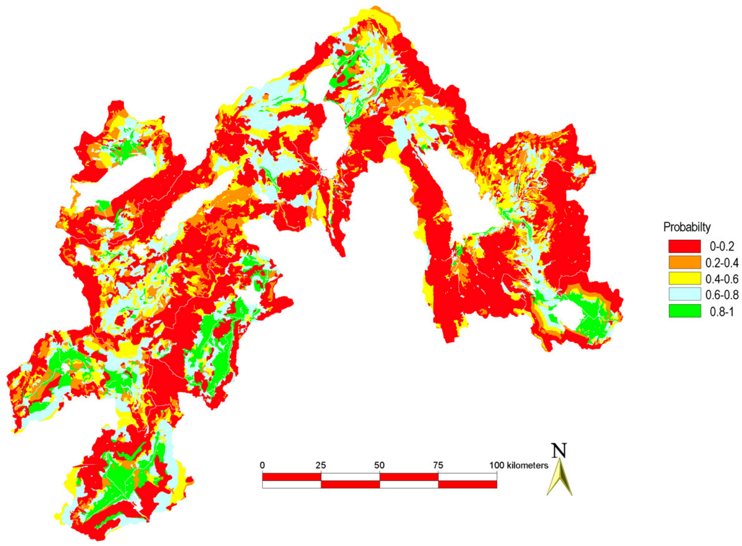


Figure 7.19 Probability surface of PM19.

$\hat{y}$=8.732+(-0.096xDtR)+(0.002xDtRd)+A

DtR: Distance to River – DtRd: Distance to Ridge

A(1):-7.603, (2):-8.623, (3):-7.742, (4):-8.679, (5):-0.055, (6):-9.780, (7):-11.354, (8):-7.370

PM18 is very similar to PM19, in terms of the characteristics of the result, and employment of independent variables. Probability decreases as the pixels get further away from rivers, and probability increases as one move away from basin ridges. As in the previous case, all Aspect directions work in contrast to 'Northwest' direction.

7.1.20 Predictive Model 20 (PM20)

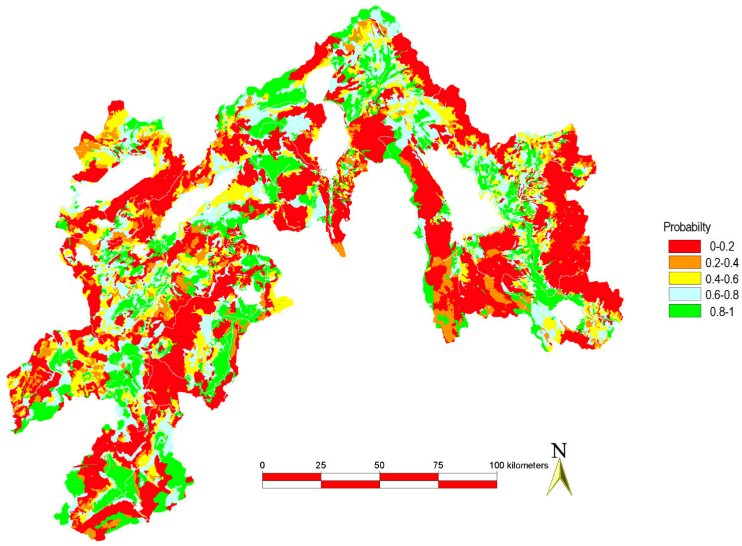


Figure 7.20 Probability surface of PM20.

$\hat{y}$ =15.352+RGH+A

RGH: Roughness – A: Aspect

A(1):-8.942, (2):-8.908, (3):-8.933, (4):-9.617, (5): 0.311, (6):-10.967, (7):-12.607, (8):-7.045
RGH(1):-6.885, (2):-5.548, (3):-4.450

PM20 has produced patches of different probability zones, dominated by maximum and minimum probabilities. Comprised of categorical variables, PM20 has covariates opposing to existence of sites Maximum probability areas then due to increased intercept value of the logistic regression equation.

7.2 Merging Predictive Models

Each predictive model employs different variables and produces different probability surfaces (Table 7.1). Thus there is a need to figure out the final predictive model so that some archaeological inferences can be made.

The first way might to examine individual predictive models visually so that any ‘different’ predictive surface can be omitted before a final decision is made.

Second way might be to use testing samples for those models and according to some criteria decide on the ones to be used for the resulting model. Again the decision on the threshold value of success will be totally subjective and different results will be obtained with different threshold values.

The other basic problem is to decide on how different predictive models will be combined to produce a final model. To give options, an average of all chosen predictive models’ pixel values can be used to produce the final model. Second, minimum of pixel values can be used to construct the resulting model. This way will produce a pessimistic model, and it will increase the gross error. A third option can be using the maximum value of predictive models. This is in contrast will produce an optimistic model with an increased wasteful error. A fourth option might to use random values of different predictive models which are later assigned to resultant model.

There might be other ways to combine different models into one big model, but considering the scope of this study only the operations of minimum, maximum and average taking will be used and the results will be given for comparison.

Then as a strategy, models to be used to construct the final model will be chosen. The later step will to combine with averaging tools. The resulting map will be reclassified into readable probabilities.

Table 7.1 Predictive Model Equations and their variables.

	Elevation	Adjusted Elevation	Slope	Aspect	Adjusted Distance to Ridges	Roughness	Distance to Roughness Junc.	Lithology	Current Land Use	Soil Particle Size	Erosion	Soil Depth	Distance to Major Class	Distance to River	Distance to Lake
PM1		X	X										X		X
PM2		X	X	X					X				X		X
PM3		X		X					X						X
PM4		X						X	X						
PM5		X		X				X	X					X	
PM6				X				X	X			X			
PM7			X	X								X			
PM8			X	X								X		X	X
PM9				X		X			X				X	X	
PM10				X					X			X			
PM11				X				X	X			X			
PM12	X							X							
PM13		X		X					X						
PM14		X	X					X							
PM15		X		X				X			X				
PM16	X			X									X		
PM17				X	X										
PM18	X			X											
PM19				X	X									X	
PM20				X		X									

Without any manipulation, 20 sub-models produce the so-called Predictive Model_All (Figure 7.21). The result looks similar to each of the other sub-model in terms of distribution of probability values.

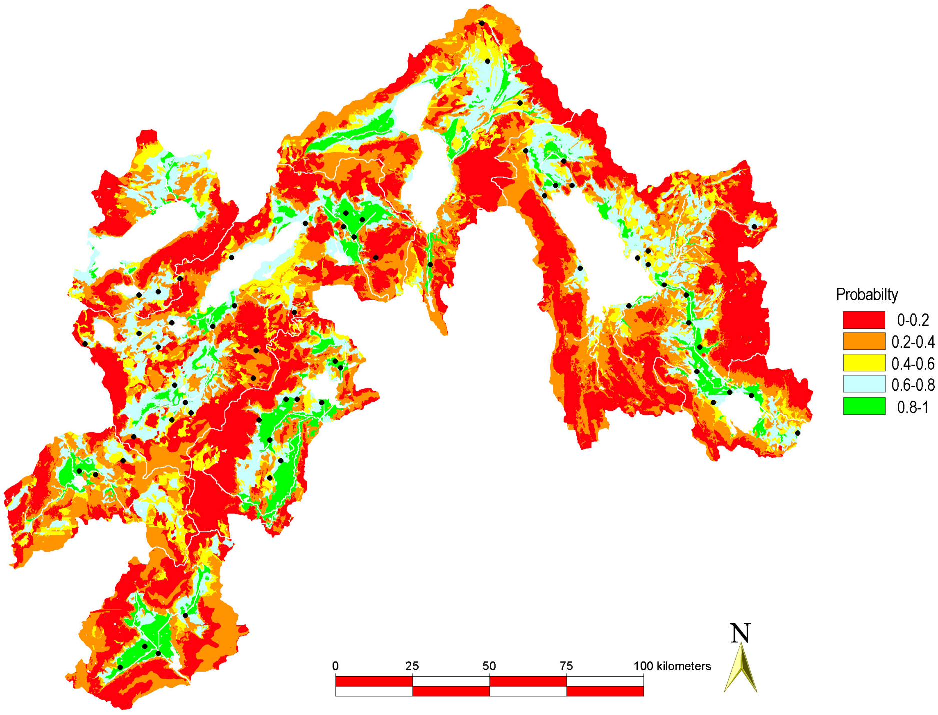


Figure 7.21 Combination of 20 Predictive Models without any filtering.

Predictive Model_All passes the threshold with 4 site pixels falling into the top probability category, and 3 site pixels falling into {0,6-0,8 }probability category (Figure 7.22). As the threshold value, archaeological inventory is used. If the final model includes 70% of the inventory in the 0.8 probability or more then the model is decided to be valid.

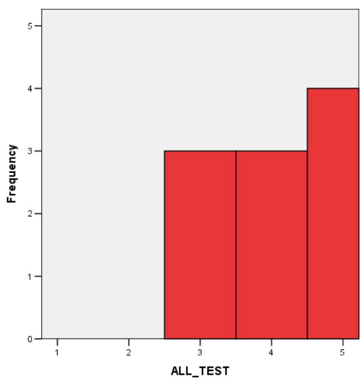


Figure 7.22 Frequencies of Testing Variables according to Predictive Model_All

7.3 Visual Inspection

As a visual inspection, models PM1, PM13, and PM16 are withheld from the final model. PM1 and PM16 are significant with their extensive distribution of maximum probabilities. PM13, on the other hand, produced so many minimum probability values so that it definitely ignores some possible site locations.

After removing those concerned sub-models, the remaining ones are combined by taking the averages of each pixel. The resulting model is given in Figure 7.23.

