PREDICTING MESOLITHIC PIONEER SETTLEMENTS IN EASTERN MIDDLE SWEDEN
- A STUDY OF PREDICTIVE MODELLING THROUGH THE APPLICATION OF ARCGIS, R AND TOPOGRAPHICAL RECONSTRUCTION

Marcus Asserstam

Two year master´s thesis 2015

Supervisor: Kjel Knutsson
Co-supervisor: Daniel Löwenborg
ABSTRACT

Asserstam, Marcus. 2015.

*Predicting Mesolithic Pioneer Settlements In Eastern Middle Sweden – a Study of predictive modelling through the application of ArcGIS, R and topographical reconstruction*

A two-year master’s thesis in Archaeology and Ancient History, Uppsala University

This master thesis investigates the possibilities of constructing a predictive model for Mesolithic pioneer settlements in the area around Torsåker, Gästrikland, Sweden. The underlying theory is that the first humans in this area originate from the Swedish west coast and the Norwegian south. This connection is made through a handle core in quartz that has been shoreline dated to the pioneering phase in the area. This handle core is evidence of the same stone tool technology being used contemporaneous in the study area as in the assumed area of origins. Furthermore, the theoretical concepts of Cultural Transmission and Habitus are central to the study. A sample for model construction was collected from the Swedish west coast and the Norwegian south.

The model was constructed using ArcGIS and R statistical software, and the statistical method chosen was logistic regression. Topographical reconstruction was performed for sites being altered by modern exploitation as a means of enhancing the quality of the input data. The result of the study is that the variables used for predicting Mesolithic settlements have limitations and additional variables are needed. When conducting field survey for validation of the model, ten potential Mesolithic settlements were located.

*Keywords: Archaeology, Mesolithic Pioneers, Predictive Modelling, ArcGIS, R, Logistic Regression, Topographical Reconstruction.*

Marcus Asserstam. Department of archaeology and ancient history, Uppsala University, Thunbergsvägen 3, Box 626, SE-751 26 Uppsala, Sweden

asserstam@hotmail.com
“essentially, all models are wrong, but some are useful”
- George Box, 1987
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ....................................................................................................................... 6

1. INTRODUCTION .............................................................................................................................. 7
   1.1 Purpose ...................................................................................................................................... 7
   1.2 Background ............................................................................................................................... 7
   1.3 Questions ................................................................................................................................. 8
   1.4 Theory ....................................................................................................................................... 8
   1.5 Method .................................................................................................................................... 10
      1.5.1 The four steps of predictive modelling .............................................................................. 10
      1.5.2 Statistical methods for predictive modelling ................................................................. 11
      1.5.3 GIS .................................................................................................................................. 12
      1.5.4 R ..................................................................................................................................... 13
      1.5.5 Scale ................................................................................................................................. 13
   1.6 The Mesolithic in Southern Northland ..................................................................................... 14
      1.6.1 Mesolithic settlement pattern ......................................................................................... 17
      1.6.2 Environment ..................................................................................................................... 18
      1.6.3 Mesolithic Torsåker ......................................................................................................... 18
   1.7 Predictive modelling .................................................................................................................. 19
      1.7.1 History of predictive modelling ....................................................................................... 19
      1.7.2 The concept of Predictive modelling ................................................................................. 20

2. REFERENCE MATERIAL FOR MODELBUILDING ..................................................................... 23
   2.1 The sample ............................................................................................................................... 23
   2.2 FMIS and Askeladden ................................................................................................................ 23
   2.3 Digital Elevation Model .......................................................................................................... 25
   2.4 Variables ................................................................................................................................ 25
      2.4.1 Slope ................................................................................................................................. 26
      2.4.2 Distance to Coast ............................................................................................................ 26
      2.4.3 Soil/Topsoil ...................................................................................................................... 27

3. QUALITY ENHANCEMENT OF TOPOGRAPHY ...................................................................... 28
   3.1 Reconstruction of pre-exploitation topography .................................................................... 28

4. STATISTICAL ANALYSIS ............................................................................................................... 34
   4.1 Logistic regression analysis .................................................................................................... 34
   4.2 Kvamme’s Gain ....................................................................................................................... 39

5. APPLICATION AND VALIDATION .............................................................................................. 41
   5.1 The predictive area .................................................................................................................. 41
   5.2 Validation of the model .......................................................................................................... 43

7. DISCUSSION ................................................................................................................................. 48
   7.1 In the context of predictive modelling critique ...................................................................... 48
   7.2 Result of the predictive model ................................................................................................ 50
   7.2 Model development ................................................................................................................. 52
   7.3 Conclusory thoughts ............................................................................................................... 53

8. SAMMANFATTNING ..................................................................................................................... 54

9. REFERENCES ................................................................................................................................. 56
ACKNOWLEDGMENTS

First, I would like to take this opportunity to thank my two supervisors, Kjel Knutsson and Daniel Löwenborg. Kjel, for always making time for discussions whenever needed. For helping me to keep on the main track, and mostly for your enthusiasm, engagement and for your ability to always send me of with a positive feeling. Daniel, for all your help and patience with technical solutions no matter how rudimentary or advanced they might have been, and also for pushing me to engage even more in all complicated issues, which hopefully have enhanced the quality of this thesis.

I am also grateful to Per Persson for taking the time to help me with all my queries and for providing me with your most valuable database. Also, thank you Detlef Klamke for all your service, commitment and help with the field survey. I am deeply impressed by your Brazilian contacts. And thank you Michel Guinard for your help at the field survey, and for interesting opinions upon the topics of this thesis. Hopefully, one day I will know the meaning of stone.

And finally, special thanks to my wife and daughter for shedding light in the darkest of days!

Marcus Asserstam

Norrköping

June, 2015
1. INTRODUCTION

1.1 Purpose
The purpose of this thesis is to investigate the possibilities of constructing a predictive model, using ArcGIS and R (statistical software) to identify archaeological remains, which can be seen as a form of digitized inventory of the cultural landscape. In this paper, the postglacial pionee- ring phase of southern Norrland (11 000 – 10 000 BP) is the focal point. The underlying theory is that the first people in the region originated from areas on the Swedish west coast, and Norwe- gian south. This is based on the evaluation of a method of lithic blade production by using pressure blade technique that originated in western Russia and the Baltic states, and large amounts of remains from this technique have been collected as stray finds in southern Norrland (Knutsson & Knutsson, 2012:16). If this theory of origins is correct, then it is likely that there are more similarities than stone artefacts between archaeological sites that chronologically coincide between the west coast and southern Norrland. One of these similarities could be choices of settlement locations.

Against this background, two main purposes are formulated; 1) to investi- gate possibilities of constructing predictive models using ArcGIS and R to locate previously unidentified archaeological sites, in this case Mesolithic pioneer settlements, and 2) to examine the theory of a west- east migration to southern Norrland.

1.2 Background
Knowledge about the post-glacial pionee- rers in southern Norrland, extending through Norrland, is limited in comparison to the rest of Scandinavia. Mainly, Mesolithic research has focused on stone technologies and the processes leading to the tools used by humans. The reason for this is partly because the permanence of stone tools and debitage in relation to organic materials (Sørensen, et al., 2012:4).

The knowledge base for the pionee- rers of southern Norrland is no exception, which is also based on the identification of a specific stone technology; namely pressure blade technology. Briefly, this technology is characterized by blade production by applying pressure to the core, using some sort of tool, e.g. antler or bone (see Kooyman, 2000, for further details). This technique is a highly advanced technique which assumes prior knowledge and practical knowhow. This technique has chronologically been traced from the Russian and Baltic states, through the northern parts of Scandinavia, down along the coast of Norway, and later into the southern parts of Norrland, spreading either by migration of humans and/or diffusion of ideas between different cultural groups. The inherent complexity of this technique has made researchers disqualify theo- ries of multiple innovation areas. A distinction is made between pressure technique and knapping techniques in which substantial knowledge are reduced to knowing the size and quality of the hammer used (Sørensen, et al., 2012:4-5).
The earliest Swedish prehistory north of river Dalälven is sparsely represented by material culture, and almost all material is from a collection of stray finds gathered by archaeologist Ragnar Lannerbro. In the northernmost part of Sweden, archaeological activity has been dependent of great lake regulations undertaken during the 20th century (Biörnstad, 2006:15-20). In these areas, the Mesolithic humans are believed to have followed the ice in a north-south direction following the edge of the ice, which is evidenced by short term settlements (Sørensen, et al., 2012:8). As previously mentioned, the earliest archaeological artefacts from southern Norrland indicates a connection with the west coast, with a possible oldest dating of 10 000 BP. Analysis shows that the two migration routes following the ice seem to have met somewhere in northern Dalarna/southern Härjedalen (Knutsson & Knutsson, 2012:15).

From the data presented above, a theory has emerged regarding the origins of the pioneers occupying this newly habitable landscape after the ice retreated. Archaeological material found in the study area show close technological resemblance with the Swedish west coast and southern Norway. This is seen as a relatively clear indication of the origins of the pioneers.

1.3 Questions

Two main research areas are the focal point in this thesis; predictive modelling and postglacial pioneer settlers in southern Norrland. On this basis, two research questions have been formulated:

1. Is there an identifiable settlement pattern for the middle Mesolithic pioneers?
2. What are the possibilities to analyse these patterns through GIS-applications for predictive models?

1.4 Theory

Recent Mesolithic research in Scandinavia has largely focused on the identity of the artefacts found at archaeological excavations. From this research, it has been possible to establish the comprehensive knowledge needed to use the above briefly described pressure blade technique to produce artefacts from stones. Archaeological materials have been analyzed using Chaîne opératoire (Sørensen, et al., 2012:3). The concept of Chaîne opératoire basically means that the remains of material culture production are used to recreate production process, and what are of interest are the processes that generated the end product (Schlanger, 1994). Schlanger emphasize that material culture has a relevant story to tell:

Humans undertake material actions in a range of situations, and for a number of purposes. It can be postulated that, being artificial and intentional, these acts follow certain belief expectations, desires and deliberations (Schlanger, 1994:147)

Dobres (2010:106) also emphasize possibilities with Chaîne opératoire analysis to find connections between prehistoric societies, material culture and its processes. That is to say that a stone tool technology is not only a result of adaption to new situations, but an expression of acquired technological knowledge.

A theoretical framework for how knowledge is acquired can be found in Cultural Transmission theory (CT). Within this framework, transmission and cultural evolution is of primary inter-
est. Most studies of CT have focused at the evolution of culture, and applications for predicting cultural change have been developed (see Cavalli-Sforza & Feldman, 1984; Bettinger & Eerkens, 1999; Eerkens & Lipo, 2007; Mesoudi & O'Brien, 2008). Cultural evolution is not of particular interest for this study. However, within the frames of CT, transmission of knowledge can be explained. CT can be defined as followed:

Cultural transmission is the process of acquisition of behaviours, attitudes, or technologies through imprinting, conditioning, imitation, active teaching and learning, or combines of these. (Cavalli-Sforza, et al., 1982:19)

And as Eerkens & Lipe puts it:

This allows for a very different type of evolutionary process to take place because the results of individual learning (i.e., behaviour modification) can be transmitted, in the modified state, to other individuals. Through individual learning and CT, organisms can continually acquire, modify, and pass on modified information. Thus, the process of CT is fundamentally based on the interaction of both individual experimentation (i.e., innovation) and social learning (i.e., copying). (Eerkens & Lipo, 2007:242)

Cultural transmission is by some seen as cultural packages transmitted. This means that not only one single part of a culture is transmitted, but multiple values. As an example, pots may be reproduced because of technological superiority, but along with this comes the reproduction of the style, colour and other traits that is not central in the transmission, it comes along as a package (Eerkens & Lipo, 2007)

Regarding similarities in cultural expression, there are hypothesis within CT to explain this phenomena. It is believed that people living in similar cultural, social and physical environments will tend to acquire similar worldviews, which consequently may lead to the adoption of similar behavioural traits, e.g. material culture. (Eerkens & Lipo, 2007:244)

In CT, the early learning hypothesis is essential – knowledge acquired early in life is more resistant to change than knowledge obtained later (Bruner, 1956: 194). Adding to this theoretical approach of CT, practice theory and Habitus is applied. This has been defined by Pierre Bourdieu:

The habitus, the durably installed generative principle of regulated improvisations, produces practices which tend to reproduce the regularities immanent in the objective conditions of the production of their generative principle, while adjusting to the demands inscribed as objective potentialities in the situation, as defined by the cognitive and motivating structures making up the habitus (Bourdieu, 1997:78).

What is explained in the above quote is that humans are not entirely governed by structures, but at the same time not liberated from these structures. Habitus is, as Eerkens & Lipo stress for CT, the package of values and norms that we are socialized into and constrained to relate to when faced with making decisions. Bourdieu (1997:79) argue that it is possible to see past human actions in present day human actions, since it is these past actions we relate to. By this, we are affected by the structure of Habitus. Reproduction of a stone tool technology over a vast period of time can be understood as the structure of the production concept remained unchanged over time, and as the result of CT.

As explained above, research on Mesolithic humans have mostly focused on Chaîne opératoire of lithics by identifying the processes leading to the end product. Through this method, the complex knowledge related to pressure blade technology, both theoretical and practical, have been made visible and functioned as a means of describing human migration or diffusion of ide-
as, as the technology can be traced in time and space. For the development of the predictive model, it is important to use the testimony of sites in a similar way that chronologically coincide with the pioneering phase in Southern Norrland, and that indicate similar stone technology.

Furthermore, CT and Habitus are important concepts when predicting settlement patterns since this helps to identify how humans have related to their history. This means that humans that have left one settlement to expand into previously unexplored areas carry knowledge and prior experience of what constitutes a good settlement site. When and if they reach landscapes that do not correspond with these prior experiences, mediation takes place. Humans do not simply adapt to new environments, and the mediation takes place from the perspective of prior experience through Habitus and the possibilities of the new environment.

This is the cornerstone of this study. If humans automatically adapted to new conditions, the possibilities of finding connections to other cultural groups would be minimal. Creating predictive models from cultural traits rests on the assumption of humans reproducing their culture, and that it is expressed in the archaeological material.

1.5 Method

1.5.1 The four steps of predictive modelling

Conolly & Lake (2010:181-186) describes the designing of predictive models as a four-step rocket; data collection, statistical analysis, application of the model and model validation. A short description of this work-flow is given here, and details in these steps will be given descriptions in later sections.

Data collection – Simply put, this is where the database is build. Two types of data are collected that in statistical terms is named the dependent variable and the independent variable. The dependent variable is what we want to predict, i.e. site or non-site. Independent variables are used to determine the likelihood of archaeological potential. These variables can be environmental (e.g. slope, drainage, vegetation, etc.), social, cultural or whatever helps predicting site location. The dataset is usually split up, called split-sampling, into training sample and testing sample. Where training sample is the part of the dataset form which the model is constructed, and the testing sample is used for testing the model.

Statistical analysis – In this step, the attributes that significantly discriminate between site and non-sites are identified. According to Conolly & Lake (2010:183), the preferred technique is logistic regression. However, there are plenty of statistical methods to use for predictions and there are today different schools of what is the “best” statistical method for archaeology.

