Considering Uncertainty in Archaeological Predictive Modelling: A Case Study in Southern Rhineland-Palatinate (Germany)

Silke BOOS\textsuperscript{a}, Sabine HORNUNG\textsuperscript{b}, and Hartmut MUELLER\textsuperscript{a1}

\textsuperscript{a}Mainz University of Applied Sciences, Germany
\textsuperscript{b}Mainz University, Germany

\textbf{Abstract.} Most archaeological predictive models lack significance because fuzziness of data and uncertainty in knowledge about human behaviour and natural processes are hardly ever considered. One approach to cope with such uncertainties is utilization of methods, which base on approaches of the probability theory like Bayes Theorem or Dempster-Shafer-Theory. In our case study we analyze an area of 50 km\textsuperscript{2} in southern Rhineland Palatinate (Germany) near the Celtic oppidum “Hunnenring” by use of Dempster-Shafer’s theory of evidence for predicting spatial probability distribution of archaeological sites. This technique incorporates uncertainty by assigning various weights of evidence to defined variables, in that way estimating the probability for supporting a specific hypothesis (in our case the hypothesis presence or absence of a site). Selection of variables for our model relies both on assumptions about settlement patterns and on statistically tested relationships between known archaeological sites and environmental factors. The modelling process is conducted in a Geographic Information System (GIS) by generating raster-based likelihood surfaces with a cell resolution of 10 m for the six selected variables ‘distance to water’, ‘distance to road network’, ‘distance to graves’, ‘slope’, ‘landforms’ and ‘geology’. The corresponding likelihood surfaces are aggregated to a final weight of evidence surface, which results in a likelihood value for every single cell of being a site or a non-site. Finally the result is tested against a database of known archaeological sites for evaluating the gain of the model. To address the high potential of soil erosion processes of the low mountain parts of our study area a model was developed which allocates erosion and deposition zones to those areas. The combination of this model with the predictive model yields a more differentiated estimation, especially with regard to the suspected potential of archaeological remains being conserved until present.

\textbf{Keywords.} Geographic Information System, Predictive Modelling, Dempster-Shafer-Theory

\textbf{Introduction}

Since the end of 2006 the area of the Celtic oppidum “Hunnenring”, which is situated on the ridge of the Dollberg Mountain in the border region of the Federal States of Rhineland Palatinate and Saarland (Germany), is examined regarding to its outstanding prominence in Celtic times. Besides the monumental construction of the oppidum, which is nowadays still reflected in its northern wall, preserved up to a height of 10 m, various sources indicate that the region used to be a centre of supra-regional importance in Celtic times [1]. Several campaigns of archaeological and geophysical prospection in the region helped to extend our knowledge on several archaeological sites in the region. Due to a wide range of important monuments and sites from both Celtic and Roman times, the area around the Hunnenring has long played a prominent part in outlining models of social and economic development. In order to support targeted prospection an archaeological predictive model was developed. This model incorporates influences from the natural environment as well as certain assumptions on settlement structures and landuse in Celtic and Roman times in order to predict the spatial distribution of different settlement preferences in our study area. For that purpose the theory

\begin{footnote}
\textsuperscript{1} Prof. Dr. Hartmut Mueller, Institute for Spatial Information and Surveying Technology, Lucy-Hillebrand-Str. 2, D-55128 Mainz, E-Mail: mueller@geoinform.fh-mainz.de
\end{footnote}
Silke BOOS, et al.: Considering Uncertainty in Archaeological Predictive Modelling

of Dempster-Shafer, a methodology derived from probability theory, is used for modelling the different influencing factors. In this context a Geographic Information System (GIS) provides for good capabilities to build the model and to visualise the results. For evaluation the model results are cross-checked against a database containing the locations of already known archaeological sites.

1. Predictive modelling

1.1 Methodology

Archaeological Predictive Modelling is a methodology, which was developed in the late 70ies in the USA in context with governmental land management projects [2]. This technique bases upon the assumption, that the choice of settlement formation and burial grounds of former societies is closely associated with natural factors and the influence of socio-cultural aspects. The objective of Predictive Modelling bases on this hypothesis by considering the spoken to incitements and enables in that way exposing significant areas with a specified probability for locating archaeological remains. Predictive Modelling can be conducted by use of certain methods, in a wide range of complexity from simple additive methods up to multivariate regression analysis [3], [4]. Presently methods of Fuzzy Logic [5] or probabilistic approaches [6], [7] are increasingly used in the field of archaeological predictive modelling. They allow for incorporating uncertainty or fuzzy knowledge about human behaviour in the model. For modelling historic settlement processes considered in this study the algorithm of Dempster-Shafer is used.