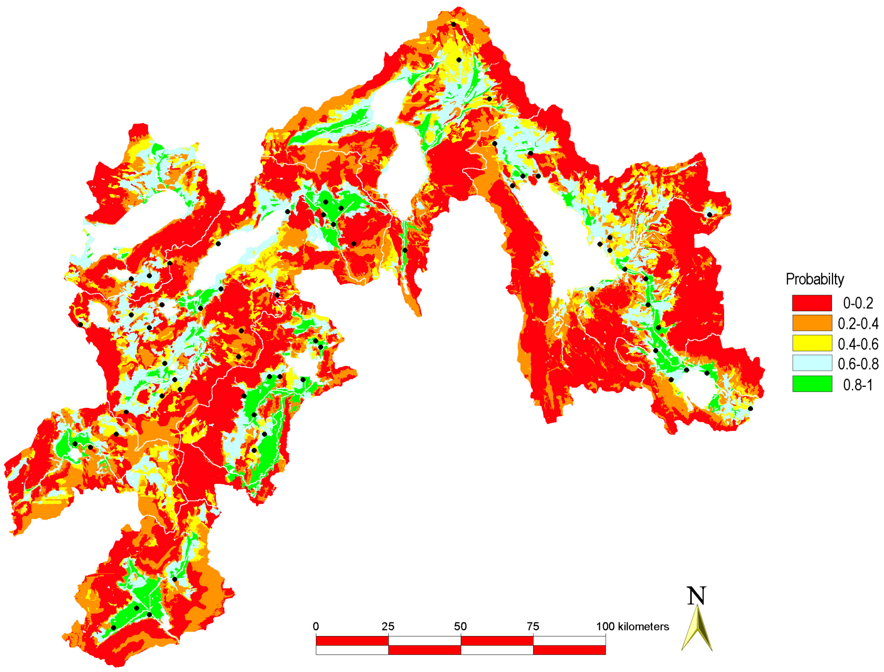


Figure 7.23 Predictive Model_Visual is obtained after screening PM1, PM13 and PM16.

Predictive Model_Visual reveals 3 site pixels at top probability category, and 4 site pixels at {0,6-0,8 }probability category (Figure 7.24).

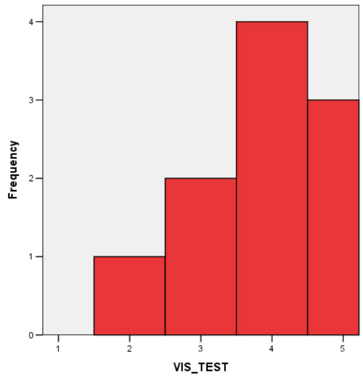


Figure 7.24 Frequencies of testing variables for Predictive Model_Visual.

7.4 Threshold Inspection

Testing samples are used as decision tools. The threshold value is the same with above, where if 70% of the testing sample falls into 80% probability or more then sub-model is included into the final model

Table 7.2 Observed frequencies of testing sample for each predictive model falling into {0-0.2}, {0.2-0.4}, {0.4-0.6}, {0.6-0.8}, {0.8-1.0} probability categories. The success rate of the model is given with percentages for each model. Also, probability categories are depicted as integers where 1 stands for {0-0.2}, 2 for {0.2-0.4}, 3 for {0.4-0.6}, 4 for {0.6-0.8}, and 5 for {0.8-1.00}.

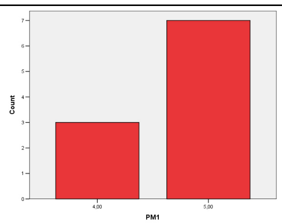
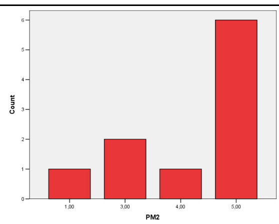
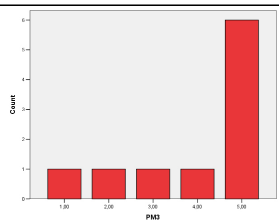
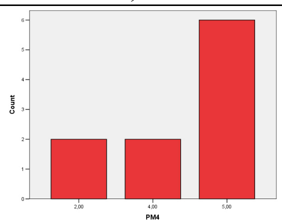
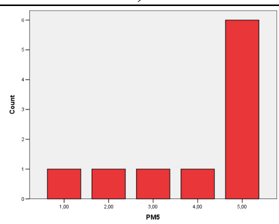
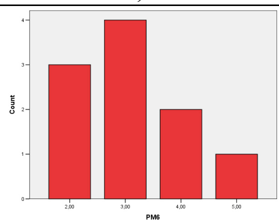
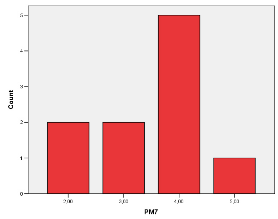
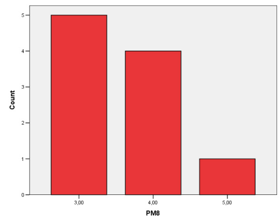
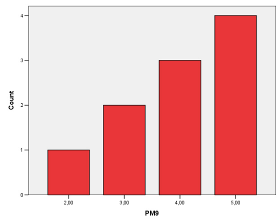
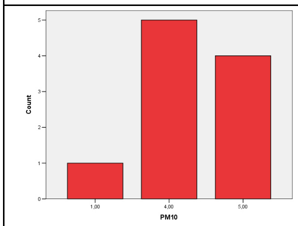
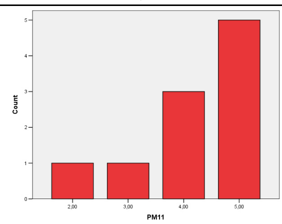
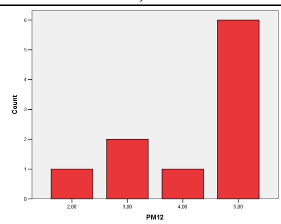
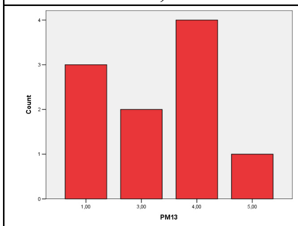
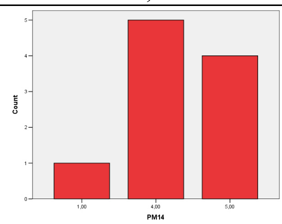
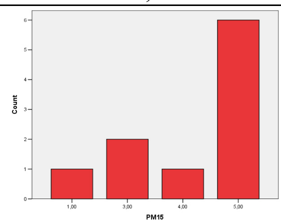
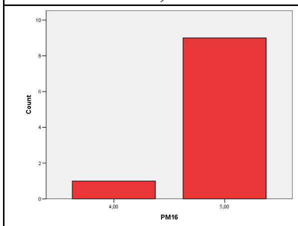
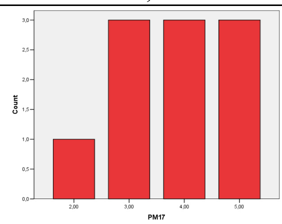
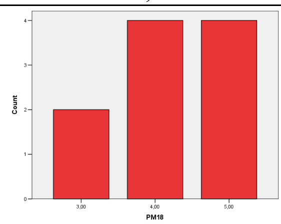
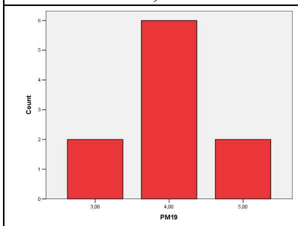
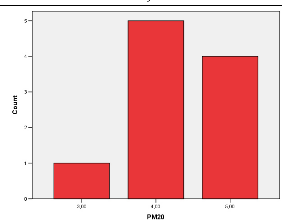
 PM1; 100%	 PM2; 70%	 PM3; 70%
 PM4; 80%	 PM5; 70%	 PM6; 30%

Table 7. 2 (Continued)

 PM7; 60%	 PM8; 50%	 PM9; 70%
 PM10; 90%	 PM11; 80%	 PM12; 70%
 PM13; 50%	 PM14; 90%	 PM15; 70%
 PM16; 100%	 PM17; 60%	 PM18; 80%
 PM19; 80%	 PM20; 90%	

Then according to the given histograms, PM6, PM7, PM8, PM13 and PM17 are excluded from the model since they do not satisfy the criterion. The resulting predictive model, called Predictive Model_Threshold is given in Table 7.2.

After obtaining the model, a testing sample is used to evaluate the efficiency of the model. 5 site pixel values fall into the top probability category, and 3 site pixels fall into the {0,6-0,8 }probability category, which makes 80% satisfaction according to the threshold value (Figure 7.26). This approach then is the one which produces the highest percentage. On the other hand, due to the small size of testing sample, this comparison is questionable.

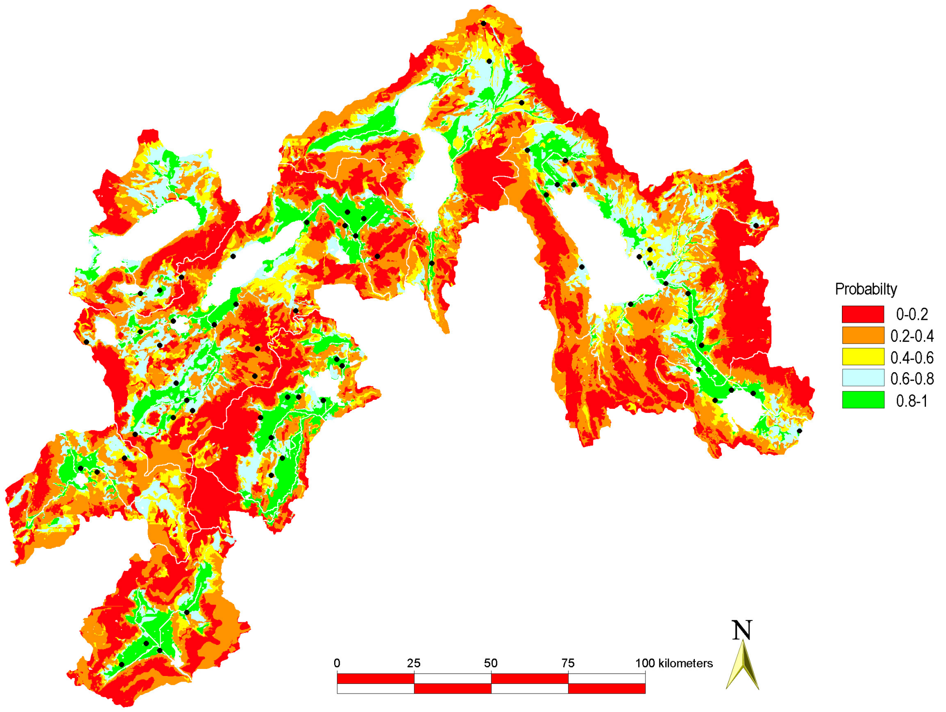


Figure 7.25 Predictive Model_Threshold is obtained after screening PM6, PM7, PM8, PM13 and PM17.