Basically, the database used for statistical analysis has two “sides”; dependent variable and independent variable. The dependent variable is what we want to predict, in this case middle Mesolithic pioneers settlements. But it can also take several expressions. It can be Binary, as in this study, meaning that the input is either site or non-site. It can be multinomial, meaning that there are more than two dependent variables. It can be based on Density, where the frequency is what is predicted. It can also be site significance, where the dependent variable in the input is ranked in order to locate site with higher archaeological importance (Svedjemo, 2014:97).

The independent variables are the characteristics recorded at each site, and predictions are made based on correlation between these variables. The characteristics are divided into four major themes; environmental variables (e.g. elevation, soil type, vegetation, etc.), social and cultural-
al factors (e.g. roads, other settlements, central places, etc.), positional characteristics (e.g. clustering, dispersal) and radiometric characteristics (identified through remotely sensed data, e.g. satellite images and aerial photos) (Svedjemo, 2014:97).

For further detail on available statistical methods, and the method chosen for this study, see section 1.5.2.

Application – When statistical analysis is completed the model is applied to the study area on a cell-by-cell basis. Using map algebra, a score for each map cell is calculated returning the archaeological potential for each cell. This will result in cells providing either a probabilistic ratio scale (i.e. 0 – 100 % probability) or a site or non-site categorization depending on the approach chosen by the researcher.

Validation – This final step is undertaken to determine the accuracy and performance of the model. The model should predict as many sites as possible in as little an area as possible. To predict all sites within a region with 100 % accuracy is actually not very hard; simply include all of the area! Now, this would not be very useful. We want something that can contribute to archaeology, and for the model to do that it needs to be as precise as possible. As an example, constructing a predictive model that will be tested through field survey needs to be effective. If we have a model that predicts 90 % of the sites in 10 % of the total study area, then it is fairly precise and can be further validated through field survey.

Conolly & Lake (2010:184-185) describes how the validation is performed using a testing sample. No matter how important this theoretically digital validation might be, it also needs to be tested in the terrain conducting field survey, as outlined by Verhagen (2005). This point is nicely summarized by Kvamme (1988:303):

> When I asked them [statisticians] about the role of statistical theory in model development, they suggested that I worry less about theory and more about how well the model works in practice.

In addition to the four steps of predictive modelling described above, a fifth step is hereby introduced: enhancing the quality of data. If this truly manifests a step of its own, or if it is a sub-step to data collection might be discussed. However, the first step is solely about data collection and the second step is statistical analysis. This new step places in between these giving the following new order: data collection, quality enhancement, statistical analysis, application and validation. The need of this step will be thoroughly described under section 5. This five step work-flow will be kept throughout this thesis, and sections divided by this definition.

1.5.2 Statistical methods for predictive modelling
To describe characteristics of the sample, statistical methods are applied. Depending on the characteristics of the sample a set of statistical methods are at disposal. Even though there seems to be plenty of statistical options, most of them are variants or analogues to regression (Warren, 1990:92).

Boolean models are easy to use, and do not require statistical knowledge to operationalize, and can be used with small samples. A threshold value is calculated and stored in separate Boolean layers. The layers contains of 1 or 0, where 1 meets the threshold-value, and 0 does not. Adding up the layers and then summimg the values gives locations with higher and lower values where higher values are expressions of prediction (Svedjemo, 2014:97).
Weighted map layer model has the same fundamentals as Boolean models, but with more complexity. Several thresholds are applied, weighting the variables according to its importance in the model (Svedjemo, 2014:97-98).

Regression is, simply put, a statistical technique to predict the score of a new case of the dependent variable, when the score of the independent variable is known. Bivariate linear regression is the simplest form, analyzing the relationship with only two variables. Extending the sample with more variables takes us to multiple regression, that can be applied when the sample consist of one continuous dependent variable and two or more independent variables (Warren, 1990:92). A number of assumptions need to be met for the sample when applying linear regression, e.g. linearity and normal distribution.

Discriminant function analysis is analogues to multiple regression. The aim of this procedure is to create an axis of discrimination that statistically separates two groups from a multivariate sample. The discriminant function is then used to make predictions about group membership (Warren, 1990:72) and is a widely used method (Svedjemo, 2014:98).

Logistic regression is a probabilistic method that is flexible and robust, and can handle variables on all scales (Nominal, ordinal, interval, and ratio). Assumptions of sample distribution do not have to be met (Warren, 1990:92; Svedjemo, 2014:98). Binary logistic regression has two dependent variables in the input data, e.g. site or non-site. Multinomial logistic regression has more than two dependent variables, e.g. Mesolithic settlement, Neolithic settlement, Iron Age settlement. Logistic regression is the statistical procedure chosen for this study.

Geographically Weighted Regression is a technique for modeling spatial relations where variables are allowed to vary over space. This differ from global statistical models, in that the same relationship is not assumed for the whole area of study (Löwenborg, 2009:1-2).

Bayesian statistics is a statistical method that has been gaining ground in recent years. It has the advantage of incorporating both subjective and objective knowledge. A priori knowledge is combined with whatever quantitative data available (Verhagen, 2007:77f). This explicit incorporation of subjective prior beliefs is what differ Bayesian statistics from classical statistics (Millard, 2005:169).

Dempster-Shaefer is an umbrella term for different approaches and interpretations aiming to model degrees of belief. It is conceptually close to Bayesian statistics. Dempster-shaefer theories differ from other probability methods in that it does not assume full information. Other probability methods assume that if there is 10 % probability of a site, then there is 90 % probability of non-site. The term ignorance is used in Dempster-shaefer, meaning that for example a 10 % probability of site can be set, as can non site at 30 %, the “left-over” 60 % is the degree of ignorance. This is an interesting part to model since no sample is perfect.

1.5.3 GIS

The use of GIS (Geographical Information Systems) applications exploded in the 90’s (Verhagen, 2007:15). GIS is a tool that helps to solve archaeological problems. It is applied by users at all levels of expertise, and has become common practice in daily based archaeological activities. A GIS can be defined as both an integrated and an integrating technology to help understanding spatial information with a set of tools. Even though GIS has its roots in digital cartography, it is not only about maps. It contributes to understanding spatial and space-time relationships between natural and human caused phenomena at all scales, from micro to macro (Conolly & Lake, 2010). The success of GIS comes from its ability to easily communicate concepts and results to
the archaeological community. GIS applications were initially developed to analyse land use and answer environmental questions. This came into conflict with the post-processual movement in archaeology in the 90’s. This forced attempts of including less tangible aspects of human behaviour in GIS analysis, attempts that up until today have experienced diversified recognition (Verhagen, 2007:16).

A number of GIS platforms are today available for spatial analysis, e.g. QGIS, GRASS GIS, and ArcGIS. For this study, ArcMap through ArcGIS have been used.

1.5.4 R
Several software programs are available for statistical computing, e.g. SPSS and stata. For this study, R was chosen and will be evaluated for its performance in an archaeological environment. R is used by many professional statisticians, and has the advantage of being open source software.

From the R-projects webpage, the following description of R is provided: “R is a language and environment for statistical computing and graphics.” Within this environment, statistical analysis can be conducted, and the system can easily be extended by usage of additional packages available. In such, it is not limited to a set of tools as other statistical software might be (The R foundation, 2015).

The reason for employing R is twofold. First, it will rationalize the process of gaining sufficient data for the model. It allows for complex, time-consuming calculations to be run “automatically” and thus saving time in comparison with hand-calculated statistics. Secondly, it serves the function of assuring the quality in the mathematical part of the analysis. By hand-calculation, small errors can severely affect the outcome of the analysis and consequently the performance of the model might be affected.

The following packages was used in R to conduct the statistical analysis; Aod, Psych, QuantPsyc & pscl.

1.5.5 Scale
The importance and the problems related to various scales (spatial, temporal and functional) are discussed by van Leusen et al. (2005:58-59), and scale is referred to as both dependent on the extent of the model, and the resolution used by the model. Most models created for archaeological heritage management (AHM) and CRM have not generated separate models for each chronological period. This is motivated by the insufficient number of observations made at each level of scale, resulting in very low confidence levels. To provide a solid base for the predictive models, large datasets are needed to strengthen the statistical correlations made. But also, choosing the scale of measurement is of importance for the outcome of the study, mainly on mathematical grounds. There is no “right” or “wrong” scales of measurement, it is determined by the research design (VanPool & Leonard, 2011:10-11).

The various scales this study employs are the following. Temporal scale is set at approximately 300 years between 10 300 cal. yrs BP – 10 000 cal. yrs BP, and represents a possible pioneering phase in Gästrikland and Torsåker (see section 1.6.3 for more detail).

The study has multiple spatial scales. The overall reference area covers a large area in Sweden and in Norway. The reason for this is the mobile nature of marine hunter gatherers, and even though all sites in the reference material is spread over such a large are, they are assumed
to, in some portion, be part of the same cultural group. Analysis is also made on high quality elevation data, capturing small scale trends within settlements based on elevation, resulting in an analysis being performed at several scales.

1.6 The Mesolithic in Southern Northland

Since the theoretical base for this predictive model is found in the knowledge about Scandinavian Mesolithic, a brief background will here be provided based on recent research developments and current knowledge.

Northern Europe offers a unique environment for archaeological research. The Weichselian glacial maximum is dated to c. 20,000 BP, and the succeeding time is characterized by the melting of the ice. The Fennoscandian area became ice-free at different times. In Norway, the southwestern part became ice-free 18,000 – 16,000 BP and the northern coast 13,000 – 12,000 BP. Denmark was ice-free c. 13,500 BP and most part of Finland c. 10,000 BP (Bergman, et al., 2004:156). The study area around Torsåker became ice-free c. 11,000 BP (Berglund, 2005:528).

Knowledge about postglacial pioneers is restricted by the low amount of excavated sites. Most of the archaeological material consists of stray finds, and most of the few excavated sites are lacking organic material. Hence, very few C\(^{14}\)-dates are available (Brinch-Petersen, 2009:1). Schmitt (1994:247) also touch upon this matter of organic material not surviving, and most dating have been made with the aid of shoreline-displacement curves. Furthermore, this poses difficulties for archaeologists to establish the economic base for the Mesolithic humans. Remains of food (bones, plants, etc.) are rarely found.

After the retreat of the ice, a habitable environment was established during the Bølling-period (13,000 – 12,000 BP). Even though biodiversity allows for human occupation during this period, no traces have been found from humans. The first sporadic evidence of humans in the area is found from the Allerød-period (11,800 – 11,000 BP), but the first evident traces of human occupation is dated to the transit in between younger Dryas and pre-boreal. (Bang-Andersen, 1996) It appears as the first areas of the Fennoscandia to be ice-free, remains uninhabited by humans for thousands of years (Bergman, et al., 2004:159). Archaeological evidence points to a colonization phase of the Swedish west coast and Norwegian coast within 200 – 300 years, preceding 9,000 BC (Bjerck, 2009: 124-125; Glørstad, 2013:59; Damlien, 2014:5).

During younger Dryas (c. 12,700 – 11,500 cal BP) the ice-sheet expanded as a consequence of colder climate. At the end of younger Dryas, there was once again a shift to warmer climate resulting in massive melting and a retreat of the ice by several hundred meters/year (Burdukiewicz, 2011:304). The colonization phase of the Swedish west coast and Norwegian coast, correlates remarkably well with the end of younger Dryas. The effect of younger dryas is apparent when examining temperatures in southern and eastern Baltic where the average temperature dropped from 13 – 16 ° in July to 12 °. The same shift in temperature have been observed at upper Volga river (Western Russia) with a 4 ° drop in average temperature in July, and 10 – 12 ° in January (Zhlin, 1996:86; Burdukiewicz, 2011:305).

Changing biodiversity offers two solutions for humans living near the edge of the ice; follow the ice as it retreats and maintain the established culture, or stay and adapt to the new biodiversity and environment. Pioneering phases includes more than the first visits to unexplored areas; inclusion of areas in a mobile resource cycle. The colonization process in post glacial northern Europe is seen as a two-step rocket. The initial phase is characterized by sporadic expedi-
tions into unexploited area, followed by a higher degree of settledness, which also is interpreted as a consequence of change in biodiversity (Bergman et al., 2004:156, 159-160). According to Brinch-Petersen (2009:4), human and animal (mainly reindeers) occupation of previously unexploited land occurs synchronously in a northerly movement along the edge of the retreating ice.

Different late Palaeolithic and early Mesolithic tool making traditions enables the possibility of identifying areas of origin and migration routes by following spread of different techno-complexes in the archaeological material. In northern Europe, a colonisation-process with two directions of movement has been identified. One westerly tool-making tradition (Post-Ahrensburg) and one easterly (Butovo) have been interpreted as two different groups of people into the area at c 11 000 BP (fig. 1) (Knutsson & Knutsson, 2012). Finland, however, is colonized later; 9 500 – 8 300 BP by Suomusjärvi cultural groups (Bergman et al., 2004:159).

During the late Palaeolithic at upper Volga – at the site Zolotoruchye 1 – regular blades have been found, manufactured by pressure blade technology. This resemble the succeeding Mesolithic culture of Butovo where tool production was conducted using both percussion and pressure techniques (Zhilin, 2005). Blade productions with pressure technique originate from western Russia and the Baltic region, and are particularly related to groups belonging to the post-swiderian techno-complex. The earliest dating for this techno-complex is with the Butovo culture (Sørensen et al., 2013:6). Through technological analysis, contextual analysis and carbon dating, a spread of this techno complex from the east into Scandinavia has been identified during pre-boreal and boreal time (Sørensen et al., 2013:28). A distinction between a western and eastern tradition in blade production can be made. These traditions are maintained, regardless of raw material availability. These expressions of different techno-complexes seen in the archaeological material are seen as manifestations of different cultures, not as functional adaptations (Sørensen et al., 2013:25-26). As a consequence of the pressure blade technology’s inherent complexity, spread is seen to take place through transmission of knowledge. Several areas of innovations are not likely due to its complexity (Sørensen et al., 2013:2).

The recently excavated sites of Sujala and Fällegohtesajeguolbba in northern Finland, contributed to a revision of migration routes and cultural change along the Norwegian coast. Archaeological material from the sites ascribes to blade production using pressure technique. Contemporary sites are known from the southern parts of Finland c. 1 000 km away. Initially there was a closer connection at the Norwegian coast. However, this techno complex at the Norwegian coast (Fosna I) is related to coastal settlements, and Sujala is an inland settlement. This founds the base for the interpretation of the northern Finnish sites being expression of migration routes originating in the post-swiderian regions in the Baltic and Western Russia (Rankama & Kankaanpää, 2011:185). Inland settlements are common at upper Volga where the economic base is interpreted as terrestrial, where moose and beaver dominate faunal remains (Zhilin, 1996:89). According to Rankama & Karkanpää (2011:204-205), expansion from to Sujala from a post-swiderian area, must have been undertaken within the span of one adult lifetime. This assumption is made due to the absence of suitable raw material for pressure blade technology, and as a consequence of the techniques complexity – knowledge needs to be transmitted between individuals. This falsified alternative theories.
The western tradition is characterized by a post-Ahrensburg techno complex (Knutsson & Knutsson, 2012:3), originated in Bromme (Schmitt, 1994), which in turn originated in the Hamburg culture (Brinch-Petersen, 2011). At the Swedish west coast, this culture is labelled Hensbacka (Schmitt, 1994), and at the Norwegian coast as Fosna phase I (Rankama & Kankaanpää, 2011:184). Archaeological material from this western tradition is characterized by long, irregular blades (Rankaama & Kankaanpää, 2011:184) produced by soft hammer percussion technique – e.g. wood, bone, antler (my note) – (Schmitt, 1994:1). This is different from the above presented eastern tradition where blades are long and regular and extracted from the core by pressure.

Around 10 000 BP, the two traditions mentioned above have been mixed. If this is a consequence of migration or diffusion, is a question without answer, even though continuity of certain technological traits strengthen theories of western groups adding pressure techniques to their toolbox (Sørensen et al., 2013:29). This does not disqualify the possibility of a parallel migration to the area, creating a mixed-zone. New evidence of the technological change favors theories of
an eastern migration route over northern Finland, Sweden and Norway for the pressure technique
groups, at the beginning of the Middle Mesolithic c 8 300 cal BC, as opposed to earlier theories
favoring the south-west migration route into Norway for this technology (Knutsson & Knutsson,
2013; Sørensen et al., 2013).

1.6.1 Mesolithic settlement pattern
One of the main goals of this thesis is to investigate if a settlement pattern of middle Mesolithic
pioneers into central Sweden can be made visible through statistical analysis. Therefore, a sec-
tion describing Mesolithic settlement patterns might seem superfluous. However, it is of im-
portance for the discussion in section 7 as to see if this study verifies or falsifies previous notions
of settlement patterns.

In this thesis, a marine economic base with water transportation is assumed, for the early
and middle Mesolithic. There are several reasons for this. First, as explained by Bergman et al.
(2004:159), the rapid colonization of coastal Norway (200 – 300 years) is an effect of a devel-
opment of arctic marine adaption and technology. For this study, this process is assumed to be
continuous for the whole colonization period of Fennoscandia.

Second, to quote Zvelebil (2008:29-30): “…dietary patterns using isotopic nitrogen and
carbon in human bone show that in many coastal areas people tended to specialize in marine
resources from the early Mesolithic onwards…” According to Nordquist (1998:192-193), this is
also true for the Swedish southwest. This enhances the theory of marine focus for pioneer settlers
during the Mesolithic.

Finally, this study includes only sites with distinct dating with radiocarbon methods. When
simulating shorelines related to the time for the sites activity period, all sites are more or less
shore-bound. There might be more evidence for pioneer settlements being shore-bound, e.g. as-
semblage of tools, but this has not been pursued. The above arguments are believed to withhold
enough evidence of a marine culture.

Rogers & Black (1976) have studied subsistence strategy for hunter-gatherers in northern
Ontario, Canada. They set up three principles (1976:20-22) as fundamental to this subsistence
strategy (see below). Even though the study has an apparent processual touch in which resources
govern settlement location, and the influence of cultural aspects are absent, there is one interest-
ing point made. In Principle II, the importance of transportation is emphasized; “…in such a
manner as to minimize time and energy spent on travel and transport…”. As can be seen for the
settlements studied here, this also seems to be true for the Swedish west coast and Norwegian
south. As mentioned above, all studied sites are shore bound. And by holding Bergman et al.’s
argument for colonization taking place rapidly due to marine technology for true, this further
strengthen the notion of a marine culture. Explicitly, if preferred transportsations are boats of
whatever model, settlements are located near the shore, and the boats, minimize time and energy
spent on travel and transport.

Principle I – To seek food resources chiefly at the time when they are most readily and
abundantly available.

Principle II – To locate and distribute the human population (providers and consumers) in
such a manner as to minimize time and energy spent on travel and transport, and regulate group
size, in accordance with resource availability (Principle I) and the existence of appropriate habi-
tat for campsites.
Principle III – To be ready with contingency plans that may override or supersede the rules as given above, when circumstances demanded it for survival.

Further evidence of importance for this thesis is that burial grounds from the Mesolithic tend to be placed in coastal areas. This might indicate the importance of these areas for the Mesolithic groups (Zvelebil, 2008:38). This further strengthens the notion of a marine culture. A study of particular interest to this thesis is Eren’s (2012) investigation of hunter-gatherers responses to environmental change. The dramatic cooling events of younger Dryas are of particular interest in his study. He concludes that, even though there are dramatic climatic shifts, hunter-gatherers do not automatically adapt to new environments emerging. Continuity is more apparent in the archaeological record than change. Constructing Predictive Models for hunter-gatherers, theories of continuity are essential. Predictions rest on the assumptions of repetitive patterns and if cultural values are more important than economic values then it is possible to predict settlement patterns based on cognition rather than environment.

1.6.2 Environment
A consequence of the extensive Weichselian ice-sheet was the massive weight pushing the earth crust downwards. After the ice retreated, the crust is moving back to its position prior to the spread of the ice. Sea level regression is a phenomena related to this process, meaning that the land is rising. Effects of the regression vary and where the ice was thickest the degree of land-rise is the highest. The opposite relation – transgression – is also commonly occurring, meaning that the land is sinking and flooded. Doggerland can be used as an example of transgression. This is the name of the big land area once situated between western Denmark and eastern England, which today is completely under water. During the melting of the Weichselian ice, a big reason for transgression was the amount of water liberated from the ices giving the effect of rising sea-levels. Today, a tilting effect is in play in the Swedish area. Highly generalized it can be described as regression on the east coast, and transgression on the west coast (Schmitt, 1994:251; Bergman et al., 2004:161; Bjerck, 2009:120). This means that the settlements under study in this thesis are today found on high altitudes in the landscape.

The environment of the pre-boreal period has been interpreted to be similar to the environment of today on Svalbard (Bjerck, 2008:66).

1.6.3 Mesolithic Torsåker
The study area around Torsåker became ice-free c. 11 000 BP (Berglund, 2005:528) and the highest located shoreline is located at 200 m a s l (Berglund, 2010:214). The area has been and still are, subject to substantial regression resulting in landscape changes due to the shoreline displacement. The earliest evidence of humans in the area is a blade core of quartz that has been dated to c 10 300 cal. BP using shore-line displacement curves (Knutsson & Knutsson, 2012:5), which is contemporaneous with introduction of the pressure blade tradition in western Scandinavia (Damlien, 2014). Using shore-lines to date archaeological sites is precarious. The reason for this is that humans could have transported the material found to higher elevation, and that the archaeological site is contemporary with a shore-line not connected to the registered site. Glørstad (2013:3) emphasize problems with the coarse quality of established shore-line displacement curves with limited precision. Also, which shore-line should be linked to the site? How close to the water was the site? In some cases, terrain and topography helps to determine this. Again,
absence of preserved organic material poses dating problems. Schmitt (1994:247) highlight this problem by saying that if dating can be established then the site could be analyzed with a probable shore-line, and inferences could be made of human landscape use. The dating c 10 300 cal. BP for the core at Torsåker from shore-lines, is enhanced by the fact that the pressure technique present is contemporary with the simulated shore-line (Knutsson & Knutsson, 2012:5).

![Shoreline displacement curve for the area of Gästrikland. From Berglund (2005).](image)

**fig. 2.** Shoreline displacement curve for the area of Gästrikland. From Berglund (2005).

Despite the issues with shore-line dating, the above mentioned core found at 155 m a s l will lay the temporal foundation for this study. It will be assumed that this core is from the pioneering phase in the study area, and that the site is connected to one of the neighboring shorelines to 155 m a s l. Since no “exact” dating is available, and since the regression is substantial, a span of shore-lines will be subject to the study ranging between 145 – 155 m a s l. This represents an archaeologically short period of time, covering c. 300 years.

### 1.7 Predictive modelling

#### 1.7.1 History of predictive modelling

Predictive modelling dates back to the 60s when settlement patterns, by American archaeologists, were believed to be determined by environmental factors (Verhagen, 2007:14). Expansion of predictive modelling is mainly traced back to the rise of processual archaeology in the late 1960s. The quantitative methods and the emphasis on natural determinants for settlement location were the main attractors of the method for the archaeologists of this era (van Leusen et al, 2005:27; Verhagen & Witley, 2012:50-51). The most influential work on which predictive models eventually built on was Gordon Willeys (1953) Prehistoric Settlement Patterns in the Viru Valley, Peru, in which he identified patterns of settlements from a large database.
The first appearance of predictive models in the USA was the outcome of legislation under the National Historic Preservation Act in 1966. This made federal agencies invest in predictive models in order to record historic properties when they need to be destroyed. And these were the first models used in Cultural Resource Management (CRM) (Verhagen & Whitley, 2012:53).

In the early 1970’s, the term predictive modelling first appeared in archaeology. By the end of this decade, the first largely produced predictive models appeared in the USA and this without the application of the still unavailable GIS technology (Verhagen & Whitley, 2012:51).

In the early 90s, GIS applications in archaeology “exploded” (Verhagen, 2007:15), and it was also about this time that European CRM and scholars began to use predictive modelling (Svedjemo, 2014:95).

Today, predictive models are routinely used in CRM planning the USA, Canada, The Netherlands, and to a lesser extent in Germany, the Czech Republic and Australia (Verhagen & Whitley, 2012:53). In Sweden, published studies on predictive modelling are almost absent within the sphere of archaeology. For CRM, I have identified two reports on predictive modelling, the first is a very rudimentary model applied in Laxå for the construction of a wind power park (Knutsson et al., 2013). The second was a pilot study employed by the highways department and was constructed without statistical base (Vägverket, 2006). Two doctoral theses apply predictive modelling; Löwenborg (2010) use Geographically Weighted Regression to explore representativity of burial grounds in Mälardalen in central Sweden and Swedjemo (2014) use Logistic Regression to predict Iron Age settlements on Gotland. One master thesis that apply a quantitative approach has been identified where Jonsson (2014) in a very rudimentary practice try to predict Mesolithic settlements around northern Anundsjöån.

Why the interest for predictive modelling is lacking in Sweden is not clear, but may be because of the same factor as Verhagen & Whitley (2012:53) describes for the UK and France where it is rejected because all archaeological sites cannot be predicted. It might also seem too environmental deterministic.

According to Verhagen & Whitley (2012:54) it has been shown that in cases where predictive models are not used, archaeologists have less opportunity to influence the planning process. Areas that do not accumulate archaeological evidence are seen as less important in the landscape, but how can we judge if there was no other cognitive aspect to the landscape, important to the prehistoric humans? (Verhagen & Whitley, 2012:54)

1.7.2 The concept of Predictive modelling

“Predictive locational models attempt to predict, at a minimum, the location of archaeological sites or materials in a region, based either on a sample of that region or on fundamental notions concerning human behavior.” (Kohler & Parker, 1986:400)

This definition is often recited in introductory chapters to studies with predictive modelling. Furthermore, a starting point for creating Predictive Models must be that certain characteristics of the landscape has attracted humans more than others, mainly as a consequent of their particular economic base (Verhagen 2009, 13). And the basic notion of predictive modelling is that archaeological remains are not randomly distributed in the landscape, but an expression of prehistoric human use of the landscape, and that certain landscape features have attracted humans more than others (Verhagen, 2007:14).
As stated by Warren (1990:91) “Predictive models are devices that make use of existing knowledge to forecast trends or events.”

As described above, the appeal of the method for processual archaeologist was the initial focus on nature as the determining factor for locational choices. Nature determinism has received plenty of critique from archaeologists, and predictive modelling has been criticized for being purely nature deterministic (see Verhagen et al., 2009; Svedjemo, 2014:96). It has been accused of lacking theoretical bases and even “predictive modelers” have emphasized this as a problem. As stated by Warren & Asch (2000): most archaeological predictive models rests on two fundamental assumptions; 1) settlement choices are strongly influenced by characteristics of the natural environment, and 2) that these influencing factors are portrayed in modern maps. However, this is not necessarily the case today. There is a growing number of complex predictive models accounting for cultural expressions, e.g. models based on Bayesian statistics (Verhagen, 2007; Millard, 2005), see section 1.5.2 for more information.

It has been assumed for prehistoric societies that the most important economic component was the environment, and at the same time, most archaeologists agree that the most important transactions in relatively complex societies are not with the environment (Kohler & Parker, 1986:400).

Two major threads of predictive models appeared in the 80’s; one being theoretical trying to identify spatial suitability by eco systemic structures and relationships, and the other based on a quantitative approach making predictions based on statistical analysis of environmental factors. This second approach is often referred to as inductive, but terms as correlative and data driven are also used. The former approach is named deductive, exploratory or theory driven. In deductive methods, archaeological site information is only used for testing purposes, and these models are based on hypothesis of settlement location preferences (Verhagen 2007:14; Verhagen & Whitley, 2012:51). Kohler & Parker (1986:400) emphasize that deductive models may not perform better than inferential models, but they may give more insight into human behavior.

This can further be distinguished with Predictive modelling being divided into two main research areas, where CRM (Cultural Resource Management) constitutes one. The main focus for CRM predictive models is called the predictive approach, identifying locational factors, and is mainly focused at locating archaeological sites. The second main area is academia, and this approach is called interpretive, meaning that the locational choices are tested, to understand why these sites were chosen (Svedjemo, 2014:95).

Even though it seems to be a distinct difference between the two major approaches, and that it is possible to construct purely deductive or inductive models, this rarely occurs. Most models are a mix of the two approaches (Warren, 1990:91).

As described above, predictive modelling has its roots in processual archaeology, with its emphasis on generalization and quantitative “objective” methods, and a lack of interest in the subjective and individual dimensions of archaeology. This has slowed down the development of predictive modelling as a scientific method, since post-processual academic archaeology have lacked in interest in using spatial technology and statistical methods (Verhagen, et al., 2009:19)

Predictive modelling have been criticized for three main issues; statistics, theory and data. The core of the critique is that predictive modelling fails to embrace the complexity of the matter. Uncritical use of statistics, a marginal role of theory concerning human and temporal factors in site placement, and that data is uncritically used despite varying quality; summarize this critique (Verhagen, et al., 2009:20).
The works of Verhagen (2007), Millard (2005) etc., concerned with locating archaeological remains, and in particular with predictive modelling, have focused in areas and time-periods that differ from that of this study. What is unique in Sweden (and other countries at the same latitude) is the isostatic rebound due to the late glacial maximum. The degree of rebound differs depending on where in Sweden you measure. The predictive area in Torsåker, Gästrikland, has experienced intense regression where the highest postglacial shore level is today found at c. 200 m a s l (Berglund, 2005:214). Since this study focuses on hunter-gatherer settlements with a marine focus, this affect has to be accounted for when constructing the model. Shoreline displacements alter the landscape, and what constitutes favorable settlement locations.
2. REFERENCE MATERIAL FOR MODELBUILDING

2.1 The sample
When studying past human cultures, archaeologists are faced with the problem of the dataset studied being only a mere fraction of the culture under study. The entirety of a culture (that we do not have available) are in statistical terms is called population. Since the population of a culture is not known, we work with a subset of the population called a sample. (VanPool & Leonard, 2011:15) Depending on what archaeological remains are under study, this sample is more or less representative of the population. In a perfect world, this subset would be a small scale copy of the entirety of the culture, which would make it easy to make confident inferences. However, this is rarely the case. As Liahaugen (2013) has shown for the Swedish archaeological database, FMIS, there are considerable biases present. Differences in how archaeological remains are registered, is one factor affecting the quality of the sample. There are also great risks that the best preserved archaeological sites can be found in areas with none or little registered remains. This will affect model building since this can be seen as a consequence of exploitation; archaeological remains are found where humans have exploited the landscape (van Leusen, 2005:31).