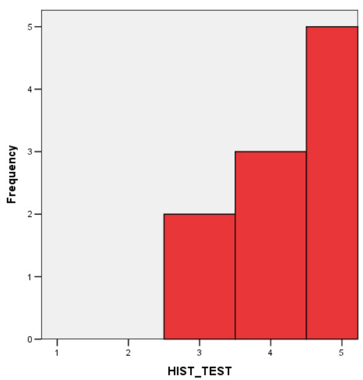


Figure 7.26 Frequencies of testing variables for Predictive Model_Threshold.

7.5 Model Diagnostics

Multicollinearity is simply the correlation between the independent variables. In other words, it is the representation, or quantification of the same happenstance with different variables. The problematic variables might have different data scales as well. The problem is evident when the model p_value is low, while contributing independent variable p_values are high.

According to Field (2000), and Neter et al.(1996), if VIF (Variance of Inflation Value) is greater than 10, or if the average of VIF values of variables considerably exceeds 1, then there is a serious multicollinearity problem.

SPSS 13.0 produces VIF and tolerance values for particular regression models. As a general rule, a model will be stated to have multicollinearity problem if tolerance value is less than 0.1 and VIF is greater than 10.

Calculated VIF values and tolerance levels (APPENDIX-E) reveal that for any predictive model constructed, there is no serious multicollinearity problem, whereas average of VIF values for each predictive model slightly passes the value of 1, but never reaches to 2.

CHAPTER 8

DETERMINANTS of SETTLEMENT LOCATIONS

This chapter aims to discuss the determinants of settlement locations from an archaeological and ecological perspective. The arguments then can be compared with the results obtained from the predictive model. Moreover, the discussion below is also used to obtain a general understanding of settlement locations as well as to determine the limits of the study area both in space and time.

In order to understand the patterning, factors correlated with the pattern should be highlighted, and 'articulation of factors' should also be examined. Then, determinants of settlement patterns are the 'classes of factors that interact with each other to produce the spatial configurations of a social group' (Trigger, 1968; 53). Independent of a specific temporal frame, determinants can be given as trade (medieval cities, Assyrian tradeposts in Anatolia, Greek colonies around the Black Sea (Trigger, 1968; 68)), warfare, and religion. Within a given geographic area, change in determinants can also be examined.

Determinants can be more simple then expected as well. The agriculturalists, for instance, were looking for water as well as Paleolithic and Mesolithic hunters, but they were having another concern of arable land, for instance in the Berbati Valley. Later in the same area, a prosperous farming economy led the people to move on plateaus. With the increased population, land clearance occurred on even bedrock slopes, then conditions and thus determinants have changed again (Zangger, 1992; 144).

As the complexity of an entire society increases, or at least as level of relations increases, then pattern can not solely be determined by subsistence factors anymore, but some advanced economical, political, and religious factors should be considered (Trigger, 1968; 67).

Effort should be made to define and use related variables with the locations of settlements. On the other hand, such an endeavor is not easy, and in fact this is the basic fallacy of the modeling studies in archaeology, but any effort will shed light to the question.

There are some studies which explicitly define the determinants of site location, whereas some other studies do not look for an answer of such a question, but still implicitly defined in the study. Zangger (1992) reveals determinants as fertile and arable soil, water, shelter from wind, protection from sun and exposure to it, availability of building stones, and strategic control from elevated heights, remoteness and inaccessibility. Again, they are not global rules, and cannot be applied to the site nearby, but at least they are helpful to construct a predictive model in a GIS. It is known that Suberde is characterized by its huge amount of animal bones (Bordaz, 1973). If such a huge consumption is detected, then it might be asserted that proximity to animal migrating zones is a determinant for the site.

In the Beyşehir-Suğla Area, Solecki (1964; 131) observed at least 7 different categories of human occupation.

1) Mounds with no evidence of underlying stone outcrops.
2) Natural summits with human occupation
3) Tumuli
4) Monuments and architectural works
5) Caves
6) Rock shelters
7) Open sites

Although there are some, like caves and rock shelters, which are determined by the existence of distinct clear geographical features (namely rocks and caves), there should be some complex motivations for some others to be located. Naturally not all of the caves and rock formations are/were suitable for inhabitation, and not all of the suitable ones were/are inhabited, but they are all matters of pragmatism. Similar discussion can also be made about natural summits to some extent. On the other hand, the selection of the initial location of a tumulus, mound or a monument or an architectural work is a response to a decision making process. Monumentality will be obtained if it is visible from many places, and if it is large enough then it will be visible. Both concerns include some criteria, such as being on a peak, or being close to raw materials for the ease of transportation, at least for the time period where transportation was a matter of concern. In the same way, locations of tumuli were selected by some rule, and might produce a non-random pattern, but maybe the most complex system is the one for the locations of mounds and flat settlements since they are the places that are chosen to be lived on and at least some criteria should be met for vital purposes.

Mounds, the source of dependent variables in the analysis, are considered to be permanent places which are occupied for a period of (from centuries to millennium or more) time (Mellart, 1972; 280) and exploiting the natural environment to some extent. A broad generalization reveals that the existence of a stream or a lake and a good arable land is common to most of the Neolithic sites, with exceptions (Mellart, 1972; 280). Suberde/Görüklük Tepe, for instance, located in a strategic position that it was possible to view both the lake and the plain, which was the reason for the hunters to be settled (Solecki, 1964; 134). According to Kökten (1952; 194), mounds are generally located at the places where modern agriculture is performed, near the lakes and streams, and on islands and at the top of the hills, which is also evident in the model. Although current lake levels in the study area hinders the situation it is highly possible that some sites were located on islands. Moreover they tend to be in water heap due to karst topography, or in shrubberies near sea or forests. On the other hand, extents of forests for the time in concern are questionable so that both simple observations and complicated models can hardly satisfy this judgment. Similar observation was made by Tringham (1971, 91) about the locations of settlements of earliest agriculturalists as being on the upper terraces of valleys, or on the edges of plateaus. Another derivation was obtained by Barker (1975, 100) on Neolithic sites of Yugoslavia as being close to major river courses, and an escape from loess and sandy soil. Allan (1972; 211) states two features are significant for earliest agricultural settlements: small size, and the community of ecology of the sites around. Also, according to the surveys made in the study area generalizations on Neolithic and Chalcolithic sites can be made (Yakar, 1994; 287):

i) Small to medium sized permanently settled unwalled settlements with individual houses, as Can Hasan III, Erbaba

ii) Small sized unwalled settlements inhabited seasonally, as Suberde

iii) Small to medium sized with no indication of 'social stratification' settlements, as Kuruçay, Hacılar II

iv) Medium to large sized, walled or unwalled settlements with some indication of 'social stratification' with some special buildings, as Hacılar I, Çatal Höyük

Those generalizations are questionable, but important in terms of showing that it is possible to create more complex models that can test deductive ideas.

Independent of any study area and independent of the researcher's stance, five basic elements can be set as the result of any generalization for agriculturalists (Chisholm, 1968; 102):

1) arable land
2) fuel
3) building material
4) grazing land
5) water

On the other hand, constructing the model merely on those elements will not be false, but lacking. Moreover those elements are not stable, and in a very simplistic manner heavy use and climatic change can affect one or more of them, and in a chain reaction exploitation and climate change reduce fertility, increase pollution, and even collapse well based settlements (Ponting, 1991). Thus the exploitable land should be left, or has to be waited for regeneration.

Four types of environment can be defined for food production, thus for the agriculturalists (Meggers, 1954; 803).

1) Areas of no agricultural potential.

A subsistence economy based on hunting-gathering, and fishing will support only small group of people in a seasonal basis. The environment permit the basics and little more in the area, but it do not mean that the specified area is totally is free of sites.

2) Areas of limited agricultural potential.

In the area, more time is available for manipulating raw materials and process of manufacture, which might introduce the pottery and loom weaving.

3) Areas of increasable (improvable) agricultural potential.
4) Areas of unlimited agricultural potential.

The potential is an implicit term so that it should be clearly defined, and even more substantial research should be made for past *potentials* of particular areas. In the model, one of the important environmental variables is the current land use. Thus, even a tentative reconstruction can reveal some for defining the patterning of the sites in the area. In any case, existing sites can be considered falling one of the above categories. In fact, similar procedures are used in modern geography to divide the land according to some criteria. The criterion in this case is the potential for agriculture, but again a researcher should keep in mind that parameters determining the potential are not static, but dynamic. A similar categorization was made for the provinces of Antalya, Burdur, and Isparta, an area of 37,000 square-km, and four ecological zones were recognized; namely, the Western Plateau, the Lakes zone, the Coastal zone, and the Mountain zone. Based on hypothetical parameters it was estimated that the total area would be able to support an overall population density of twelve per square-km, but the specific numbers were 24.3, 16.5, 11.3, and 1.8 per square-km accordingly (Allan, 1972; 215). Then those numbers can be used in proportions to obtain a hypothetically estimated number of sites in the area. Definitely, it will not be exactly true, but it will be informative for the potentials of zones.

Food production is depending on environmental conditions, and at the same time production determined the density, population of a region, and to some extent socio-politics of the region. The degree of relation is the concern and this degree is the one to be predicted, tested and judged (Meggers, 1954; 802).

But the general picture would not have been so straightforward. For instance, backswamp soils were very close to the settlements of Çatal Höyük, and Nea Nikomedeia despite the fact that that type of soils were marshy, and the marsh was probably a source of malaria. Maybe it was the stock, but not the crop were the factor for a decision of site location, where domesticated fauna was the interest (Barker, 1975; 90).

As a determinant, land use is important but not enough by itself. Another concern should be the locations of trade routes and their relations with the settlements. On the other hand, it will not be wrong to assume that the locations of settlements had created the network, and not the other way around. Not surprisingly, the routes can be natural and the locations can be determined by considering the use of natural paths, such as Calycadnus Valley which leads from the Konya Plain to the Mediterranean Coast. This issue is again related with the problem of transhumance and determined by the cost of being between the sites, and there is evidence of such transhumance as early as the second half of the ninth millennium B.C. at Mureybit and Jericho (Mellart, 1972; 280).