For this study, the population is set to all middle Mesolithic pioneer settlements in Sweden, and the sample is all excavated sites form the pioneering phase. This would have the potential of providing a rather large sample. However, most of the excavated sites related to the pioneering phase have been dated by means of typology. Consequently, in terms of landscape reconstructions, this gives a large degree of uncertainty whether or not the site actually was active during the time period under investigation. Since most parts of Sweden experience post-glacial effects (regression, transgression), the importance of being able to capture the sites contemporary landscape is of utmost importance, and even more so since the assumption is that the pioneers travelled by sea and that the settlements were shore bound.

This uncertainty has led to the decision to only include the earliest C¹⁴-dated sites in the database, which constitutes the sample. With the courtesy of Per Persson at Oslo University, a database containing 4 907 C¹⁴-dated sites from the Swedish west-coast and Norwegian south was obtained. Sites dating from 8 500 BP and earlier were selected.

2.2 FMIS and Askeladden
The selected sites from the database described above, was searched for in FMIS and Askeladden. FMIS is the Swedish antiquarian database and it was initiated in January 2006, as a digital version of the Swedish ancient monument registry. The main purpose of this digitalization was to make the database more accessible (Liahaugen, 2013:16-17). It is available at www.fmis.raa.se. Askeladden is the Norwegian counterpart, and it was initiated in 2009. The purpose of the database is to be a tool for cultural heritage management (Riksantikvaren, 2015). Askeladden is available at www.askeladden.ra.no.
Fig. 3. The 14 middle Mesolithic sites from the west coast of Sweden and the south of Norway that comprise the site end of the input. The green square is the survey area around Torsåker. © Lantmäteriet, i2012/921
All Swedish sites were easily located, and extracted from FMIS. Finding sites in Askeladden was a bit more problematic. There are many registered posts with the same name, making it difficult to find the sites searched for. Unfortunately, because of this, all selected sites could not be found.

After extracting the sites from FMIS and Askeladden, the database with the input data for the analysis comprised a total of 15 sites. These sites were then visually studied, and compared to available reports to ensure that the extracted area locations were representative of the Mesolithic settlement. The reason for this is that the polygons for area locations sometimes represent large time spans, and could range from early Mesolithic to present day. However, not all of the area can be classified as Mesolithic. Because of this, some of these sites had to be modified, both the size of the polygon to represent the Mesolithic activity area and the surrounding landscape which is described under section 3. However, one of these sites (Norra Råda 32:1) had to be excluded as a consequence of extensive water regulations in the area, making landscape reconstruction almost impossible, or at least too time consuming. The final database consisted of 14 sites.

2.3 Digital Elevation Model

Digital Elevation Models (DEM) plays an important role in this study. For Sweden, a national LiDAR survey was initiated in 2009 by Lantmäteriet to create a New National Elevation-model (NNH in Swedish). This project is now in the finalizing stages. The measurements have been conducted using aerial laser scanning, with a point density up to > 0.5 points/m² (Lantmäteriet, 2012). With this data, it is possible to create high resolution topographical models of the Swedish landscape. This elevation model is delivered as both pre-processed 2 x 2 meter grid, and as raw data containing all points from the scan. For this study the grid 2+ data was used.

In Norway, no LiDAR survey has been conducted at a national scale. LiDAR scans are conducted at small scale, in relation to specific projects, e.g. archaeological excavations (Per Persson, orally). Available DEM for Norway is a 10 x 10 meter grid, which is provided by Kartverket (http://kartverket.no/kart/gratis-kartdata/). In order to perform analysis based on DEM at the same scale, the Norwegian grid was converted into a 2 x 2 m grid to match the Swedish data.

2.4 Variables

At the outset of this study, ambitions where high regarding the inclusion of high qualitative variables. However, as a consequence of poor prior knowledge in statistics, the amount of time invested in acquiring sufficient statistical knowledge was immense. Adding on the time spent in learning the basics in R software, there was simply no time to follow these ambitions through. One original thought was to include qualitative variables in the form of natural harbors, as described by Petterson & Wikell (2013:74). The basic concept is that these harbors provide calm water to land small boats/canoes without capsizing. Furthermore, these locations provide shelter for wind. These attractive attributes of the landscape are effects of the specific topographic features of the site. What would provide more power to a constructed predictive model, is if these attributes could be quantified as a 3D feature with set boundaries, and then digitally searched for in the landscape. Methods for performing this sort of studies in archeology are starting to appear,
but it is still in its infancy. It has been explored by Verhagen & Dragut (2012) using eCognition software, and object based image analysis in the Netherlands.

To predict Mesolithic settlement locations, three variables have been included in the model; slope, distance to coast and soil type. The basic idea is that certain characteristics of these variables are perceived as favorable for settlement location, and consequently ruling out other characteristics of these variables. As described above, the theoretical base for this is that the pioneers in the previously unexplored area around Torsåker, arrives with a set of ideas of what constitutes a good settlement location. This knowledge originates from the area of departure, and is the result of knowledge passed on through previous generations.

Analysis of the reference material in this study will give knowledge of what was perceived as a good settlement location, and consequently knowledge about what information that was passed on through generations. The basic theoretical idea of this study is that these pioneers do not simply adapt to a new environment. They use their prior experience from the area of origin, which enables the analysis in this study, even though the development area and prediction area are many miles apart.

For the analysis, soil maps from SGU and elevation models from Lantmäteriet for Sweden was downloaded from the GET-portal from SLU, and for the Norwegian part of the study it was downloaded from Kartverket.

2.4.1 Slope
One factor that might influence the choice of settlement location is the slope of the site. Slope is derived from a DEM and by calculating the maximal change in degrees at a given location (Connell & Lake, 2006:190). In this analysis, slope was calculated on a 2 x 2 m grid base for both Swedish and Norwegian sites.

2.4.2 Distance to Coast
This variable represents how close to the shore-line the pioneers chose to settle. This is an important factor since pioneers in this time and area are assumed to be of a marine economic base, and always in need of being close to their means of transportation, i.e. their sea vessels (see section 1.6.1). This is the main reason why this variable has been limited to include areas within 100 meters from the shore-line. A central phenomena to this variable is shore-line displacement, which is the interplay between glacio-isostatic uplift, sea-level rise due to melting glaciers and paleogeography (Berglund, 2005:519). As a consequence of immense regression in Gästrikland, pioneers settlements are found at high elevation in the landscape.

With $^{14}$C-dates from the chosen sites and with the aid of shoreline displacement curves, shorelines were generated to represent contemporary landscape for each individual settlement. For the reference material in Norway, shoreline displacement curves from Stabell (1980) and Kjensted (1984) was used. For the Swedish area, a combination of Påsse (2001) and SGU’s map generator was used (SGU, 2015). For the application area in Gästrikland, Berglund (2005) was used.
2.4.3 Soil/Topsoil

A great variety of soil types are present in the landscape, and some are, and have been more habitable than others. This data unfortunately have some inherent problems, mainly with the quality and resolution. Soil maps from SGU are produced in several layers. The layer chosen for this study was JG2, which represent soil types expected to a depth of 0.5 m and with a thickness > 0.5. The soil map comes with a resolution of the smallest unit being 50 meter in diameter (SGU, 2014). This is rather large, and not very well suited for settlement identification since many settlements might be smaller than these units. Also, classification errors are present which can be seen when studying archaeological reports in comparison with soil maps. This has led to reclassification for some sites when needed. Archaeological reports are detailed when describing soil types, and these have been used when reclassification have been necessary. As an example, a large part Skee 1212:1 is according to the soil map situated on bare rock, but in the report it was placed on fine sand. Bare rock was then converted into fine sand to match the actual soil type.

All the different soil categories where converted into interval. The reason for this was problems with nominal values using means for the sites. Some of the sites were spread over more than one category. When extracting only one mean value for each variable, conversion to a ratio scale, is a necessary step. A table provided by SLU (2007; see table 1) where soil types are hierarchically ordered with respect to degree of fraction or grain size was used. With this table, all nominal values were converted into interval values ranging from 0-9. The resulting conversion can be found in appendix 1.

<table>
<thead>
<tr>
<th>Code</th>
<th>Classification</th>
<th>Grain size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Boulder in pit</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>Clatter and stone/Boulder and stone/Flat rock</td>
<td>&gt; 20 mm</td>
</tr>
<tr>
<td>2</td>
<td>Gravel</td>
<td>20 - 2 mm</td>
</tr>
<tr>
<td>3</td>
<td>Coarse sand</td>
<td>2 - 0.6 mm</td>
</tr>
<tr>
<td>4</td>
<td>Medium sand</td>
<td>0.6 - 0.2 mm</td>
</tr>
<tr>
<td>5</td>
<td>Coarse fine sand</td>
<td>0.2 - 0.06 mm</td>
</tr>
<tr>
<td>6</td>
<td>Fine fine sand</td>
<td>0.06 - 0.02 mm</td>
</tr>
<tr>
<td>7</td>
<td>Silt</td>
<td>0.02 - 0.002 mm</td>
</tr>
<tr>
<td>8</td>
<td>Clay/Mud</td>
<td>&lt; 0.002 mm</td>
</tr>
<tr>
<td>9</td>
<td>Turf</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. The classification table used when coding soil categories (author’s translation).
3. QUALITY ENHANCEMENT OF TOPOGRAPHY

3.1 Reconstruction of pre-exploitation topography

One of the main three issues of critique against predictive modelling is the uncritical use of data (Verhagen, 2009:20), see section 1.10. From a visual examination of the sites for the input dataset, the problem with an uncritical use of data became apparent. As can be seen in Fig. 4, a constructed road runs straight through the activity area of the site Morlanda 89:1, situated on the island Orust on the Swedish west coast. This clearly manifests the problem with conducting micro topographic landscape studies of sites that has been altered by modern exploitation.

![Fig. 4. Visualization through hill shade of the topography at Morlanda 89:1 before reconstruction. The causeway running through the area is clearly visible. © Lantmäteriet, i2012/921](image)

When creating a predictive model it is presupposed that the analyzed variables are representative for the period under study. Now, this becomes problematic regarding these excavated sites. Taking into account that many of the thoroughly excavated sites in Sweden have been so due to exploitation, it is fairly reasonable to say that many of these are no longer representative of the landscape at the period of activity. A first visual examination of Morlanda 89:1 (Fig. 4), and its activity area makes clear that it is not the same as during the Mesolithic period under study, as a consequence of this road. And also, it is fairly easy to see that the outcome of a statistical analysis of the sites topography will be heavily affected due to this causeway.
How should sites like Morlanda 89:1 be treated when statistical analysis is supposed to be conducted? In this study, this problem was treated by reconstructing the topography of the site. A search for published literature dealing with this issue of employing topographical reconstruction to enhance the quality of landscape analysis has been undertaken, but without success. However, landscape reconstruction in terms of visual relations between archaeological sites is commonly occurring to make inferences of landscape use (Conolly & Lake, 2010). Since no published literature has been found, this might be one of the first attempts at reconstructing the landscape to enhance the quality of the statistical analysis. Or at least one of few studies that is explicit with this treatment.

Topographical reconstruction was conducted for five of the 14 sites in the sample – Morlanda 89:1, Forshälla 140:1, Skee 1212:1, Lur 438:1 and Lur 438:2. Simply put, the present day topography had to be excluded, and the past topography reconstructed. This was achieved by the following procedure.

![Fig. 5. Layer with contour-lines with an equidistance of 1 m. Area subject to reconstruction has been extracted from the contour-layer (white area). This blank area was then “filled” by connecting contour lines with the same elevation values. © Lantmäteriet, i2012/921](image)

The first step was to generate contour lines from a DEM using the contour tool. Equidistance was set to 1 meter, which made visible the effects of landscape modifications. Second, a polygon was created that covered the total area to be reconstructed. This polygon was used to exclude this area from the DEM, generating a “hole” in the raster. The next step was to create new contour layer with equidistance 1 meter from this modified raster, with the end product being a contour layer with a hole in the middle (Fig. 5).
The next step is the most crucial, and involves the actual topographical recreation. The edit function and the straight segment tool were used to connect contour lines with the same topographical value. Maps displaying the topographical landscape prior to exploitation were used as is exemplified in fig. 6 for Morlanda 89:1. This provided a guide line for how the contour lines should be recreated. When all contour lines have been connected, the topography is reconstructed as a contour map. The final step is to convert this contour map into a raster map, enabling statistical analysis for the recreated topography. The interpolation tool topo to raster was used to obtain this raster map.

In Fig. 7 the reconstructed topography is displayed as a hillshade. It is obvious that the landscape is markedly different from the present day topography (Fig. 4). To test if this “new” topography is more representative of the Mesolithic landscape, a comparative analysis was conducted between constructed sites and sites that have not been affected by modern constructions. This offers a possibility to highlight the change in values as a consequence of this method.

Before making this comparison, analysis is made of the individual sites before and after the reconstruction. Slope at the sites will be compared in this analysis. As can be seen at the site of Morlanda 89:1, a distinct shift in slope values appear for the reconstructed topography (Fig. 8). This shift is further concretized in table 2. Before reconstruction, slope of the site is mainly distributed over five categories, with an emphasis on 4.59 – 9.43 °, comprising 37 % of the total distribution. After reconstruction, the distribution is mainly at the three categories with the least slope values. The two categories with the lowest degree of slope comprise 73 % of the total distribution. This trend for pre- and post-
reconstruction is similar for all sites that underwent this procedure, which can be seen in Fig. 10.

![Fig. 8. Visualization of slope analysis before and after reconstruction. A shift towards lower slope values can be seen (darker green areas) © Lantmäteriet, i2012/921](image)

**Table 2.** Distribution of slope within the site of Morlanda 89:1. “Slope_old” is the distribution before reconstruction, and “slope_new” is the distribution after reconstruction. The table makes visible a shift in concentration for slope values with values concentrated to fewer categories after reconstruction.

<table>
<thead>
<tr>
<th>Slope old</th>
<th>Slope new</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>Count</td>
</tr>
<tr>
<td>1.45 - 4.59°</td>
<td>543</td>
</tr>
<tr>
<td>4.59 - 9.43°</td>
<td>1482</td>
</tr>
<tr>
<td>9.43 - 14.78°</td>
<td>657</td>
</tr>
<tr>
<td>14.78 - 20.39°</td>
<td>792</td>
</tr>
<tr>
<td>20.39 - 26.25°</td>
<td>454</td>
</tr>
<tr>
<td>26.25 - 32.62°</td>
<td>39</td>
</tr>
<tr>
<td>32.62 - 40.02°</td>
<td>16</td>
</tr>
<tr>
<td>40.02 - 40.02°</td>
<td>10</td>
</tr>
<tr>
<td>48.69 - 69.99°</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>4004</td>
</tr>
</tbody>
</table>

The resulting shift in appearance of the site as a consequence of this topographical modification is very interesting. To see if this shift is closer to the characteristics of Mesolithic settlement sites, the altered sites was compared to sites that have not been subject to exploitation. To exemplify this comparison, a site (Bro 195:1) was chosen that has not been affected by exploitation and where the topography has not been actively altered.

As can be seen in table 3, slope values for the site Bro 195:1 are mainly distributed over three classes with the range of 4.59 – 20.39 °, with a modest concentration to the category 9.43 – 14.78 ° with 33 % of the total distribution.

There are similarities between the reconstructed site of Morlanda 89:1, and Bro 195:1. Even though the distribution over categories is not exactly the same, characteristics of the distribution span is similar. For both sites, slope values are concentrated to three categories, even
Table 3. Distribution of slope values within the site of Bro 195:1. As can be seen in the table, slope values are concentrated to degrees between 4.59° - 20.39°.

<table>
<thead>
<tr>
<th>Degrees</th>
<th>Count</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 4.59°</td>
<td>62</td>
<td>0.065331928</td>
</tr>
<tr>
<td>4.59 - 9.43°</td>
<td>235</td>
<td>0.247629083</td>
</tr>
<tr>
<td>9.43 - 14.78°</td>
<td>311</td>
<td>0.327713383</td>
</tr>
<tr>
<td>14.78 - 20.39°</td>
<td>257</td>
<td>0.27081138</td>
</tr>
<tr>
<td>20.39 - 26.25°</td>
<td>75</td>
<td>0.079030558</td>
</tr>
<tr>
<td>26.25 - 32.62°</td>
<td>9</td>
<td>0.009483667</td>
</tr>
<tr>
<td>32.62 - 40.02°</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>40.02 - 48.69°</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48.69 - 85.3°</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Though there is a slight shift in what values these three categories cover. From this illustration, it can be argued that the undertaken topographical reconstruction makes the area more representative of the sites activity period.

To further analyze the effect of this approach a comparison between all reconstructed sites before and after reconstruction was conducted (Fig. 10), and a comparison between reconstructed sites and sites that have not been affected by modern construction was also conducted (Fig. 11).

In the boxplot in Fig. 10, the effect of topographical reconstruction is illustrated through a comparison of the sites before and after reconstruction. What can be seen here is a shift from a wider span of values to a more concentrated distribution. This will affect the outcome of a statistical analysis. If values within a site cover a wide span, and are relatively evenly distributed, precision of the model will be affected. By reconstructing the topography at the sites, distribution is narrowed down to fewer categories, and enhancing possibilities of creating a more representative model of Mesolithic pioneer settlements.

In Fig. 11, a comparison between the reconstructed sites and the sites not affected by modern constructions is visualized by a boxplot. The five sites to the left in the boxplot are the reconstructed sites, and the nine sites to the right are the sites that have not been affected by modern construction. Even though the exact values of the sites are not equal, there is a similar tendency of the span of the distribution. With the exception of Bro and Hovland, slope values for both reconstructed sites and unaltered sites cover a smaller amount of values. This can be seen as evidence of the benefit of this method. It takes the characteristics of the sites affected by modern construction closer to the characteristics of the unaltered sites. In other words, the topographical reconstruction undertaken here, takes the sites expression closer to the sites “original” Mesolithic appearance. Consequently, this affects the result of any statistical analysis conducted on the material, and consequently it also affects the result of the predictive model.
Fig. 10. Comparison between sites before reconstruction (the five to the left) and sites after reconstruction (the five to the right) based on slope values. The affect is distinct in that slope values are more concentrated after reconstruction than before.

Fig. 11. Comparison between reconstructed sites (the five to the left) and sites that have not been subject to reconstruction (the nine sites to the right) on the basis of slope values. Apart from Bro and Hovland, there is homogeneity in the sample. All sites have concentrated slope values even though the actual values differ between sites.
4. STATISTICAL ANALYSIS

4.1 Logistic regression analysis

Identifying appropriate methods for analysis is tricky without the statistical background needed in order to be selective between methods for the specific case and material. However, this study landed at logistic regression with some confidence, even though it is somewhat problematic in regards of the reference material.

Due to the nature of logistic regression, a rather large sample is needed for the researcher to make strong claims about correlation. A second problem is the structure of the reference material. As described earlier, logistic regression analysis needs a binary input (e.g. site or non-site), or a multinomial input (e.g. Mesolithic, Neolithic, iron age), to work. At my disposal was 14 middle Mesolithic sites spread over a rather large area stretching from Gothenburg at the Swedish west coast to the west side of the “ending” of the Oslofjord. These are the sites for the input, but at the outset of the study, there were no non-sites available. There has not been enough time or funding to conduct a proper field survey as described by Warren & Ash (2000) to obtain this end of the data.

The solution to this was to create hypothetical non-sites, see table 4. The area covered by the sites was separated into 14 zones. Means was calculated for the variables (slope, distance, soil) for each of the zones, and the result was used as the non-sites in the input. This binary input was then coded. Sites were given the value of 1, and non-sites the value of 0. Since one of the assumptions of logistic regression is an even number of 1s and 0s, the number of non-sites is equivalent to the number of sites.

The sample is presented in table 4, and comprises a total of 28 objects. All values in the table are means. The reason for treating and using means for sites of the input is simply because the sizes of the excavated and/or defined settlements vary. By ignoring this fact, and for example using a non-random system of points from the area, one would implicitly label the different sites as more settlement or less settlement. A large settlement containing thousands of points and values would take on a dominating role in the analysis, and the small sites would have a marginal effect on the result.

The reasons for a sites being registered as larger, or smaller than others can be a consequence of biases of which we have no control. It might be a consequence of exploitation-process, meaning that how much of a site is excavated is determined by what is to be exploited. Contractors do not pay for extensive excavation just to capture the whole size of the archaeological site. Also, the size of the polygons available via FMIS and Askeladden might be biased. These polygons do not, in many cases, represent the actual settlement. There might be several time-periods bunched into one single polygon, as is the case with Morlanda 89:1 described above. These are a few examples of biases, and there might be more present.
Table 4. The sample used as input for the logistic regression analysis. A total of 14 sites and 14 non-sites were included. All numbers represent means within the sites and non-sites areas calculated.

<table>
<thead>
<tr>
<th>Site Code</th>
<th>Slope (degree)</th>
<th>Soil Distance (meters)</th>
<th>Distance (meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bro</td>
<td>1 14.079504091508</td>
<td>3.79734848484848</td>
<td>9.6773405681819</td>
</tr>
<tr>
<td>Forshälla</td>
<td>1 6.765840592</td>
<td>4.781321185</td>
<td>38.92736019</td>
</tr>
<tr>
<td>H96</td>
<td>1 2.479674783</td>
<td>4.093758746</td>
<td>13.63741336</td>
</tr>
<tr>
<td>H97</td>
<td>1 3.66801498</td>
<td>4</td>
<td>18.57872478</td>
</tr>
<tr>
<td>Hovland</td>
<td>1 13.85057471</td>
<td>4</td>
<td>9.039883315</td>
</tr>
<tr>
<td>Lur1</td>
<td>1 6.18722937</td>
<td>5.755241935</td>
<td>28.84143106</td>
</tr>
<tr>
<td>Lur2</td>
<td>1 7.119217417</td>
<td>5</td>
<td>37.18718582</td>
</tr>
<tr>
<td>Morlanda</td>
<td>1 5.166365527</td>
<td>4.719068413</td>
<td>20.61187243</td>
</tr>
<tr>
<td>Rodbol</td>
<td>1 3.883295195</td>
<td>4</td>
<td>74.61909493</td>
</tr>
<tr>
<td>Skee</td>
<td>1 16.06298744</td>
<td>5.211267606</td>
<td>4.5625957</td>
</tr>
<tr>
<td>Torkop</td>
<td>1 4.925773063</td>
<td>4</td>
<td>51.26286138</td>
</tr>
<tr>
<td>Torstvet</td>
<td>1 4.446278858</td>
<td>7.16</td>
<td>21.83253759</td>
</tr>
<tr>
<td>Vestby</td>
<td>1 6.649096794</td>
<td>4</td>
<td>16.84462831</td>
</tr>
<tr>
<td>Zon 1</td>
<td>0 4.334759223</td>
<td>3.828571429</td>
<td>44.4623728</td>
</tr>
<tr>
<td>Zon 2</td>
<td>0 5.018517884</td>
<td>3.821138211</td>
<td>46.41532351</td>
</tr>
<tr>
<td>Zon 3</td>
<td>0 4.324356175</td>
<td>4.015312787</td>
<td>46.3458681</td>
</tr>
<tr>
<td>Zon 4</td>
<td>0 5.716114936</td>
<td>3.768965517</td>
<td>48.5984432</td>
</tr>
<tr>
<td>Zon 5</td>
<td>0 4.275881631</td>
<td>3.875912409</td>
<td>44.48810997</td>
</tr>
<tr>
<td>Zon 6</td>
<td>0 4.585095399</td>
<td>3.012738854</td>
<td>48.40727855</td>
</tr>
<tr>
<td>Zon 7</td>
<td>0 7.344513018</td>
<td>1.818181818</td>
<td>45.7038544</td>
</tr>
<tr>
<td>Zon 8</td>
<td>0 8.199317873</td>
<td>3.748427673</td>
<td>44.17246426</td>
</tr>
<tr>
<td>Zon 9</td>
<td>0 9.918832837</td>
<td>4.332075472</td>
<td>46.83937636</td>
</tr>
<tr>
<td>Zon 10</td>
<td>0 9.20964042</td>
<td>3.928571429</td>
<td>47.42411444</td>
</tr>
<tr>
<td>Zon 11</td>
<td>0 14.36752643</td>
<td>3.773913043</td>
<td>51.52465518</td>
</tr>
<tr>
<td>Zon 12</td>
<td>0 14.42624899</td>
<td>3.93673469</td>
<td>48.44754402</td>
</tr>
<tr>
<td>Zon 13</td>
<td>0 11.51599081</td>
<td>3.536231884</td>
<td>47.33538659</td>
</tr>
<tr>
<td>Zon 14</td>
<td>0 11.83860634</td>
<td>3.882312457</td>
<td>47.17253432</td>
</tr>
</tbody>
</table>

One of the critical, and most problematic parts of this study, is the use of non-sites. These were obtained by the following procedure. First, the area from the Swedish west-coast to west of Oslofjorden was divided into 14 somewhat equally large zones. Second, a 500x500 m fishnet was created with attached label points that covered the whole area. Third, a buffer zone was created that stretched 100 meters from the simulated shoreline. This represents a possible “settlement zone” since the pioneers have had a need of staying close to their transportation means (see section 1.6.3). Fourth, values from points within this buffer (see Fig. 12) were collected and means was created for each zone, ending up with the non-site end of the database. As a consequence of varying proportions of land within each zone, different amount of points was extracted, and the amount in each of the zones ranged between 115 – 296 points.
This non-site part of the input is not completely random, even though it has a touch of randomness. The fishnet is a regular construction with squares covering the whole area, with a point centered in each square. However, shorelines are irregular to its nature hence the points located within the buffer-zone for each shore-line can be labelled as semi-random, as is exemplified in Fig. 12.

For the first attempt at obtaining these values for the non-sites, a buffer of 200 m was used. However, this was not very successful, since it appeared to be a perfect separation between sites and non-sites and the model was inflated. In other words, the distance means calculated for the non-sites was too far away from the distance means of the settlements and for the analysis to work there need to be some sites and non-sites overlapping.

Statistics and statistical test is summarized in table 5 and 6, additional statistics and tests can be found in appendix 2. In table 5, the predictive model is summarized. The values in the column to the right are the p-values. Simply put, p-values are the probability of obtaining the values obtained, given that the null hypothesis is true. When conducting this sort of hypothesis testing, a null hypothesis and an alternative hypothesis is formulated. The null hypothesis states that there is no relationship between the dependent variable and the independent variables. When constructing predictive models, we want to reject the null hypothesis and claim that there is a relationship between the dependent and the independent variable. The p-value is a measure of statistical significance. A commonly used threshold value is smaller or equal to 0.05 (5 %) for rejecting the null hypothesis and retaining the alternative hypothesis.

As a consequence of not initially analyzing more variables for inclusion in the model, I was “forced” to include the analyzed variables regardless of how well each variable perform.

Looking at the p-values for each variable presented in table 5, it is tempting to keep only
the distance to water variable with a p-value of .0454, which is lower than the 0.05 significance level. Rejection of slope (.3104) and soil (0.1174) as not statistically significant would be based on the generally agreed significance level of .05 at which the null hypothesis is rejected, and limiting the possibility of committing a Type I error (when the null hypothesis have been incorrectly rejected) (Conolly & Lake, 2010:123).

Table 5. Estimates, intercept and statistics for the predictive model.

| Parameter | Estimates | Std. | z value | Pr ( > |z| | I ) |
|-----------|-----------|------|---------|--------|
| Intercept | -14.93682 | 12.78394 | -1.168 | 0.2426 |
| Slope     | -0.20066  | 0.19783 | -1.014 | 0.3104 |
| Soil      | 4.99659   | 3.19113 | 1.566  | 0.1174 |
| Distance  | -0.09425  | 0.04711 | -2.001 | 0.0454 |

Despite the risk of committing Type I error, all three variables were included in the model. Tests of the overall performance of the model are summarized in table 6. The chi-square test is used to see how well the model fits the data and a p-value of 0.00005485861 is considered a good fit.

When looking at model performance for ordinary linear regression analysis, R²-values are used. These values are the percent of variance explained by the model. R² statistics for ordinary linear regression does not apply to logistic regression. Because of this, several pseudo R² have been developed, and the McFadden R² is presented here. The “pseudo” expression is used because these look like ordinary R² values and since they are on a similar scale, ranging from 0 to 1, where higher values represent greater fit. McFadden R²-values for logistic regression as goodness of fit, varies from R²-values for ordinary linear regression. R²-values for logistic regression are considerably lower than for ordinary regression. Values between .2 – .4 represent excellent fit (McFadden, 1977). For the constructed model in this study, the R² McFadden is .58. With the above described excellent fit for values between .2 – .4 for logistic regression, the constructed model is statistically performing very well. However R² values for logistic regression should be used with caution, since they only give a hint of the fit of the model (Svedjemo, 2014:102).

Table 6. Statistical tests for the predictive model.

<table>
<thead>
<tr>
<th>Observations</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Likelihood</td>
<td>-8.2274485</td>
</tr>
<tr>
<td>Pearson's Chi-square (Predictive model)</td>
<td>16.455</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>24</td>
</tr>
<tr>
<td>Pearson's Chi-square (Null model)</td>
<td>38.816</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>27</td>
</tr>
<tr>
<td>Chi-square Test</td>
<td>5.485861e-05</td>
</tr>
<tr>
<td>R² McFadden</td>
<td>0.5760822</td>
</tr>
</tbody>
</table>

A low p-value for the overall model and a high R² McFadden is somewhat in conflict with the high p-values of slope and soil in the model. It is recommended that Pseudo R², like McFadden, is used with caution. The high value of 0.58 does not by itself imply that it is a good model. Other analyzes can be more informative of model performance.
Fig. 13. Regression line (blue line) for each of the included variables with calculated 95% confidence intervals for each variable (darker grey area). The grey area represents possible alternative regression lines based on confidence intervals. For the variable slope it can be concluded that a regression line can be positive, negative or have no regression by drawing alternative regression lines within the dark grey area. For soil the regression line will remain positive no matter how we draw this line, and distance will remain negative on the same basis.
One such concept is confidence intervals. In Fig. 13, each predictor is visualized with scatterplots with a 95% confidence interval (darker grey area). One factor that might give a high R² McFadden is a low sample size. The problem with small sample size is particularly clear with the slope variable. The dark grey area that represents the confidence interval should be interpreted as possible regression lines. This means that for slope, possible scenarios of regression lines are many. Regression for slope could be positive or negative, and it could also be that there is no regression at all, e.g. a flat line. The main reason for this is the three outliers (the sites of Bro, Hovland and Skee) in the top right corner. With a larger sample size, the impact of outliers like these would be smaller and the confidence level of 95% will cover a much smaller interval.