The effect of environment, thus the effect of determinants of settlement locations on the construction and evolution of a culture is speculative (Meggers, 1954). On the other hand, if a population is moved from one place to another for any reason the lifestyle of that society will adapt or change. Moreover, as Sollars states (2005; 258)

> … to move from the plains to settle in a mountainous region, their lifestyle would, of necessity, come to reflect their location, rather than the influence their choice of it. Their choice of location would reflect the broader social situation of conflict.

CHAPTER 9

CONCLUSION

In the study area, 64 settlement mounds are transformed into 95 site pixels. Transformation is based on available information on sizes of mounds. 85 of the pixels are used to train model, and 10 of them are withheld from the analysis to test the model. Then a total of 15 independent variables and 1 dependent variable are put to the analysis, and logistic regression equations are obtained. Considering the immense size of the study area, and limited number of positive responses for the model, more than one regression equation is run, and they are used to construct different predictive models. At the end, with given criteria the resulting predictive model is obtained.

There are two major outcomes of the study. First, individual environmental variable and site location relations are revealed via graphs and descriptive statistics. Second, by obtaining a predictive surface, possible site locations are detected so that those areas can be used in future studies.

To begin with, in terms of topography, it can be claimed that sites tend to be located around 1400m. After making an adjustment with respect to lake levels, it had been found that sites tend to gather around lakes, and in few cases they are located far away from lakes. Also, sites are located at flat areas, or to explicitly define, at areas having less than 3 degrees of slope. There is an interesting recognition on Aspect variable. Although sites are expected to be located at the South faces, in the study area sites are located at non-South faces. This fact was also significant at regression equation so that Aspect is used in most of the predictive models. Adjusted Distance to Ridges revealed that sites are close to ridges in terms of unit distance. This is both due to the fact that locations of lakes, in which they are close to ridges, and distribution of pixel values in the overall study area.

According to the measure created for Roughness, sites avoid very smooth areas, and people had looked for some undulation. On the other hand, this undulation should be noted as a proxy, where it can be claimed that such areas are suitable for rich flora and fauna, but after some degree of undulation it is not feasible to settle down.

In terms of lithology, ancient settlers really favored quaternary alluvium, where rock categories are evenly distributed in the study area except for volcanic class.

Modern agricultural fields contain many of the existing settlement mounds. This was also stated by Kökten in the year of 1952. Although such observations are clear for some, there is still a need for a formalization of whole understanding to obtain a general perspective.

Very few sites are found at areas where very heavy erosion is observed. This can be related with survival of sites, and thus can be used for future field surveys. Most of the sites are located on alluvium. This is another indicator of relation of locations of sites and where modern agriculture is practiced. Moreover, such an assertion can be used to put the relation of rivers, or soils around rivers and settlement locations.

Sites are mostly located within 5km range of river courses, but as stated earlier, high order rivers are put to the analysis, and so domestic use of water is omitted. Distance to Lakes produces an interesting result. There are three groups of sites in relation to lakes. One group with around 40% is found to be located very close to lakes. Another group is found to be located at a distance of 20km, and the last group is having an average distance of 40km. Those distances are significant when adjacent distances are considered. That is, those distance values are making peaks when compared with other values which are close to those distances. Considering again the size of the study area, those results divulge general patterning of sites rather than specific responses of site locations to their immediate environments.
Second outcome of the study was to obtain particular places that deserve more archaeological investigation (Figure 9.1).

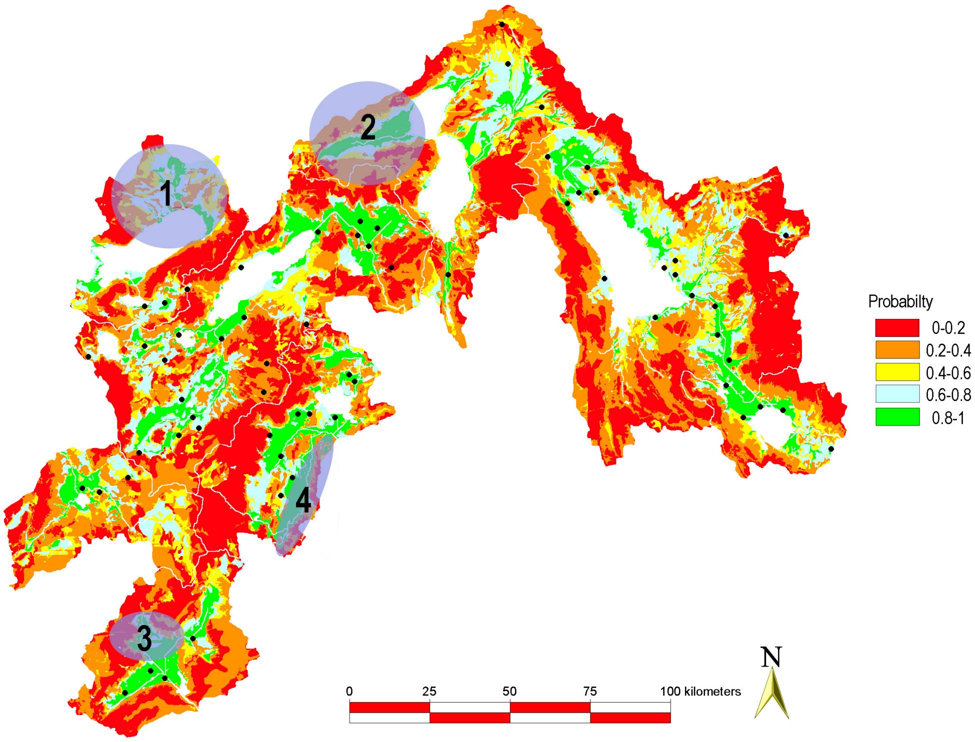


Figure 9.1 Predictive Model_Threshold with particular areas of interest.

Areas marked as Number 1 and Number 2 contains no site in the archaeological inventory, but claimed to be high potential areas, which means it is highly possible to find other sites according to the model developed. As revealed in layers, those specific zones are relatively flat when compared to immediate vicinities, and close to river systems. Areas marked as Number 3 and Number 4 also shows high potential areas. Although some sites exist in their close district, more research can be conducted in those areas since they are still broad zones.

Model suffers from a number of limitations. First of all, the model is heavily influenced from the existing archaeological inventory. The bias in the inventory is directly reflected into the model, such as the surveyors prejudice for data collection, mistakes in the existing coordinates or mistakes for dating sites. Secondly, data obtained from various sources have various fallacies. For instance, soil data obtained from Directorate of Rural Services have missing values for particular places, thus a random filling for those missing values had to be employed. Moreover, some attributes are not totally reliable at all, such as erosion. Topographical data obtained from SRTM imagery have some gaps so that a built-in algorithm is used to fill those gaps in the elevation data. Since elevation data is the primary coverage for other data types, the errors in this layer is automatically introduced to the secondary coverage. Although water is claimed to be very important for the sites in concern, only major river courses are used in the model since it is nearly impossible to reconstruct the locations of springs and streams carrying water for that particular period of time. Not all of the ancient lake levels are available so that only existing ones are used. This situation introduced some bias to the model since relative locations of sites in concern are not totally depicted in the model.

There are also some other implicitly defined limitations in the model stemming from the assumptions used to create the model. Definitely, projection of current data to the past is not exactly true. For example, modern land use can hardly be utilized instead of ancient land use, and any reconstruction to do this is beyond the scope of this study. The second assumption used to construct the model is about human locational behavior stating that activities are not arbitrary but caused. This is based on energy minimizing and profit maximizing ideology, whereas such a terminology cannot be entirely true for all human history.

In any case, efforts for such modeling studies in any area for any time period are an asset for the discipline and should not be underestimated. The collaboration of computational power with the well defined questions equipped with proper data can reveal more than expected.

REFERENCES

Allan, W.
1972. *Proceedings of a meeting of the Research Seminar in Archaeology and Related Subjects held at the Institute of Archaeology*, Duckworth: London University.

Allison, P.D.
2000. *Logistic Regression Using the SAS System, Theory and Applications*, North Carolina: SAS Institute Inc.

Altınkale, S.
2001. *Eğirdir ve Burdur Göllerinin Hidrojeokimyasal ve İzotop Jeokimyasal Karşılaştırılması*, Isparta: Jeoloji Mühendisliği AnaBilimDalı Yüksek Lisans Tezi.

Altschul, J.H.
1989. *Man, Models and Management: An Overview of the Archaeology of the Arizona Strip and the Management of its Cultural Resources*, Report prepared for USDA Forest Service, Denver: USDI Bureau of Land Management.
1990. "Red flag models: the use of modeling in management contexts." In *Interpreting Space: GIS and Archaeology*, edited by K.M. Allen, S.W. Green, S.W., and E. Zubrow, 226-238. London: Taylor & Francis Ltd.

Field, A.
2000. *Discovering Statistics using SPSS for Windows: Advanced Techniques for the Beginner,* London: Thousand Oak, Sage Publications.

Baird, D.
2000. "Konya Plain Survey" *Anatolian Archaeology* 6:15.

Barker, G.
1975. *Early Neolithic Land Use in Yugoslavia*, Proceedings of the Prehistoric Society (PPS) 41: 85-104.

Barker, G.W. et al.
1999. "Environment and Land Use in the Wadi Faynan, Southern Jordan: The Third Season of Geoarchaeology and Landscape Archaeology 1998" *Levant* 31: 255-292.

Bintliff, J.L.
1982. "Palaeoclimatic Modeling of Environmental Changes in the East Mediterranean Region since the Last Glaciation." In *Palaeoclimates, Paleoenvironments and Human Communities in the Eastern Mediterranean Region in Later Prehistory,* edited by J.L. Bintliff, and W. Van Zeist. *BAR Int. Series*.

Blalock,
1979. *Social Statistics,* Mc Graw-Hill Series in Sociology Aukland: Mc Graw-Hill.

Bordaz, J.
1965. "Suberde Excavations 1964" *Anatolian Studies* 15:30-32.
1966. "Suberde 1965" *Anatolian Studies* 16:32-33.
1968. "The Suberde Excavation, Southwestern Turkey: An Interim Report" *TAD* 17(2):43-71.
1973. "Current Research in the Neolithic of South Central Turkey: Suberde, Erbaba and Their Chronological Implications" *American Journal of Archaeology, Vol.77, No.3*: 282-288.