Putting the model to work is fairly simple. From the statistics above, the following equation was formulated and fed into the raster calculator in ArcGIS:

\[
1.0/(1.0 + \exp(-14.93682 - 0.20066 \times \text{"slope"} + 4.99659 \times \text{"soil"} - 0.09425 \times \text{"distance"}))
\]

with the end result being a raster based on probabilities of finding Mesolithic settlements.

4.2 Kvammes Gain

One important concept when determining the performance of the predictive model, is to calculate gain as described by Kvamme (1988), a procedure applied by many others after the publication of his study (e.g. Verhagen, 2007; Svedjemo, 2014). The basic principles of gain are the following: predictive modelling is not only about predicting sites, but also non-sites. It would be easy to predict all sites in a region with 100% accuracy; simply include all of the area. However, this would not be very meaningful. But, “If a model is able to predict 90% of the sites in 50% of the total land area, then something is gained” (Kvamme, 1988:327). In other words, the predictive model must have some gain over a random model without predictive capacity. This gain is defined as

\[
\text{Gain} = 1 - \left( \frac{\text{percentage of total area covered by the model}}{\text{percentage of total sites within model area}} \right)
\]

When gain approaches 1, predictive power increases. If there is a gain approximating 0, then there are very little, or no predictive power. And if gain is negative, a reverse predictive power is present (Kvamme, 1988:329)

A common procedure at the outset of constructing predictive models is to extract a subset from the total dataset that will be used for validation. The area where the predictive model is constructed is called the development area. The area of the subset is called verification area. And finally, the area where the predictive model is to be applied is called the predictive area (Rose & Altschul, 1988:242f). This can also be referred to as split-sampling as described by Kvamme (1988:395).

As a consequence of the small dataset, a subset was not extracted and consequently no verification area could be used, and Kvammes gain was calculated for the development area only. This is not a preferred scenario since independence between the two subsets is necessary in order to see how well the model performs. However, this was followed through as a means of testing the method within the frames of this thesis.

Gain was calculated for three levels for the development area based on the probability scale (Table 7) of the predictive model. The chosen breakpoint was 0.50, which covers a total of
22 % of the total area. Within this area, 76 % of the sites can be found. Using the equation above, a gain value of 0.71 is obtained.

This level was chosen for two reasons. First, the amount of predicted sites at this level is relatively large. Even though the other two levels presented in table 7 has a higher gain, they predict fewer sites. And to be able to predict 76 % of the sites within 20 % of the area is satisfying. The other reason comes from when the model is applied to the prediction area. There was too little area indicated at the other two levels when applying the model based on ten different shorelines, as will be discussed in the following section.

Table 7. Summary of gain calculation with the chosen breakpoint level at 0.5 illustrated in grey.

<table>
<thead>
<tr>
<th>Breakpoint</th>
<th>Area %</th>
<th>Sites %</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.83</td>
<td>0.18394</td>
<td>0.67711</td>
<td>0.72834</td>
</tr>
<tr>
<td>0.67</td>
<td>0.20334</td>
<td>0.73335</td>
<td>0.72272</td>
</tr>
<tr>
<td>0.50</td>
<td>0.21812</td>
<td>0.76387</td>
<td>0.71445</td>
</tr>
</tbody>
</table>
5. APPLICATION AND VALIDATION

5.1 The predictive area

The predictive area is situated in the vicinity of the urban center of Torsåker, in eastern middle Sweden. Torsåker is situated approximately 90 km from the present eastern shoreline, but in the pioneering phase under study, this was an archipelago. The area where the predictive model have been applied, comprise a total of 230 000 ha. Since this study focuses on predicting marine hunter gatherers, the model has been applied to simulated shore-lines ranging from 146 – 155 m a s l, which represents an interpreted pioneering phase approximately 10 300 – 10 000 cal. BP
(see section 1.6.3). And since a buffer of 100 meters from each shoreline have been modelled (see section 2.4.2), the total area covered by the predictive model is 16260 ha, which is only 7% of the total study area (see Fig. 14). As described above, the breakpoint chosen was 0.5 and the area covered by the model at this breakpoint is 7773 ha. For the resulting model area, this covers 15%.

The predictive model was applied to each of the ten shorelines between 146 – 155 m a s l. In Fig 15, the predictive model calculated for 155m shoreline is visualized. This clearly exemplifies the drawbacks of the model when calculated for the individual shorelines. There is simply too much area defined as high probability zones, which is a consequence of small variation mainly in the variables of slope and soil. The solution to this was to aggregate the ten predictive models calculated for the shorelines between 146 – 155 m a s l, by adding them together using the raster calculator. The result is exemplified in Fig. 16, and as can be seen, the precision of the model is better then what can be seen in Fig 15, in that it indicates smaller portions of area as high probability.

![Fig. 15. Extract from the predictive model applied at 155 m shoreline, with a breakpoint of 0.5, to exemplify the low resolution for the individual shore-line models. The background hillshade illustrates the topographical features of the landscape. © Lantmäteriet, i2012/921](image)

It is important to be explicit about these results, and what they represent. It is not a representation of the most suitable locations for Mesolithic settlements. It represents the cumulative effect of intersecting areas between the ten predictive layers. It implies that it has been a suitable location for settlements over a longer period of time since each shoreline represents a certain time in history. It is merely an expression of probabilities being higher for finding middle Mesolithic settlement as a cause of the cumulative effect, instead of the best suited location for a settlement.
Fig. 16. Extract from the final predictive model with all ten “shoreline specific” models combined with a breakpoint of 0.5 for the same area as in fig 15, to exemplify the combining effect. The background hillshade illustrates the topographical features of the landscape. © Lantmäteriet, i2012/921

5.2 Validation of the model

There are several approaches for validating predictive models, and these can be divided into two main groups – statistical and field survey.

When validating a model, it is preferable to test the model using a subset of the data. In other words, before starting the analysis, a part of the dataset is extracted from the main dataset creating a subset. When the model is constructed, it is tested against this subset to see how well it performs in predicting the sites. The percentage of predicted sites by the model is reported and with these numbers the model is evaluated in terms of precision and the amount of sites the model explains for (see section 4.2).

As a consequence of the limited dataset in this study, this could not be executed correctly. Instead, emphasis came to be on the second approach, which is to validate the model by field survey. The purpose of a Predictive Model is to predict previously unknown archaeological sites, features and/or material. No matter how many statistical test you perform for the model, it all comes down to showing that the model actually works in the field. In order for predictive modelling to be a powerful archaeological tool there has to be confidence in model performance in the landscape. Of outmost importance is the models ability to predict site presence/absence. There are several ways of performing field surveys in order to validate the model; trenching, core sampling, test pit sampling, watching briefs and walking surveys (for more detail on the different prospection strategies in predictive modelling see Verhagen, 2007:101).
Test pit sampling was chosen as the method for conducting field survey. It allows for large areas to be covered intrusively, without large costs. It has the drawback of uncovering only small areas in comparison with trenching. The result of this drawback could be that even though a site is correctly located in the terrain, no archaeological remains will be found because the test pits are made in between the areas with archaeological remains. Consequently, the site will not be found due to the drawbacks of the method.

As a consequence of the models binary nature with site or non-site classification in the input, and the problems of no “real” non-sites available for the model construction, not only high probability zones needed to be tested. It was important to test whether the hypothetical non-sites in the model actually are non-sites. For the field survey, the solution was to test both high and low probability zones. This was achieved by walking the stretch covered by the model and making test-pits randomly in both high and low probability zones.

For the field survey, a Garmin eTrex30 GPS was purchased to navigate in the terrain. However, this GPS unit turned out to have some drawbacks. First, it cannot handle Esri shape files. This was solved using DNR GPS application developed by the Minnesota Department of Natural Resources (Minnesota Department of Natural Resources, 2015) which transfers data between Garmin handheld GPS receivers and GIS software. Second, the amount of information this unit can manage is limited. The total area modelled is large, and could not be transferred to the GPS as a shape file. Even though the model was broken down into small pieces, it was still too much information for the GPS unit to handle. The solution to this was to manually create points along the stretch of the model that helped to orientate in the terrain. However, this meant that the full extent of the model could not be located in the terrain via the GPS unit. Consequently this affected the precision of the field survey. Even though the quality of the field survey was affected by this factor, outcome of the survey was acceptable, and most of the test-pits were located within the model area.

Field work was conducted for three days, March 23 – 25 in 2015. Spring had not yet come to the area and in some places there were still deep snow. Frozen ground was present, but for most of the area it was not a problem, except at clear cut areas where the frozen ground went deep. Field work was undertaken with the aid of local amateur archaeologist Detlef Klamke, and also the first day with the aid of doctoral student Michel Guinard.

During the survey, approximately 300 test pits was made, of which 169 were registered using the above mentioned GPS unit. All test pits was not registered because of logistic reasons. One GPS unit was available, and we separated out to cover as much area as possible. Also, if two test pits were made very close, only one was registered since it was believed to be representative of the location registered.

During the survey, no evidence of Mesolithic activity was found, but a total of 10 possible settlement locations were identified with the aid of the model (Fig. 17 & table 9). As an additional variable, natural harbors (as described under section 2.4) were included when searching for these possible settlement locations. The relationship for the total model between high (19 %) and low probability (81 %) zones is quantified in table 8, as is the amount of high (25 %) and low probability (75 %) covered by the possible settlement locations is quantified in table 8. The percentage of high probability covered by the possible settlement locations is higher than the overall percentage of high probability for the entire model. But, it is only a small difference. The possible settlement locations identified from field survey are to 25 % situated in high probability zones, and the overall amount of high probability zones is 19 % for the entire model. This is a
small tendency of settlement locations being located more in high probability zones than in low probability zones. However, 75 % of the area covered by the possible settlement locations is situated in low probability zones, which indicates that there is more work to be done before the quality of the model is satisfying.

Table 8. Relationship between high and low probability zones covered by the total predictive model, and to what extent the possible settlement location areas cover high and low probability zones.

<table>
<thead>
<tr>
<th>Settlement location</th>
<th>Total model</th>
</tr>
</thead>
<tbody>
<tr>
<td>High probability</td>
<td>25 %</td>
</tr>
<tr>
<td>Low probability</td>
<td>75 %</td>
</tr>
</tbody>
</table>

Table 8 shows the relationship between high and low probability zones covered by the total predictive model, and to what extent the possible settlement location areas cover high and low probability zones.

Some problems came to attention during the field survey. The main drawback of the model is the great influence of the soil map. As described above, resolution in these maps are poor. In field, further problems with these maps arose when miss-classifications became evident. Areas classified as fine grained sand can in fact be areas with big boulders of rock, and vice versa. Using variables with poor quality is problematic, and even more so when the total number of variables is few, since each variable then have great influence over the result.

Walking the terrain around Torsåker also make visible issues with the variable slope. It becomes apparent that there is not much variation in the landscape. Long, flat stretches make it hard to differentiate between suitable and not suitable slope for settlement locations since most of the terrain has the right slope characteristics as identified by the statistical analysis.

![Fig. 17. The ten identified possible settlement locations from field survey. The area is situation southwest of Torsåker. © Lantmäteriet, i2012/921](image-url)
Table 9. List of the possible settlement locations identified through field survey with a description of the characteristics of the surrounding terrain.

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
<th>masl min</th>
<th>masl max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obj I</td>
<td>Settlement location, ca 100x40-60 m (NW-SE), consisting of sandy beach terraces. Defined by small stream in N and by big rocks in W, S, E.</td>
<td>152</td>
<td>160</td>
</tr>
<tr>
<td>Obj II</td>
<td>Settlement location, ca 82x10-65 m (NW-SE), sandy morain in SE slope. Defined by rocks and boulder in SW, NW, NE and in SE by old farming ground.</td>
<td>144</td>
<td>157</td>
</tr>
<tr>
<td>Obj III</td>
<td>Settlement location (SW-NE), 73x20-40 m, sandy morain shelf in NE slope. Boulder present in and around settlement area. Defined by road in E, and Rock in SE, NW.</td>
<td>154</td>
<td>158</td>
</tr>
<tr>
<td>Obj IV</td>
<td>Settlement location (NW-SE), 120x40-65 m, sandy shelf in NW slope. Defined by rocks in NE, SE and S.</td>
<td>146</td>
<td>157</td>
</tr>
<tr>
<td>Obj V</td>
<td>Settlement location (NW-SE), 192x55-78 m, sandy morain shelf in NW slope. Defined by rocks in SW, SE and NE.</td>
<td>149</td>
<td>170</td>
</tr>
<tr>
<td>Obj VI</td>
<td>Settlement location (SW-NE), 130x30-60 m, morrain shelf with pockets of fine sand in NE slope. Defined by steep hills in SW, W and NW.</td>
<td>145</td>
<td>164</td>
</tr>
<tr>
<td>Obj VII</td>
<td>Settlement location (W-E), 100x30-45 m, sandy morrain shelf in W slope. Defined by rocks in E and SE.</td>
<td>153</td>
<td>158</td>
</tr>
<tr>
<td>Obj VIII</td>
<td>Settlement location (SW-NE), 50x40 m, sandy morain shelf in SW slope. Defined by rocks in NE.</td>
<td>154</td>
<td>157</td>
</tr>
<tr>
<td>Obj IX</td>
<td>Settlement location (W-E), 50x50 m, Sandy shelf in E slope. Defined by big rocks in N, NE, W and SW.</td>
<td>152</td>
<td>159</td>
</tr>
<tr>
<td>Obj X</td>
<td>Settlement location (NW-SE), 40x35 m, sandy pocket between rocks. Defined by rocks in W, S, E.</td>
<td>153</td>
<td>156</td>
</tr>
</tbody>
</table>
Fig. 18. Possible settlement location exemplified by Obj X. © Lantmäteriet, i2012/921

Fig. 19. The terrain at Obj I. Beach terraces is seen to the right in the photo. Photo: Marcus Asserstam
7. DISCUSSION

7.1 In the context of predictive modelling critique

The following discussion can be divided into four sections. First, it will be discussed how well this thesis meets the three issues (statistics, theory, data quality; see section 1.7.2 pp 23) stressed by Verhagen as the improvements needed for predictive modelling to take the next step. Second, the result of the particular model in this thesis will be discussed. Third, suggested improvements needed to enhance the quality of the model will be put forth. This will be followed by some conclusory thoughts.

One of the major critiques put forward against predictive modeling is the uncritical use of statistics. And the statistical side of this thesis is also problematic. No statistician was engaged in the process of constructing the model. The statistical analysis is the end product of knowledge acquired during the progress of this thesis. The backside to this is that not enough knowledge could be acquired to actively choose between appropriate methods for statistical analysis. This however does not per se mean that the method chosen and the statistical analysis conducted are incorrect. What is does mean is that I have not been in total control of this critical part of the study. The choice of logistic regression for statistical analysis comes from the characteristics of the input data and that other studies with similar data have been applying logistic regression. Therefore it has been followed through with some confidence.

Predictive modelling has been criticized for not engaging theory in the process of model construction, or at least for not emphasizing it more. In terms of theory, I would argue that this thesis have embraced theory and thus meeting the demands of this issue. A clear theory for the origins of the pioneers in the study has been formulated based on the material culture of these pioneers. What is central to this study, is the complex concept of pressure blade technology and the fact not more than one place of origin is likely. Furthermore, the fact that the migration of this technology and the people using it can spatially and temporally be traced is of utmost importance.

Since reference material was not collected in the surroundings of the predictive area, a distinct theory of the origins of the pioneers was necessary. Who were the first people in the area? And where did they come from? The quartz core found at Torsåker that clearly relates to the pressure blade technology, provides the bridge for these questions. The massive ice-sheet obstructed a northern origin and no evidence for an eastern migration has been found. At the same time, the quartz core at Torsåker relates to the west coast and contemporaneous cultural groups in terms of stone tool technology.

Central to the theoretical side of the thesis is these pioneers perception of identity and history. As have been mentioned earlier in this thesis, quartz is not a very well suited raw material for pressure blade technology. Raw material that breaks regularly is preferred when employing this technology. Quartz does not break regularly. Still, this technology was applied to quartz instead of percussion techniques that are better suited for quartz. This can be seen as evidence of
the ambition by pioneers to keep close connections to their area of origins and history, and their ancestry can be seen in the material culture.

This is where this thesis meets the issue of including more theory. In terms of cultural transmission and habitus, prehistoric cognition constitutes the cornerstone of this study. Reproduction of a certain stone tool technology in badly suited raw material is evidence of the application of prior knowledge acquired through past generations. This prior knowledge is consulted and employed when faced with new situations as an expression of habitus. The assumption is that this is not restricted to stone tool technology, and that it can also be seen in for example settlement patterns. As will be discussed further down, similarities in settlement locations is present in the reference material. There are traces of human cognition present in these settlements and it is this perception of suitable locations for settlements that have been quantified in the Swedish west-coast and southern Norway and applied in the predictive area around Torsåker.

Regarding the issue of uncritical use of data, several steps have been taken to ensure high quality of the input data. The first step was to ensure that only settlements contemporaneous with the shore-line dated quartz core from Torsåker were included. This was achieved by using only 14C-dated sites from Swedish west-coast and South Norway for the input dataset. This allows for more precise landscape reconstruction in terms of shoreline displacement since one of the variables in this study is how close to the shoreline that the Mesolithic humans chose to settle.

Second, site polygons downloaded from FMIS were closely examined to see if these polygons were representative of the Mesolithic period under study. In some cases, polygons were modified due to incorrect shapes in comparison with archaeological reports. This could be an effect of polygons being registered in FMIS at an early stage of the archaeological process, before excavations were initiated and consequently before the actual extent of the settlements was known. The extent of these polygons has in some of these cases not been corrected after excavations have finished. Also, sites with only one archaeological identification number can contain remains from a long period of time including more than one phase and archaeological period. In these cases, the Mesolithic part of the polygon was extracted and used for the analysis.

Thirdly, and perhaps one of the most interesting parts of this study, is the use of topographical reconstruction. As have been shown, many sites have been affected by modern exploitation, such as removal of hard rock for road construction. An uncritical use of data could render a statistical analysis of the topographical characteristics of the constructed road, instead of the Mesolithic landscape. This is not representative of a Mesolithic settlement site, and it would heavily affect the result of any statistical analysis. By reconstructing topography for the sites that have been actively altered by modern activity, the original site characteristics were approached. It can never be argued with full confidence that this represents a perfect reconstruction of the sites and that the Mesolithic characteristics are to its fullest extent embraced. However, it takes us closer and perhaps very close to the original site and thus enhancing the quality of the statistical analysis and consequently enhancing performance of the predictive model.

I have shown that there are advantages of performing this sort of data modification. For the sites in this study that have been subject to reconstruction, the result is that the characteristics of the reconstructed sites appear more like the unaltered sites after reconstruction than before reconstruction. One issue with this method is that it is time consuming, and when applying statistical analysis to large datasets, it might not be manageable. However, most likely not all sites in the input data set have been altered beyond recognition by modern activity. Applying the method to the altered sites has positive effects and should be considered when constructing predictive
models, or when conducting other analysis that includes variables based on topographical features.

What is interesting with topographical reconstruction is that no published literature has been found where this method is applied. There can be several reasons for this. It can be that I have simply not been able to identify papers or books that apply this method. It could also be that this method has not been applied prior to this study. Or it could simply be because no author has been explicit about this method when it has been employed.

7.2 Result of the predictive model
Related to the actual results of the model, is how the sample was obtained. The result of model relies on the dependent and independent variables that constitutes the input dataset. In this study, 14 Mesolithic settlements constitute the sites of the input dataset and 14 random locations constitute the non-sites of the dataset. This is a small dataset and most literature recommends large datasets when applying logistic regression. Regardless of what statistical method chosen, there seem to be little consensus regarding sample size, and these questions are still debated in statistical forums. However, even though there is no general agreement on how large the sample should be for logistic regression, it can be concluded that the sample size in this study is smaller than generally recommended.

The non-site end of the sample is problematic since these are not actual non-sites. These sites have been randomly selected and are what could be called hypothetical non-sites, in a worst case scenario many of these are actual sites. This would obviously affect the performance of the model, and documented non-sites are preferable. This could have been obtained by field survey prior to statistical analysis where non-sites could have been documented. Studies have applied this method to obtain this side of the sample, for example Warren & Asch (2000). Time was the limiting factor here, and this pre-survey could not be performed within the frame of this thesis.

As discussed above, the quality of the dependent variable (site and non-site) have been ensured and are at a satisfactory level. At the same time, there are issues with the independent variables (slope, distance to coast, soil). As have been described, great effort was put into acquiring statistical knowledge to be able to conduct preferred analysis with the material at hand. This had consequences when it came to variable selection. There was not enough time to test and include more variables to the model than the three presented above. Also, plenty of time was spent with the topographical reconstructions further limiting explorations of the inclusion of more variables.

The effect of only including three variables in the model is that each variable takes a more dominating role than they would have if more variables had been included, and the negative effect to this model is apparent. Identifying a settlement pattern based on slope, distance to coast and soil is possible, and wherever there might be uncertainties, archaeological reports can be consulted. In other words, the issue is not with the development of the model. The issue is with application of the model.

Resolution of soil maps available for Sweden is of low and of poor quality. In the process of constructing the model it was possible to take control over this limiting factor for the reference material. By consulting archaeological reports and modifying the data, a valid result was obtained. The issue is with application of the model. The smallest units in the soil maps are 50 meters in diameter. These are rather large units considering the variation in the landscape, and
small pockets of for example suitable sand soils are not identified that could have been attractive to the pioneers. Also, there can be miss classifications. There are regional differences to the quality of these maps. In densely populated areas these maps tend to have higher quality in comparison with sparsely populated areas like the study area around Torsåker.

The problem of low quality of the soil map became apparent when field survey was conducted. Miss-classifications were more common than not, and consequently, this is not a reliable variable. If these soil maps would be of higher quality, it would be very useful since the statistical analysis shows that fine grained soil types was preferred by the Mesolithic humans when choosing settlement locations.

There are also issues with the slope variable. As the statistical analysis has shown, gentle slope was preferred by the Mesolithic humans on the Swedish west-coast and Norwegian south. The problem with the area around Torsåker regarding this variable is that the terrain is generally very flat. Large areas can be categorized as being within the frames of the preferred slope, consequently making it hard to distinguish areas suitable for settlements from not suitable areas.

The problem with small variations in the landscape affected the design of the final predictive model. It has been described above that the effect of the model is accumulative since ten predictive models for ten simulated shorelines was added into one model. The two main reasons for this procedure were the lack in precision of the ten individual models, and the rapid regression in the area. None of the individual models was strong enough to clearly separate between high probability and low probability. The models had too much area categorized as high probability, and consequently failing in precision. By adding all models into one model, a model was obtained that pointed to less area with high probability, and resulting in smaller areas with archaeological potential. Also, since regression at the time under study was rapid, it resulted in a fast shore-line displacement process. The ten shorelines included in the model cover 250 – 300 years, which are supposed to capture the whole possible pioneering phase in the area.

The accumulative effect of the model is problematic. At one hand, the model points to areas where there could have been longer periods of Mesolithic activity, consequently increasing chances of finding settlement sites through artefacts since more material might have accumulated over time. But on the other hand, the model might ignore many potential settlement sites through this procedure. Optimal settlement sites that have only been available for a short period of time due to the regression process will be discriminated from the model. Unfortunately, the model fails at including these sites and therefore lack in precision.

However, when validating the model, ten possible settlement locations were identified, and most of these are to some extend situated in high probability zones since these areas overlap both high and low probability zones. These sites were tested by using test-pit sampling without any human remains being found. This does not mean that these places were not used for settlement. The period under study need to be taken into consideration. We do not know how many humans entered the area in this pioneering phase. It is not unlikely to assume that it was small groups exploring these new lands. A combination of the rapid regression with each potential site being a potential site for a short period of time, and a small population in the initializing stage of the area results in small quantities of archaeological remains. These are factors that limit the potential of finding human remains from the pioneering phase, and might be one reason why no artefacts were found during field survey. Also, more time should have been invested in testing the model in field, something that might have led to finding remains of the pioneers.
What can be concluded is that there is a pattern visible for the settlements in this study based on the three variables incorporated in the model. This pattern can be summarized as 1) Settlements are located on low slope ground with an average value of approximately seven degrees, 2) sandy soil has been preferred and 3) settlements have been located close to the shoreline with an average distance at 27 meters. However, it has not been confirmed that this settlement pattern is valid for Mesolithic pioneers. In order to draw this conclusion, sites need to be confirmed using the predictive model, and no such finds was made during fieldwork.

Even though there have been obvious problems with application of the model as have been discussed above, this analysis has taken us closer to a measurable settlement pattern for middle Mesolithic cultural groups, and closer to what these humans perceived as a good settlement location.

7.2 Model development

Constructing predictive models is always a work in progress. New technical solutions, new statistical solutions, new input data available, are some examples of why no definite model can be constructed. For the model in this study, a few suggestions for development will now be presented.

First, the need of excluding the soil variable. As have been presented above, the resolution of this data does not meet demands for locating settlements on a micro scale level. This is unfortunate, since the outcome of the statistical analysis emphasize the importance of a specific type of soil for middle Mesolithic human settlement location choice. Until new high resolution soil maps are available, this data is not recommended to be included in similar studies.

Further development of the model could be accomplished by feeding the model with more variables. Besides the fact that more variables important for choosing settlement location can be found, it would also decrease the impact of individual variables to the model. It would also provide the opportunity of dropping variables from the model that are not statistically significant.

Of particular interest to predictive models regarding more variables for early and middle Mesolithic settlements, is the progress of software enabling identification of 3D-objects such as eCognition that have been described earlier in this thesis. Since certain topographical features have been crucial for settlement location choice, this might be the key to constructing accurate predictive models for this type of human remains. Hopefully there will soon be user friendly applications providing this technique. If not, it would be interesting to search for other solutions of quantifying these characteristics within current GIS applications. This would make up an interesting base for future projects.

One limiting factor that the model in this study has suffered from is the small input dataset. As have been described above, logistic regression needs a large sample, and the sample size in this study is not large enough. During the field survey for validating the model, I engaged in a conversation with doctoral student Michel Guinard regarding the structure of the database. It came to my attention that it would have been possible to include some of the sites that I had rejected. These were sites that did not have $^{14}$C-dates, which was one of the criteria that had to be met for settlements to be included in the model construction process. The reason for this was to capture the sites contemporary environment, mainly in terms of shore-line displacement, in times of rapidly changing landscapes due to regression. However, in some areas, regression has been
modest and it might be possible to capture a contemporary shore-line for sites even though no $^{14}$C-dates are available.

7.3 Conclusory thoughts

Quantifying human actions is somewhat tricky. Identifying patterns of human behavior when behavior can be irrational is hard. However, as have been shown in this study, there are patterns of human activity to be found and quantified, even though these patterns are within a certain span of variation. Possibilities of finding and using visible patterns in the archaeological material are the advantage of predictive modeling, and what it can contribute to the archaeological community. It has the ability, if used correctly, to identify areas with archaeological potential. This could be a useful tool for rescue archaeology in Sweden as it has been in other countries for decades. And as Verhagen and Whitley (2012:54) states that by using predictive maps, archaeologist have the opportunity of influencing the exploitation process at an early stage.

It is important to be aware of the problems that come with predictive modelling. One issue is the fact that predictive models might be a tool to reproduce what is known at present, meaning that the constructed model is based on what has been found so far. The whole picture might not be available to us. If predictive models are not treated with caution, they might in fact become truths, obstructing our way of finding new sites or archaeological remains that lie outside the patterns that are visible today.

Despite apparent drawbacks of the method, I am positive that this method has plenty to contribute to Swedish archaeology. Hopefully, predictive modelling will soon be incorporated into Swedish archaeology, guiding and aiding archaeologist in their quest for knowledge about past human societies.
8. SAMMANFATTNING

Två huvudsakliga syften har formulerats i denna master uppsats 1) att undersöka möjligheterna att skapa en prediktiv modell med hjälp av ArcGIS och statistikprogrammet R för att lokalisera tidigare okända arkeologiska platser, i detta fall Mesolitiska pionjärbosättningar, och 2) att undersöka teorin om en väst-östlig migration in i södra Norrland. För att undersöka detta har två frågeställningar formulerades; finns det identifierbara bosättningsmönster för Mesolitiska pionjärer? Och vad är möjligheterna för att analysera dessa mönster genom GIS applikationer för att skapa prediktiva modeller?

Den teoretiska utgångspunkten har varit att pionjärerna i södra Norrland har sitt ursprung på den svenska västkusten och södra Norge. Denna koppling har antagits till följd av den spånkärna i kvarts som hittats i närheten av Torsåker, Gästrikland (vilket även utgör undersökningsområdet) och som har strandlinjedaterats till en pionjär fas i området (c. 10 300 – 10 000 cal. BP). Denna kärna uppvisar spår av en avancerad tryckteknik för tillverkning av verktyg i sten och som återfinns vid samma tid på den svenska västkusten och i södra Norge. Kvarts är ett råmaterial som är dåligt lämpat för denna teknik då den spricker oregelbundet. Att föredra är material som spricker regelbundet, exempelvis flinta. För kvarts är diverse slagtekniker en bättre lösning. Trots detta har alltså denna tryckteknik tillämpats på kvarts.

Genom detta är det då möjligt att göra en koppling till ett potentiellt ursprungsområde där det har varit av vikt för dessa pionjärer att behålla traditionen av verktygstillverkning, istället för att anpassa sig till ett lokalt råmaterial. Kulturell Transmission och Habitus är de teoretiska ramverk som anlagts i uppsatsen. Kulturell transmission innebär att kunskap och värderingar förs vidare från en individ till en annan och är en effekt av mänsklig interaktion och social inlärning. En viktig del i denna teoribildning är de kunskap och värderingar som förvärvas tidigt i livet och som är mer motståndliga mot förändringar. Kopplat till denna teori är begreppet Habitus, som ses som de kognitiva strukturer som vi bär med oss, och även om vi inte är helt styrd av dessa, så är det något som vi hela tiden förhåller oss till. Detta innebär exempelvis att när vi ställs inför nya utmanande situationer så konsulteras Habitus undermedvetet, varefter beslut fattas. Att tryckteknik har applicerats på kvarts i södra Norrland kan således ses som ett uttryck för kulturell transmission och habitus. Dessa kognitiva strukturer blir tydliga då beslut har fattats i detta nya landskap utifrån medförd kunskap och att kopplingen till ursprungsgruppen vidhålls.