Bordaz, J. & L.L. Alpers - Bordaz
1976. "Erbaba Excavations 1974" *TAD, 23(2)*: 39-43
1982. "Erbaba. The 1977 and 1978 Season in Perspective" *TAD,* 26(1): 85-93

Brandt, R., B.J. Groenewoudt & K. Kvamme
1992. "An Experiment in Archaeological Site Location: Modeling in Netherlands Using GIS Techniques", *World Archaeology,* Vol.24, No.2: 268-282

Butzer, K.W.
1982. *Archaeology as Human Ecology: Method and Theory for a Contextual Approach,* Cambridge.
1964. *Environment and Archaeology: an Introduction to Pleistocene Geography,* Chicago: Aldine Pub. Co.

Chisholm, M.
1968. *Rural Settlement and Land Use*, London: Hutchinson University Library.

Church, B., J. Brandon & G.R. Burgett
2000. "GIS Applications in Archaeology: Method in Search of Theory." In *Practical Applications for Archaeologist: A Predictive Modeling Kit,* edited by K.L. Westcott, and R.J. Brandon, London: Taylor & Francis

Cohen, H.R. & O. Erol
1969. "Aspects of Palaeogeography of Central Anatolia" The Geographical Journal, *The Royal Geographic Society* 135:388-389

Cohen, H.R.
1970. "The Palaeoecology of South Central Anatolia at the end of the Pleistocene and Beginning of the Holocene" *Anatolian Studies* 20: 119-137.

Dalla Bona, L
1997. *Cultural Heritage Resource Predictive Modeling Project* Ontario Ministry of Natural Resources.

Duncan, R.B. & K.A. Beckman
2000. "The Application of GIS Predictive Site Location Models with Pennsylvania and West Virginia." In *Practical Applications for Archaeologist: A Predictive Modeling Kit,* edited by K.L. Westcott, and R.J. Brandon, London: Taylor & Francis

Duru, R.
1994. *Kuruçay Höyük I 1987-1988 Kazılarının Sonuçları , Neolitik ve Erken Kalkolitik Çağ Yerleşmeleri*, Türk Tarih Kurumu Yayınları V. Dizi, Sayı 44.
1995. "Höyücek Kazıları 1991/1992" *Belleten* 59: 447-476.
1996. *Göller Bölgesi'nde Neolitik Köyden Kasabaya Geçiş, Tarihten Günümüze Anadolu'da Konut ve Yerleşme,* Türkiye Ekonomik ve Toplumsal Tarih Vakfı Yayınları, Ofset Yapımevi, 49-59.
1999. "The Neolithic of the Lake District." In *Neolithic in Turkey*, edited by N.Başgelen, and M. Özdoğan, 165-191. İstanbul: Arkeoloji ve Sanat Yayınları.

Eastwood, W.J., N. Roberts & H.F. Lamb
1998. "Palaeoecological and Archaeological Evidence for Human Occupance in SW Turkey: The Beyşehir Occupation Phase" *Anatolian Studies* 48: 69-86.

Eastwood, W., N. Roberts, H. Lamb & J. Tibby
1999. "Holocene environmental change in southwest Turkey: a palaeoecological record of lake and catchment-related changes" *Quaternary Science Reviews* 18: 671–696.

Ebert, J
2000. "The State of the Art in "Inductive" Predictive Modelling: Seven Big Mistakes (and Lots of Smaller Ones In *Practical Applications for Archaeologist: A Predictive Modeling Kit,* edited by K.L. Westcott, and R.J. Brandon, London: Taylor & Francis

Ejstrud, B.
2002. "Indicative Models in Landscape Management" (last retrieved 26 July) web.sdu.dk/ejstrud/forskning/GIS/ejstrud_wunsdorf_2001.pdf

Erinç, S.
1978. "Changes in the Physical Environment in Turkey since the End of the Last Glacial" In *The Environmental History of the Near and Middle East since the Last Ice Age*, edited by W.C. Brice, 87–110. London: Academic Press.

Erol, O.
1971. "Konya, Tuzgölü, Burdur Havzaları'ndaki Pluvial Göllerin Çekilme Safhalarının Jeomorfolojik Delilleri" *Coğrafya Araştırmaları Dergisi* 3/4:13-52.
1978. "The Quaternary History of the Lake Basins of Central and Southern Anatolia.", In *The Environmental History of the Near and Middle East since the Last Ice Age*, edited by W.C. Brice, 111-139. London: Academic Press.

Farrand, W.R.
1964. “Geology and Physiography of the Beyşehir -Suğla Depression ,Western Taurus Lake District, Turkey” *Türk Arkeoloji Dergisi*, *Sayı* 13/1: 149-154.

Fontugne, M., C. Kuzucuoğlu, M. Karabıyıkoğlu, C. Hatte & J.F. Pastre
1999. **“**From Pleniglacial to Holocene: A 14C chronostratigraphy of Environmental Changes in the Konya Basin” *Quaternary Science Reviews* 18: 573-592.

French, D.H.
1970. “Notes on Site Distribution in the Çumra Area” *Anatolian Studies* 20: 139-148.

Fry, G.L.A., B. Skar, G. Jerpasen, V. Bakkestuen, & L. Erikstad
2004. “Locating archaeological sites in the landscape: a hierarchical approach based on landscape indicators” *Landscape and Urban Planning* 67: 97–107.

Gaffney, V. and M. vanLeusen.
1995. “Postscript GIS, environmental determinism and archaeology: a parallel text” In *Archaeology and Geographical Information Systems: A European Perspective,* edited by G. Lock, and Z. Stancic, London: Taylor and Francis

Harmankaya, S, Tanındı, O., Özbaşaran, M.
1997. *TAY-Neolitik*, İstanbul: Ege Yayınları.

Hayden, B.
2003. “Were Luxury Foods the First Domesticates? Ethnoarchaeological Perspectives from Southeast Asia” *World* Archaeology, Vol.34(3): 458-469.

Hodder, I.
1990. *The Domestication of Europe*, Oxford: Blackwell.

Hole, F.
2004. “Nature’s Role in the Beginning of the Agriculture” *TUBA-AR* 7, Tubitak

Judge W.J. & L. Sebastian
1988. *Quantifying the present and predicting the past: Theory, method, and application of archaeological predictive modeling*, U.S. Government Printing Office, Washington

Japan Aerospace Exploration Agency
2006. What is SRTM?, (last retrieved 26 July)
http://iss.sfo.jaxa.jp/shuttle/flight/sts99/mis_srtm_e.html#souchi, 2006 (Figure 6.2).

Kincaid, C.
1988. “Predictive Modeling and its Relationship to Cultural Resource Management Applications.” In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling,* edited by W.J. Judge, and L. Sebastian., 549–569. U.S. Government Printing Office, Washington

Kohler, T. A. & S. Parker
1986. “Predictive Models for Archaeological Resource Location.” In *Advances in Archaeological Method and Theory* Vol. 9, edited by M.B. Schiffer. 397-452

Kohler, T.
1988. “Predictive Locational Modeling: History and Current Practice.” In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling,* edited by W.J. Judge, and L. Sebastian. U.S. Government Printing Office, Washington

Kökten, I.K.
1952. “Anadolu' da Prehistoric Yerlerin Dağılışı Üzerine Bir Araştırma” *DTCFD* 10:167-207.

Kuzucuoğlu, K.
2001. Environmental setting and evolution from the 9th to the 5th millennium cal BC in central Anatolia: An introduction to the study of relations between environmental conditions and the development of human societies (last retrieved 26 July 2006) http://www.canew.org/leckuzubox.html

Kvamme, K.
1988. "Development and testing of quantitative models." In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling,* edited by W.J. Judge, and L. Sebastian., 325-428. U.S. Government Printing Office, Washington.

Lahn,E.
1948. *Türkiye Gölleri Jeolojisi ve Jeomorfolojisi Hakkında Etüt,* Ankara: MTA.

Leveau, P., F. Trement, K. Walsh, & G. Barker
1999. *The Archaeology of Mediterranean Landscapes 2: Environmental Reconstruction in Mediterranean Landscape Archaeology,* Oxbow Books: Oxford

Llobera, M.
1996. “Exploring the Topography of Mind: GIS, Social Space and Archaeology” *Antiquity* 70:612-622.

Lobeck, A.K.
1939. *Geomorphology: An Introduction to the Study of the Landscapes*, New York and London: McGraw-Hill Book Company.

McCorriston, J. & F. Hole
1991. “The Ecology of Seasonal Stress and the Origins of Agriculture in the Near East” *American Anthropologist,* Vol.93, No.1, 46-69.

Meggers, B.J.
1954. “Environmental Limitation on the Development of the Culture.” *American Anthropologist*, New Series, Vol.56, No.5, Part 1, 801-824.

Mellaart, J.
1954. “Preliminary Report on a Survey of Preclassical Remains of Southern Turkey.” *Anatolian Studies* IV, 188.
1961. “Early Cultures of South Anatolian Plateau.” *Anatolian Studies* 11,159-184.
1970. *Excavations at Hacılar I,* Edinburgh: University Press.
1972. *Anatolian Neolithic Settlement Patterns, Man, Settlement, and Urbanism*, Proceedings of a meeting of the Research Seminar in Archaeology and Related Subjects held at the Institute of Archaeology, London University, Duckworth.

Moon, Heather
1993. “Archaeological Predictive Modelling: An Assesment.” *The Earth Sciences Task Force Resources Inventory 3.*

Muller-Beck, H.
1979. *Prehistoric Swiss Lake Dwellers, Readings from Scientific American Hunters, Farmers and Civilisations, Old World Archaeology*, San Francisco: W.H. Freeman and Company.

Neter, J., M.H. Kutner, C.J. Nachtsheim, & W. Wasserman
1996. *Applied Linear Statistical Models,* WCB/MCGraw-Hill, Fourth Edition.

Özdoğan, M.
2002. “Devrimlerin Atası Neolitik Çağ”, *ArkeoAtlas*, DBR Dergi Yayıncılık ve Pazarlama A.Ş.

Özsait, M.
1986. “Pisidya Bölgesi'nde Yeni Prehistorik İskan Yerleri (III)” *Anadolu Araştırmaları* 10:73-99.