Med detta som utgångspunkt har i denna uppsats ett antagande gjorts att detta även är applicerbart för val av boplatssläge. Pionjärerna som anlände till området för med sig inlärda kunskap om vad som utgör ett lämpligt boplatssläge och har sedermera tagit dessa lägen i anspråk.

För att undersöka detta har ett referensmaterial samlats in från den svenska västkusten och södra Norge i syfte att försöka identifiera ett bosättningsmönster som kan tillämpas i undersökningssammanhang. Totalt uppgår antalet analyserade boplatser till 14 stycken som är samtida med den pionjärer fasan vid Torsåker. Detta har säkerställts genom att endast inkludera 14C-daterade boplatser. För att säkerställa kvalitén på referens materialet har topografiska rekonstruktioner genomförts för de boplatser där modern exploatering har modifierat landskapet, exempelvis genom vägkonstruktioner.
En logistisk regressionsanalys har sedan genomförts på materialet genom R utifrån variablene lutning, jordmån och närhet till kust för att skapa den prediktiva modellen. Analysen visar att en svag lutning och sandigt underlag, samt ett genomsnittligt avstånd till kust om 27 meter har föredragits. Resultatet applicerades sedan på undersökningsområdet runt Torsåker utifrån tio simulerade strandlinjer (155 – 145 m ö h) då landhöjningen har inneburit en snabbt landskapsförändrande kraft i området, vilket resulterade i tio prediktiva modeller. Dessa adderades sedan ihop för att erhålla den slutliga modellen. Effekten av detta är att den presenterade modellen i uppsatsen har ett ackumulativt uttryck där lägen indikerats utifrån att de varit lämpliga boplatsslägen under längre tid. Modellen i sig lyftes således inte fram ytor som varit bra boplatsslägen under kortare tid, vilket är en begränsning i modellen.


Utifrån uppställda syften och frågeställningar i uppsatsen kan följande konstateras. Ett bosättningsmönster av mesolitiska pionjärer har inte kunnat påvisas, även om ett mesolitiskt bosättningsmönster har identifierats och kvantifierats utifrån referensmaterialet. För att med säkerhet kunna säga att detta bosättningsmönster är rådande för pionjärbosättarna hade detta behövt styrkas med hjälp av fynd vid fältarbete, något som inte framkom. Gällande den andra frågeställningen har detta mönster kunnat analyseras och kvantifieras med hjälp av GIS-applikationer och applicerats i undersökningsområdet, om än med vissa tillkortakommanden.


Berglund, M. 2010. Litorina sea shore displacement and pollen analytical indications of forest succession during the Mid-Holocene in Gästrikland, east central Sweden. *GFF* 132:3-4, 213-226


Burdikiewicz, J. M., 2011. Late Glacial hunter-gatherer reactions to the Younger Dryas cooling event in the southern and eastern Baltic regions of Europe. *Quaternary International* 242: 302-312


56


Eerkens, Jelmer W., Lipo, Carl P. 2007 Cultural transmission theory and the archaeological record: providing context to understand variation and temporal changes in material culture. *Journal of Archaeological Research* 15:239-274


59


www

The R Foundation. 2015. [http://www.r-project.org/about.html](http://www.r-project.org/about.html) (2015-06-12)


Minnesota Department of Natural Resources. 2015. [http://www.dnr.state.mn.us/mis/gis/DNRGPS/DNRGPS.html](http://www.dnr.state.mn.us/mis/gis/DNRGPS/DNRGPS.html) (2015-06-12)


Unpublished

SGU. 2014. Accompanying document with soil map order.
10. ILLUSTRATIONS

**Fig. 1.** Illustration of the postglacial western Post-Ahrensburgian migration, and the eastern Butovo migration. Illustration from Knutsson & Knutsson (2012:3).

**Fig. 2.** Shoreline displacement curve for the area of Gästrikland. From Berglund (2005).

**Fig. 3.** The 14 middle Mesolithic sites from the west coast of Sweden and the south of Norway that comprise the site end of the input. The green square is the survey area around Torsåker. © Lantmäteriet, i2012/921

**Fig. 4.** Visualization through hill shade of the topography at Morlanda 89:1 before reconstruction. The causeway running through the area is clearly visible. © Lantmäteriet, i2012/921

**Fig. 5.** Layer with contour-lines with an equidistance of 1 m. Area subject to reconstruction has been extracted from the contour-layer (white area). This blank area was then “filled” by connecting contour lines with the same elevation values. © Lantmäteriet, i2012/921

**Fig. 6.** Map from the archaeological report of Morlanda 89:1 where topography prior to landscape modification is represented. This topography illustrated as contour lines was used as a template when reconstructing the topography (From Nordqvist, 2005).

**Fig. 7.** Reconstructed topography for the site of Morlanda 89:1 visualized with a hillshade. © Lantmäteriet, i2012/921

**Fig. 8.** Visualization of slope analysis before and after reconstruction. A shift towards lower slope values can be seen (darker green areas) © Lantmäteriet, i2012/921

**Fig. 9.** Topography at the site of Bro 195:1 illustrated by a hillshade. © Lantmäteriet, i2012/921

**Fig. 10.** Comparison between sites before reconstruction (the five to the left) and sites after reconstruction (the five to the right) based on slope values. The affect is distinct in that slope values are more concentrated after reconstruction than before.

**Fig. 11.** Comparison between reconstructed sites (the five to the left) and sites that have not been subject to reconstruction (the nine sites to the right) on the basis of slope values. A part from Bro and Hovland, there is homogeneity in the sample. All sites have concentrated slope values even though the actual values differ between sites.

**Fig. 12.** Creation of non-site values exemplified. The points extracted were the points that fell within the 100 meter buffer zone, and these are symbolized as green points. © Lantmäteriet, i2012/921

**Fig. 13.** Regression line (blue line) for each of the included variables with calculated 95 % confidence intervals for each variable (darker grey area). The grey area represents possible alternative regression lines based on confidence intervals. For the variable slope it can be concluded that a regression line can be positive, negative or have no regression by drawing alternative regression lines within the dark grey area. For soil the regression line will remain positive no matter how we draw this line, and distance will remain negative on the same basis.

**Fig. 14.** The study area and the model area based on ten simulated shorelines. The model area in grey is a representation of the area covered by the model for ten simulated shore-lines with a buffer of 100 meters for each shore-line. Torsåker is situated in the middle of the area. © Lantmäteriet, i2012/921

**Fig. 15.** Extract from the predictive model applied at 155 m shoreline, with a breakpoint of 0.5, to exemplify the low resolution for the individual shore-line models. The background hillshade illustrates the topographical features of the landscape. © Lantmäteriet, i2012/921

**Fig. 16.** Extract from the final predictive model with all ten “shoreline specific” models combined with a breakpoint of 0.5 for the same area as in fig 15, to exemplify the combining effect. The background hillshade illustrates the topographical features of the landscape. © Lantmäteriet, i2012/921

**Fig. 17.** The ten identified possible settlement locations from field survey. The area is situation southwest of Torsåker. © Lantmäteriet, i2012/921

**Fig. 18.** Possible settlement location exemplified by Obj X. © Lantmäteriet, i2012/921

**Fig. 19.** The terrain at Obj I. Beach terraces is seen to the right in the photo. Photo: Marcus Asserstam

**Table 1.** The classification table used when coding soil categories.
Table 2. Distribution of slope within the site of Morlanda 89:1. “Slope_old” is the distribution before reconstruction, and “slope_new” is the distribution after reconstruction. The table makes visible a shift in concentration for slope values with values concentrated to fewer categories after reconstruction.

Table 3. Distribution of slope values within the site of Bro 195:1. As can be seen in the table, slope values are concentrated to degrees between 4.59° - 20.39°.

Table 4. The sample used as input for the logistic regression analysis. A total of 14 sites and 14 non-sites were included. All numbers represent means within the sites and non-sites areas calculated.

Table 5. Estimates, intercept and statistics for the predictive model.

Table 6. Statistical tests for the predictive model.

Table 7. Summary of gain calculation with the chosen breakpoint level at 0.5 illustrated in grey.

Table 8. Relationship between high and low probability zones covered by the total predictive model, and to what extent the possible settlement location areas cover high and low probability zones.

Table 9. List of the possible settlement locations identified through field survey with a description of the characteristics of the surrounding terrain.
## 11. APPENDIXES

### Appendix 1. Re-classification tables

<table>
<thead>
<tr>
<th>Description</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvsediment</td>
<td>4</td>
</tr>
<tr>
<td>Berg</td>
<td>1</td>
</tr>
<tr>
<td>Blockmark</td>
<td>1</td>
</tr>
<tr>
<td>Flygsand</td>
<td>4</td>
</tr>
<tr>
<td>Fyllning</td>
<td>2</td>
</tr>
<tr>
<td>Glacial grov-silt--finsand</td>
<td>5</td>
</tr>
<tr>
<td>Glacial lera</td>
<td>8</td>
</tr>
<tr>
<td>Glacial silt</td>
<td>6</td>
</tr>
<tr>
<td>Gyttja</td>
<td>8</td>
</tr>
<tr>
<td>Gyttjelera</td>
<td>8</td>
</tr>
<tr>
<td>Isälvssediment</td>
<td>4</td>
</tr>
<tr>
<td>Isälvssediment, grus (1)</td>
<td>2</td>
</tr>
<tr>
<td>Isälvssediment, sand</td>
<td>4</td>
</tr>
<tr>
<td>Kärrtorv</td>
<td>9</td>
</tr>
<tr>
<td>Klapper</td>
<td>1</td>
</tr>
<tr>
<td>Lera</td>
<td>8</td>
</tr>
<tr>
<td>Lera--silt</td>
<td>8</td>
</tr>
<tr>
<td>Morän</td>
<td>4</td>
</tr>
<tr>
<td>Mossetorv</td>
<td>9</td>
</tr>
<tr>
<td>Oklassat område, tidvis under vatten</td>
<td>9</td>
</tr>
<tr>
<td>Postglacial finlera</td>
<td>6</td>
</tr>
<tr>
<td>Postglacial finsand</td>
<td>5</td>
</tr>
<tr>
<td>Postglacial lera</td>
<td>8</td>
</tr>
<tr>
<td>Postglacial sand</td>
<td>4</td>
</tr>
<tr>
<td>Rösberg (1)</td>
<td>1</td>
</tr>
<tr>
<td>Sandig morän</td>
<td>4</td>
</tr>
<tr>
<td>Sandig-silt morän (1)</td>
<td>5</td>
</tr>
<tr>
<td>Silt</td>
<td>7</td>
</tr>
<tr>
<td>Svämsediment, grov-silt-finsand (1)</td>
<td>5</td>
</tr>
<tr>
<td>Svämsediment, ler--silt</td>
<td>8</td>
</tr>
<tr>
<td>Svämsediment, sand</td>
<td>4</td>
</tr>
<tr>
<td>Svallsediment, grus</td>
<td>2</td>
</tr>
<tr>
<td>Torv</td>
<td>9</td>
</tr>
<tr>
<td>Torv, tidvis under vatten</td>
<td>9</td>
</tr>
<tr>
<td>Urberg</td>
<td>1</td>
</tr>
<tr>
<td>Vatten</td>
<td>9</td>
</tr>
<tr>
<td>Lossmassetype JORDART (Norway)</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Morenemateriale, sammenhengende dekke, stedvis med stor mektighet</td>
<td>3</td>
</tr>
<tr>
<td>Morenemateriale, usammanhengende eller tynt dekke over berggrunnen</td>
<td>4</td>
</tr>
<tr>
<td>Randmorene/randmorenebelte</td>
<td>4</td>
</tr>
<tr>
<td>Breelvavsetning (Glasifluvial avsetning)</td>
<td>2</td>
</tr>
<tr>
<td>Hav- og fjordavsetning, sammanhengende dekke, ofte med stor mektighet</td>
<td>4</td>
</tr>
<tr>
<td>Marin strandavsetning, sammanhengende dekke</td>
<td>4</td>
</tr>
<tr>
<td>Hav- og fjordavsetning, usammanhengende eller tynt dekke over berggrunnen</td>
<td>4</td>
</tr>
<tr>
<td>Elve- og bekkeavsetning (Fluvial avsetning)</td>
<td>3</td>
</tr>
<tr>
<td>Vindavsetning (Eolisk avsetning)</td>
<td>5</td>
</tr>
<tr>
<td>Forvitringsmateriale, ikke indelt etter mektighet</td>
<td>2</td>
</tr>
<tr>
<td>Forvitringsmateriale, sammanhengende dekke</td>
<td>2</td>
</tr>
<tr>
<td>Forvitringsmateriale, usammanhengende eller tynt dekke over berggrunnen</td>
<td>2</td>
</tr>
<tr>
<td>Skredmateriale, ikke indelt etter mektighet</td>
<td>2</td>
</tr>
<tr>
<td>Skredmateriale, sammanhengende dekke, stedvis med stor mektighet</td>
<td>2</td>
</tr>
<tr>
<td>Skredmateriale, usammanhengende eller tynt dekke over berggrunnen</td>
<td>2</td>
</tr>
<tr>
<td>Torv og myr (Organisk materiale)</td>
<td>9</td>
</tr>
<tr>
<td>Humusdekke/tynt torvedekke over berggrunn</td>
<td>7</td>
</tr>
<tr>
<td>Usammanhengende eller tynt løsmassedekke over berggrunnen, flere løsmassetyper i tett veksling</td>
<td>3</td>
</tr>
<tr>
<td>Fyllmasse (antropogent materiale)</td>
<td>1</td>
</tr>
<tr>
<td>Bart fjell</td>
<td>4</td>
</tr>
</tbody>
</table>
Appendix 2. Summary of statistics

**R output**

Call:
glm(formula = bpl$boplats ~ bpl$lutning + bpl$jordart +
    bpl$vatten,
    family = binomial)

Deviance Residuals:

Min      1Q     Median      3Q     Max
-1.30896 -0.51150  -0.00201   0.24694  2.37633

Coefficients:

Estimate Std. Error z value Pr(>|z|)
(Intercept)  -14.93682   12.78394  -1.168  0.2426
bpl$lutning  -0.20066    0.19783  -1.014  0.3104
bpl$jordart   4.99659   3.19113   1.566  0.1174
bpl$vatten  -0.09425    0.04711  -2.001  0.0454 *

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 '.' 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 38.816  on 27  degrees of freedom
Residual deviance: 16.455  on 24  degrees of freedom
AIC: 24.455

Number of Fisher Scoring iterations: 7

**Statistical tests**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>28</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-8.2274485</td>
</tr>
<tr>
<td>Pearson's Chi-square (Predictive model)</td>
<td>16.455</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>24</td>
</tr>
<tr>
<td>Pearson's Chi-square (Null model)</td>
<td>38.816</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>27</td>
</tr>
<tr>
<td>Chi-square Test</td>
<td>5.485861e-05</td>
</tr>
<tr>
<td>R2 McFadden</td>
<td>0.5760822</td>
</tr>
<tr>
<td>R2 Maximum Likelihood</td>
<td>0.5500503</td>
</tr>
<tr>
<td>R2 Crag and Uhler's</td>
<td>0.7334004</td>
</tr>
</tbody>
</table>