Perkins, P.
2000. “A GIS Investigation of site Location and Landscape relationships in the Albegna Valley, Tuscany” In *CAA96 BAR International Series, Archaeopress,* edited by K. Lockyear

Ponting, C ,
1991. *Dünyanın Yeşil Tarihi: Çevre ve Uytgarlıkların Çöküşü*,
Sabancı Üniversitesi , Çeviren Ayşe Başcı-Sander

Raikes, R.
1967. *Water, Weather and Prehistory*, New York: Humanities Press Inc.

Richerson, P.J., Boyd, R. and Bettinger, R.L.
2001. “Was Agriculture Impossible During the Pleistocene but Mandatory During the Holocene? A Climate Change Hypothesis” *American Antiquity* 66(3), 387-411

Rodriguez et al.
2005. “An Assessment of the SRTM Topographic Products”
http://www2.jpl.nasa.gov/srtm/SRTM_D31639.pdf

Roberts, N., M. Meadows & J. Dodson
2001. “The history of mediterranean-type environments: climate, culture and landscape” *Holocene* 11, 631–634.

Saraçoğlu, H.
1962. *Mevki Sınırla Yüzey Şekilleri, Denizler, İklim Bitki Örtüsü, Akarsular ve Göller*, Cilt II, İstanbul.

Shermer, S. J. & J. A. Tiffany
1985. “Environmental Variables as Factors in Site Location: An Example from the Upper Midwest” *Midcontinental Journal of Archaeology* 10(2):215-240.

Singh, P.
1974. *Neolithic Cultures of Western Asia*, London and New York: Seminar Press.

Solecki, R.S.
1964. “An Archaeological Reconnaisance in the Beyşehir-Suğla Area of Southwestern Turkey” *Türk Arkeoloji Dergisi*, 13(1), 129-148

Sollars, L.
2005. “Settlement in the Prehistoric Mediterranean” In *The Archaeology of Mediterranean Prehistory*”, edited by E. Blake and, A. B. Knapp, 253-269 Oxford: Blackwell Publishing.

Tabachnick, B.G. & L. S. Fidell
1996. *Using Multivariate Statistics*, Harper Collins College Publishers, Third Edition.

Thomas, David, H.
1986. *Refiguring Anthropology: First Principles of Probability and Statistics. Prospect Heights,* Illinois: Waveland Press Inc.

Thorpe, I.J.
1996. *The Origins of Agriculture in Europe*, London: Routledge.

Tosta, N.
1991. “More than maps: asking the right questions” *Geo Info Systems*, February: 46-48.

Trigger, B.
1968. Determinants of Settlement Patterns” In *Settlement Archaeology*, edited by K.C. Chang, California: National Press Books

Tringham, R.
1971. *Hunters, Fishers & Farmers of Eastern Europe 6000-3000 B.C.,* London: Hutchinson & Co.

Van Dalen, J
1999. “Probability Modeling: a Bayesian and a Geometric Example” In *Geographical Information Systems and Landscape Archaeology (The Archaeology of Mediterranean Landscapes 3*), edited by Gillings, M et al., 117-124. Oxford: Oxbow Books.

Van der Veen, M.
2003. “When is food a luxury?” *World Archaeology* Vol. 34(3): 405–427.

Van Leusen
1993. “Cartographic modeling in a cell-based GIS.” In *Computing the Past: computer applications and quantitative methods in archaeology*, edited by J. Andresen, T. Madsen, and I. Scollar, pp. 105-123, Denmark: Aarhus University Press.
2001. *Pattern to Process: Methodological Investigations into the Formation and Interpretation of Spatial Patterns in Archaeological Landscapes* (Chapter 5), doctorat in de Letteren aan de Rijksuniversiteit Groningen.

Van Zeist, W., H. Woldering & D. Stipert
1975. “Late Quaternary Vegetation and Climate of Southwestern Turkey” *Palaeohistoria,* 17: 53-143

VanZeist, W. & H. Buitenhuis
1982. “A Palaeobotanical Study of Neolithic Erbaba, Turkey” *Anatolica* 10:47-89

Warren, Robert E
1990. “Predictive modelling of archaeological site location: a case study in the Midwest” In *Interpreting Space: GIS and Archaeology*, edited by K.M. Allen, S.W. Green, S.W., and E. Zubrow, 226-238. London: Taylor & Francis Ltd.

Warren, R. & D. Asch
2000. “A Predictive Model of Archaeological Site Location in the Eastern Prairie Peninsula” In *Practical Applications for Archaeologist: A Predictive Modeling Kit,* edited by K.L. Westcott, and R.J. Brandon, London: Taylor & Francis

Watson, J.
1983. *Geology and Man: An Introduction to Applied Earth Science*, English Language Book Society, George Allen & Unwin

Westcott, K.L. & R.J. Brandon,
2000. *Practical Applications for Archaeologist: A Predictive Modeling Kit,* London: Taylor & Francis.

Wescott, K.L., Kuiper, J.A.
2000. "Using a GIS to Model Prehistoric Site Distributions in the Upper Chesapeake Bay." In *Practical Applications for Archaeologist: A Predictive Modeling Kit,* edited by K.L. Westcott, and R.J. Brandon, London: Taylor & Francis

Wheatley, D. & M. Gillings
2002. *Spatial Technology and Archaeology: The Archaeological Applications of GIS*, Taylor and Francis.

White, A.M.
2002. Archaeological Predictive Modeling of Site Location Through Time: An Example from the Tucson Basin, Arizona, http://www.ucalgary.ca/~amwhit/Thesis/A%20M%20White%20MSc%20Thesis%202002%20(300dpi).pdf

Whitley, T.G.
2001. *Using GIS to Model Potential Site Areas at the Charleston Naval Weapons Station, South Carolina: An Alternative Approach to Inferential Predictive Modeling*, Paper presented at the conference "GIS and Archaeological Predictive Modeling: Larger-Scale Approaches to Establish a Baseline For Site Location Models"

Wikipedia
Problem of induction, (last retrieved 26 July 2006)
http://en.wikipedia.org/wiki/Problem_of_induction

Wikipedia
Dempster-Shafer Theory, (last retrieved 26 July 2006)
http://en.wikipedia.org/wiki/Dempster-Shafer

Wilkinson, T.J.
2003. *Archaeological Landscapes of the Near East*, Tucson: The University of Arizona Press

Williams, L., Thomas, D.H., and Bettinger, R.
1973. "Notions to Numbers: Great Basin Settlements as Polythetic Sets." In *Research and Theory in Current Archaeology* edited C.L. Redman, 215-237. John Wiley & Sons, New York.

Woldring, H. & S. Bottema
2002. "The vegetation history of east-Central Anatolia in relation to archaeology: the Eski Acigöl evidence compared with the Near Eastern environment" *Palaeohistoria:* 43–44

Yakar, J.
1994. *Prehistoric Anatolia The Neolithic Transformation and the Early Chalcolithic Period*,139-180, Tel Aviv University, Sonia and Marco Nadler Institute of Archaeology, Monograph Series, Number 9.

Yuan, M.
1996. Temporal GIS and Spatio-Temporal Modeling (last retrieved 26 July 2006) http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/yuan_may/may.html

Zangger, E
1992. "Neolithic to Present Soil Erosion in Greece, Past and Present Soil Erosion" In *Archaeological and Geographical Perspectives*, edited by M. Bell and J. Boardman, Oxford: Oxbow Monograph 22; Oxbow Books

APPENDICES

APPENDIX-A

List of categorical variables used in the logistic regression analysis and their value labels.

SITE PRESENCE	
1	Site
0	No Site

ASPECT	
1	Flat
2	337,5 – 22,5
3	22,5 – 67,5
4	67,5 – 112,5
5	112-5 – 157,5
6	157,5 - 202,5
7	202,5 – 247,5
8	247-5 – 292,5
9	292,5 – 337,5

EROSION (ERZ)	
1	No Erosion
2	Moderate Erosion
3	Heavy Erosion
4	Very Heavy Erosion

SOIL DEPTH (TOK_D)	
1	Deep Soil
2	Moderately Deep
3	Shallow
4	Very Shallow
5	No Soil

SOIL INCLUSION SIZE (TOK_B)	
1	Fine
2	Medium
3	Coarse
4	Mixed

CURRENT LAND USE (CLU)	
1	Agriculture
2	Yard and Garden
3	Grassland
4	Shrubbery and Forest

LITHOLOGY	
1	Quaternary
2	Volcanic
3	Melange
4	Neritic
5	Clastic
6	Carbonate Clastic

ROUGHNESS	
1	Smooth
2	Considerably Smooth
3	Average
4	Moderately Rough
5	Rough

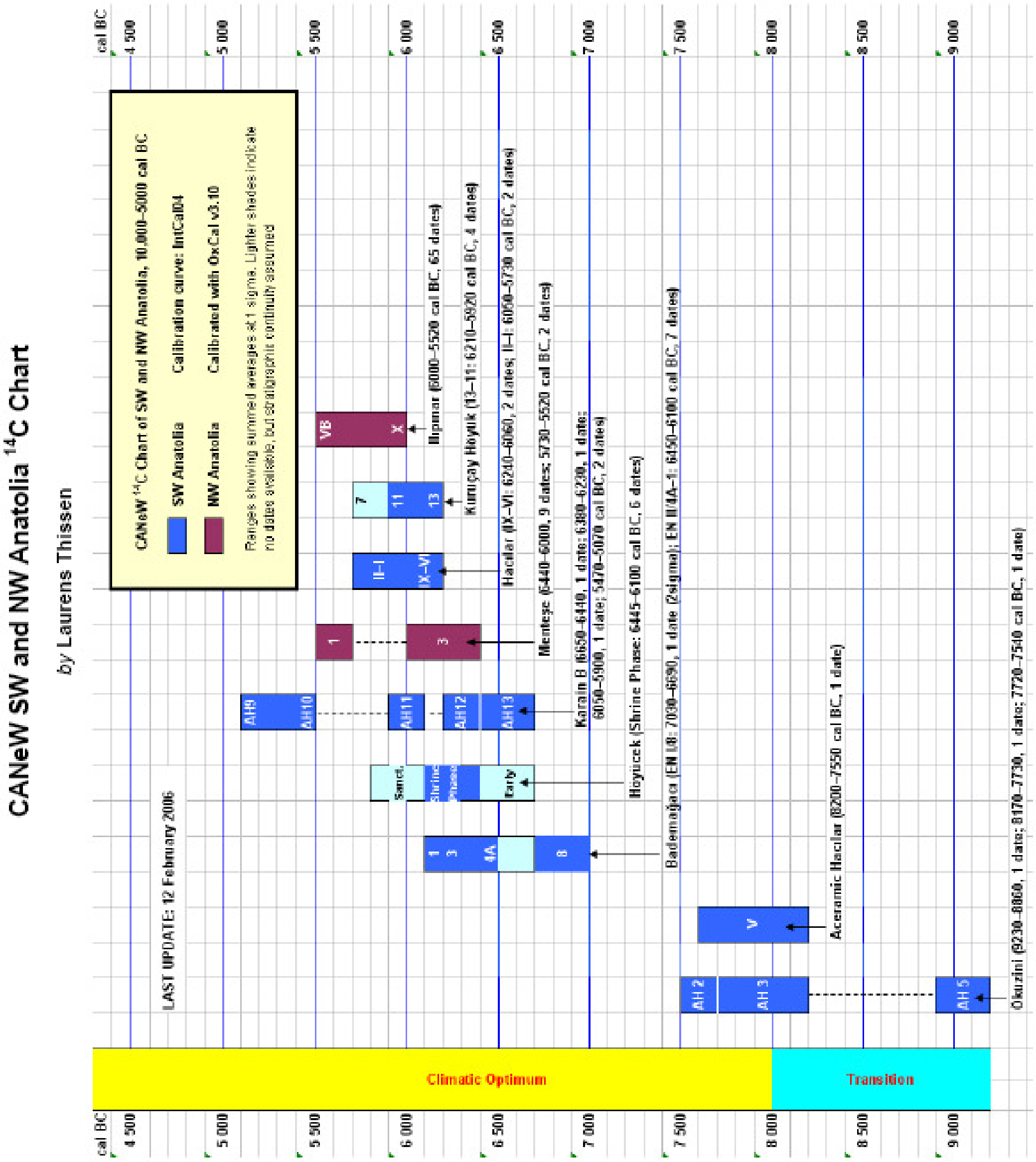
APPENDIX-B

List of sites used in the analysis

NAME	N	E	NAME	N	E
Akcay1	36,58	29,74	KayaliI	37,30	29,93
AlanHoyuk	37,63	31,61	Kecili	37,36	30,39
Aziziye	37,42	30,23	KizilHoyuk	37,88	30,63
Baskuyu1	37,51	29,88	Kizilviran	37,86	32,07
BektemurHoyuk	37,66	31,82	KizlarHoyuk4	37,24	30,29
Beykoy	38,05	31,37	Kurucay	37,63	30,16
BeysehirHoyukC	37,79	31,68	Kuyucak	38,22	31,21
Cavdir	37,14	29,65	Leylekbeleni	37,61	30,38
Cemcem	37,69	31,74	Murseller	37,40	29,94
Derekoy1	37,66	29,81	OrtaKaraviran	37,37	32,06
Duden	37,55	29,81	Pamuklu	37,77	30,68
Efeoglu	37,13	30,29	Salda	37,52	29,61
Erbaba	37,75	31,68	Sarikayali_AyasCiftlik	37,95	31,30
EskikoyYeri	37,30	30,25	Sazak	37,58	29,93
EvregiII	37,58	31,83	SazliHoyuk	37,26	32,23
Gokpinar	36,73	29,98	Seydiler	37,25	29,79
Golde	37,50	30,24	SeydisehirHoyuk	37,44	31,86
Hacilar	37,57	30,08	Sorkun	37,15	29,59
Hacimusalar_Beyler	36,64	29,83	Suberde_GoruklukTepe	37,35	31,92
HanvakfiEskiII	37,51	31,87	Tekke	36,62	29,88
Heybeli	37,36	30,35	Teknepinar	38,34	31,09
Hoyucek	37,45	30,55	Tepeli	37,75	30,88
IlyasI	37,77	30,15	Topraktol	37,74	31,43
Incirdere	37,47	30,53	YagcaTasHoyuk	37,18	30,33
Incirlipinar_IncirliII	37,87	30,42	Yakaemir	38,08	31,23
KanalHoyuk	37,38	31,98	Yakalar	37,32	30,00
KanliHoyuk	37,83	30,60	Yarikkaya_Isparta	38,45	31,07
Kanlitepe	37,71	29,96	YazirHoyuk	37,67	29,88
Karaaliler	37,35	30,48	Yenice	37,35	29,98
Karamusa	37,18	29,75	YenikoyHoyuk	37,98	31,34
Karayaka	37,98	31,40	YilanHoyuk	37,77	31,64
KarayugHoyuk	37,90	30,57	Yug_Bozanonu	37,86	30,56

APPENDIX-C

Some C14 dates for major sites



APPENDIX-D

Model Summaries of the predictive models.

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM1	0.619	0.833

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	116	4	96.7
	Site	6	79	92.9
			Overall Percentage	95.1

Variables

	S.E.	Wald Statistics
S	0.048	3.997
EA	2.375	16.447
DtL	0.029	10.589
DtMSC	0.215	9.342
Constant	1.293	30.952

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM2	0.555	0.748

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	112	8	93.3
	Site	11	74	87.1
			Overall Percentage	90.7

Variables

	S.E.	Wald Statistics
CLU(1)	0.618	11.899
CLU(2)	1.489	0.927
CLU(3)	0.956	0.009
EA	1.921	12.604
DtL	0.028	6.832
DtMSC	0.212	3.859
A(1)	19.799	0.090
A(2)	0.960	0.215
A(3)	0.966	0.287
A(4)	1.076	2.872
A(5)	1.156	5.246
A(6)	1.342	4.700
A(7)	1.336	1.952
A(8)	1.021	0.717
Constant	1.221	9.059

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM3	0.477	0.643

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	106	14	88.3
	Site	17	68	80.0
			Overall Percentage	84.9

Variables

	S.E.	Wald Statistics
CLU(1)	0.521	16.167
CLU(2)	20.510	0.153
CLU(3)	0.791	0.216
EA	1.498	21.023
DtL	0.021	5.898
A(1)	0.997	0.038
A(2)	0.862	2.639
A(3)	0.818	0.121
A(4)	0.932	0.849
A(5)	0.929	4.205
A(6)	0.987	5.964
A(7)	1.006	3.660
A(8)	0.876	0.002
Constant	0.871	2.976

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM4	0.415	0.559

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	107	13	89.2
	Site	21	64	75.3
			Overall Percentage	83.4

Variables

	S.E.	Wald Statistics
CLU(1)	0.494	5.874
CLU(2)	19.637	0.166
CLU(3)	0.746	0.455
EA	1.436	17.237
L(1)	0.623	8.469
L(2)	60.439	0.002
L(3)	0.663	7.457
L(4)	0.716	2.390
L(5)	0.623	0.005
Constant	0.708	17.237

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM5	0.386	0.520

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	101	19	84.2
	Site	21	64	75.3
			Overall Percentage	80.5

Variables

	S.E.	Wald
CLU(1)	0.545	10.450
CLU(2)	18.860	0.262
CLU(3)	0.858	1.117
L(1)	0.658	4.089
L(2)	42.638	0.035
L(3)	0.711	1.407
L(4)	0.800	2.356
L(5)	0.703	0.837
EA	1.510	10.470
DtR	0.062	9.235
A(1)	0.993	2.649
A(2)	0.740	3.475
A(3)	0.707	0.774
A(4)	0.783	1.451
A(5)	0.855	0.019
A(6)	1.016	0.452
A(7)	0.931	0.003
A(8)	0.741	0.133
Constant	0.933	6.023

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM6	0.298	0.401

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	98	22	81.7
	Site	27	58	68.2
			Overall Percentage	76.1

Variables

	S.E.	Wald
CLU(1)	0.543	10.616
CLU(2)	0.839	0.880
CLU(3)	0.836	0.670
L(1)	0.753	4.501
L(2)	0.811	2.965
L(3)	0.857	2.819
L(4)	0.748	0.114
TOK_D(1)	0.499	0.599
TOK_D(2)	0.536	10.321
A(1)	0.724	1.548
A(2)	0.652	0.666
A(3)	0.728	4.035
A(4)	0.763	0.395
A(5)	0.721	0.618
A(6)	0.843	1.848
A(7)	0.787	0.923
A(8)	0.693	0.010
Constant	1.004	11.551

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM7	0.195	0.263

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	90	30	75.0
	Site	34	51	60.0
			Overall Percentage	68.8

Variables

	S.E.	Wald
S	0.033	5.242
TOK_D(1)	0.455	0.680
TOK_D(2)	0.444	5.141
A(1)	0.674	0.504
A(2)	0.676	1.038
A(3)	0.659	0.202
A(4)	0.665	1.018
A(5)	0.692	2.636
A(6)	0.807	5.107
A(7)	0.758	2.678
A(8)	0.668	0.084
Constant	0.618	0.063

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM8	0.244	0.328

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	96	24	80.0
	Site	33	52	61.2
Overall Percentage				72.2

Variables

	S.E.	Wald
TOK_D(1)	0.492	0.680
TOK_D(2)	0.476	2.990
S	0.037	7.032
DtR	0.035	9.075
DtL	0.017	7.280
A(1)	0.669	0.170
A(2)	0.647	0.672
A(3)	0.690	2.458
A(4)	0.683	0.859
A(5)	0.716	0.428
A(6)	0.832	3.806
A(7)	0.766	2.563
A(8)	0.669	0.677
Constant	0.717	0.355

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM9	0.293	0.394

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	99	21	82.5
	Site	24	61	71.8
			Overall Percentage	78.0

Variables

	S.E.	Wald
CLU(1)	0.470	15.203
CLU(2)	0.825	4.856
CLU(3)	0.748	0.588
R(1)	1.299	0.336
R(2)	1.293	0.052
R(3)	1.475	0.742
DtR	0.040	5.514
DtMSC	0.160	4.080
A(1)	0.817	0.058
A(2)	0.702	0.062
A(3)	0.719	0.120
A(4)	0.745	4.093
A(5)	0.790	3.205
A(6)	0.865	7.556
A(7)	0.850	3.046
A(8)	0.756	1.896
Constant	1.461	0.071

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM10	0.297	0.400

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	98	22	81.7
	Site	31	54	63.5
Overall Percentage				74.1

Variables

	S.E.	Wald
CLU(1)	0.436	16.424
CLU(2)	0.848	4.229
CLU(3)	0.718	0.016
TOK_D(1)	0.472	3.006
TOK_D(2)	0.468	10.477
A(1)	0.771	3.406
A(2)	0.679	0.910
A(3)	0.695	0.955
A(4)	0.695	0.300
A(5)	0.725	1.831
A(6)	0.832	4.122
A(7)	0.844	0.632
A(8)	0.705	0.134
Constant	0.687	10.751

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM11	0.363	0.401

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	105	15	87.5
	Site	27	58	68.2
			Overall Percentage	79.5

Variables

	S.E.	Wald
CLU(1)	0.447	12.087
CLU(2)	0.882	0.285
CLU(3)	0.812	0.139
L(1)	0.708	3.861
L(2)	99.634	0.005
L(3)	0.706	0.517
L(4)	0.722	0.030
L(5)	0.713	0.066
TOK_D(1)	0.512	0.000
TOK_D(2)	0.493	7.407
A(1)	21.224	0.154
A(2)	0.708	1.571
A(3)	0.657	0.097
A(4)	0.728	0.674
A(5)	0.797	0.486
A(6)	0.908	2.557
A(7)	0.824	1.597
A(8)	0.757	0.002
Constant	0.875	6.249

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM12	0.294	0.397

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	109	11	90.8
	Site	35	50	58.8
			Overall Percentage	77.6

Variables

	S.E.	Wald Statistics
E	0.990	8.452
L(1)	0.591	15.215
L(2)	13.506	0.104
L(3)	0.595	1.488
L(4)	0.612	0.002
L(5)	0.560	0.165
Constant	1.280	3.701

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM13	0.351	0.472

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	98	22	81.7
	Site	24	61	71.8
			Overall Percentage	77.6

Variables

	S.E.	Wald
CLU(1)	0.438	9.574
CLU(2)	0.974	3.384
CLU(3)	0.762	0.331
EA	1.204	10.400
A(1)	0.902	3.780
A(2)	0.695	4.063
A(3)	0.707	3.466
A(4)	0.724	0.089
A(5)	0.715	0.815
A(6)	0.810	3.290
A(7)	0.771	1.449
A(8)	0.666	0.019
Constant	0.591	0.531

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM14	0.298	0.401

Classification Table

<table>
<tr><th colspan="2" rowspan="3">Observed</th><th colspan="3">Predicted</th></tr>
<tr><th colspan="2">Presence-Absence</th><th rowspan="2">Percentage Correct</th></tr>
<tr><th>Non-Site</th><th>Site</th></tr>
<tr><td>Presence-Absence</td><td>Non-Site</td><td>98</td><td>22</td><td>81.7</td></tr>
<tr><td></td><td>Site</td><td>27</td><td>58</td><td>68.2</td></tr>
<tr><td colspan="4" align="right">Overall Percentage</td><td>76.1</td></tr>
</table>

Variables

	S.E.	Wald
EA	1.274	9.807
S	0.039	5.570
L(1)	0.621	4.292
L(2)	20.314	0.072
L(3)	0.611	2.054
L(4)	0.658	1.144
L(5)	0.598	0.351
Constant	0.628	1.433

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM15	0.431	0.580

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	105	15	87.5
	Site	23	62	72.9
			Overall Percentage	81.5

Variables

	S.E.	Wald
ERZ(1)	0.993	2.601
ERZ(2)	0.650	12.389
ERZ(3)	0.580	7.932
L(1)	0.801	4.895
L(2)	15.010	0.055
L(3)	0.697	1.729
L(4)	0.732	0.024
L(5)	0.686	0.717
EA	1.425	7.927
A(1)	1.375	3.030
A(2)	0.743	2.436
A(3)	0.762	1.619
A(4)	0.840	0.338
A(5)	0.859	0.673
A(6)	0.969	0.359
A(7)	0.834	0.119
A(8)	0.823	0.000
Constant	1.000	2.248

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM16	0.474	0.638

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	104	16	86.7
	Site	18	67	78.8
			Overall Percentage	83.4

Variables

	S.E.	Wald
DtMSC	0.121	5.704
E	1.303	5.561
A(1)	1.401	1.901
A(2)	1.303	6.022
A(3)	1.366	1.589
A(4)	1.369	5.294
A(5)	24.477	0.036
A(6)	1.427	10.439
A(7)	1.348	16.266
A(8)	1.551	0.043
Constant	2.154	10.517

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM17	0.446	0.600

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	104	16	86.7
	Site	19	66	77.6
			Overall Percentage	82.9

Variables

	S.E.	Wald
DtRd	0.001	5.517
A(1)	20.070	0.114
A(2)	20.067	0.172
A(3)	20.069	0.137
A(4)	20.068	0.190
A(5)	31.755	0.000
A(6)	20.073	0.233
A(7)	20.069	0.306
A(8)	20.078	0.105
Constant	20.063	0.154

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM18	0.403	0.542

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	97	23	80.8
	Site	27	58	68.2
			Overall Percentage	75.6

Variables

	S.E.	Wald
E	1.151	5.623
A(1)	19.878	0.138
A(2)	19.875	0.172
A(3)	19.875	0.190
A(4)	19.879	0.175
A(5)	31.515	0.000
A(6)	19.881	0.264
A(7)	19.879	0.276
A(8)	19.902	0.078
Constant	19.922	0.331

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM19	0.436	0.588

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	102	18	85.0
	Site	19	66	77.6
			Overall Percentage	82.0

Variables

	S.E.	Wald
DtR	0.044	4.716
DtRd	0.001	4.272
A(1)	19.606	0.150
A(2)	19.603	0.194
A(3)	19.604	0.156
A(4)	19.606	0.196
A(5)	30.823	0.000
A(6)	19.610	0.249
A(7)	19.606	0.335
A(8)	19.611	0.141
Constant	19.604	0.198

Model Summary

	Cox & Snell R Square	Nagelkerke R Square
PM20	0.461	0.621

Classification Table

Observed		Predicted		
		Presence-Absence		Percentage Correct
		Non-Site	Site	
Presence-Absence	Non-Site	104	16	86.7
	Site	20	65	76.5
			Overall Percentage	82.4

Variables

	S.E.	Wald
RGH(1)	63.666	0.012
RGH(2)	63.666	0.008
RGH(3)	63.670	0.005
A(1)	32.253	0.077
A(2)	32.252	0.076
A(3)	32.252	0.077
A(4)	32.255	0.081
A(5)	50.624	0.000
A(6)	32.256	0.166
A(7)	32.254	0.153
A(8)	32.267	0.048
Constant	71.367	0.046

APPENDIX – E

VIF and tolerance levels for the predictive models.

PM1	***VIF***	***TOLERANCE***
Adjusted Elevation	2,044	0,489
Slope	1,422	0,703
Distance to Major Soil Classes	1,422	0,693
Distance to Lake	1,247	0,802
PM2	***VIF***	***TOLERANCE***
Slope	1,598	0,626
Aspect	1,094	0,914
Adjusted Elevation	1,737	0,576
Current Land Use	1,746	0,573
Distance to Major Soil Classes	1,132	0,883
Distance to Lake	1,161	0,861
PM3	***VIF***	***TOLERANCE***
Aspect	1,039	0,963
Adjusted Elevation	1,247	0,802
Distance to Lake	1,066	0,938
Current Land Use	1,158	0,864
PM4	***VIF***	***TOLERANCE***
Adjusted Elevation	1,331	0,751
Lithology	1,208	0,828
Current Land Use	1,265	0,790
PM5	***VIF***	***TOLERANCE***
Aspect	1,509	0,944
Adjusted Elevation	1,737	0,576
Lithology	1,184	0,844
Current Land Use	1,349	0,741
Distance to River	1,468	0,681
PM6	***VIF***	***TOLERANCE***
Aspect	1,026	0,974

Lithology	1,149	0,870
Current Land Use	1,223	0,818
Soil Depth	1,163	0,860
PM7	***VIF***	***TOLERANCE***
Slope	1,243	0,804
Aspect	1,024	0,977
Soil Depth	1,220	0,819
PM8	***VIF***	***TOLERANCE***
Aspect	1,032	0,969
Slope	1,297	0,771
Soil Depth	1,256	0,796
Distance to Lake	1,032	0,969
Distance to River	1,085	0,922
PM9	***VIF***	***TOLERANCE***
Aspect	1,038	0,963
Surface Roughness	1,109	0,902
Current Land Use	1,040	0,961
Distance to Major Soil Class	1,082	0,924
Distance to River	1,118	0,842
PM10	***VIF***	***TOLERANCE***
Aspect	1,008	0,992
Soil Depth	1,176	0,850
Current Land Use	1,177	0,849
PM11	***VIF***	***TOLERANCE***
Aspect	1,035	0,967
Lithology	1,175	0,851
Current Land Use	1,139	0,878
Soil Depth	1,195	0,837
PM12	***VIF***	***TOLERANCE***
Distance to Surface Roughness Junctions	1,011	0,989
Elevation	1,011	0,989
PM13	***VIF***	***TOLERANCE***
Adjusted Elevation	1,278	0,783
Current Land Use	1,293	0,773
Aspect	1,024	0,977

PM14	***VIF***	***TOLERANCE***
Slope	1,751	0,571
Adjusted Elevation	1,737	0,576
Lithology	1,237	0,808
PM15	***VIF***	***TOLERANCE***
Aspect	1,043	0,959
Adjusted Elevation	1,449	0,690
Lithology	1,207	0,828
Erosion	1,478	0,677
PM16	***VIF***	***TOLERANCE***
Elevation	1,567	0,638
Aspect	1,126	0,888
Distance to Major Soil Classes	1,420	0,704
PM17	***VIF***	***TOLERANCE***
Aspect	1,069	0,935
Adj. Distance to Basin Ridges	1,069	0,935
PM18	***VIF***	***TOLERANCE***
Elevation	1,154	0,866
Aspect	1,154	0,866
PM19	***VIF***	***TOLERANCE***
Aspect	1,051	0,951
Adj. Distance to Basin Ridges	1,106	0,904
Distance to River	1,114	0,898
PM20	***VIF***	***TOLERANCE***
Aspect	1,007	0,993
Surface Roughness	1,007	0,993