1.2 Dempster-Shafer’s Theory of Uncertainty

The Dempster–Shafer theory [8], [9] is a mathematical theory, which uses evidences instead of probabilities for modelling uncertainty. The quintessence of Dempster-Shafer’s theory can be summarized in a way, that each domain of knowledge implies uncertainty and that the complement of a hypothesis must not automatically be assigned to its negation, but has to be allocated to the factor uncertainty. An aggregation rule is used to include numerous pieces of information (evidence) with varying weight into a decision making process thus supporting or excluding defined hypothesis. A model based on Dempster-Shafer’s theory can be expressed mathematically in the following way:

- The model is composed of a set of hypothesis $H = h_1...h_n$, which comprises all possible and mutual excluding outcomes and all of their combinations. This set is called Frame of Discernment ($\Omega$).
- The theory of evidence assigns a belief mass $m(A)$ (also called Basic Probability Assignment -> BPA) to each element of a set of interest $A$, which is a given member of the Frame of Discernment, and expresses the proportion of all relevant and available evidence that supports the claim that the actual state belongs to $A$ but to no particular subset of $A$. The BPA fulfils the following two conditions:

$$m(\emptyset)=0$$
$$\sum_{A \subseteq \Omega} m(A)=1$$

- Belief $Bel(A)$ for a set of interest $A$ is defined as the sum of all the masses of subsets $m(B)$ of the set of interest $A$:

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

- Another important quantity is plausibility $Pl(A)$, which is the sum of all the masses of the sets $B$ that intersect the set of interest $A$. Consequently this quantity describes any belief in spaces, which are consistent with a specific hypothesis:
The difference between *belief* and *plausibility* is another important quantity of the Dempster-Shafer formalism and is referred to as *belief interval*. It represents in this way the range of maximum uncertainty.

By use of Dempster’s Rule of aggregation the single belief sets can be combined pairwise and in that way it is possible to aggregate them to a total belief:

\[
m(Z) = \sum m(X) \cdot m(Y) \quad \text{if} \quad (X \cap Y) = Z
\]

\[
m(Z) = 1 - m(X) \cdot m(Y) \quad \text{if} \quad (X \cap Y) = \Phi
\]

2. The study area

An area of about 50 km² size located in the vicinity of the village Hermeskeil (Figure 1), which is situated 4 km northwest of the Hunnenring, serves as test area for applying the model. The area stands out due to its relatively high number of predominant Roman settlement finds and a multiplicity of graves (Figure 2), which are both dated in Roman and Celtic times.

3. Development of the model

3.1. Modelling Assumptions and Input Parameters

Modelling settlement strategies of Roman and Celtic people concentrates on typical forms of open settlements like small farms or villages [10]. Fortified settlements cannot be considered because we are still lacking knowledge on their precise function. For the same their existence cannot be predicted with certainty. Due to the natural characteristics of the low mountain range position of our study area it can be assumed, that motivation for a choice of site hardly differed between Celtic and Roman times. For this reason analysis of the natural environment is conducted as a diachrone consideration and does not differentiate between epochs.
Data processing and modelling using the concept of Dempster-Shafer was conducted in a GIS. For building the model the data basis was composed of a historic map, a Geological Map and a Digital Elevation Model (DEM) with a resolution of 50 cm, which served for derivation of a significant part of the model parameters. Even though the DEM allows for a higher spatial resolution, a resolution of 10 m was sufficient, since it can be assumed, that processes on a smaller scale had less influence on historical settlement strategies. Official documents on archaeological finds from the region and several publications served as a basis for a database, comprising all known finds. This database contains a total of 22 Roman finds and 53 Celtic finds, which were consulted for parameterisation of the degrees of belief and which served as input parameters for the final evaluation of the model.

3.2. Modelling

In compliance with the theory of Dempster-Shafer the Frame of Discernment of the predictive model has to test the hypothesis \{Presence of site\} and \{Absence of site\} as well as the hypothesis \{Presence of site, Absence of site\}, which expresses uncertainty about presence or absence of a site. Figure 3 shows the Frame of Discernment comprising of these hypothesis and all variables, which were used in the model and which support one of the hypothesis.

![Figure 3. Frame of Discernment with the single hypothesis and their supporting variables](image)

Quantification of the single variables was carried out on basis of a statistical analysis of known sites in the study area but also on estimation by archaeological experts and their knowledge about human behaviour of the former population. In the following section the single variables and their influence on the model are described.

3.2.1. Variables, which support the hypothesis \{Presence of site\}

- **Slope**
  Soils in our study area are characterized by nutrient poorness and stagnant moisture. Thus it can be assumed that slightly sloped areas should have been preferred for settling, whereas flat locations and also distinct slopes should have been avoided by settlements. Statistical analysis of the known sites confirms this estimation, because 80 % of the sites are located in a slope range between 2 – 4°, 15 % in a range of 4 – 6° and only 5 % on flat or steep slopes.

- **Landform**
  Slope as single criterion for characterization of the landscape does not reveal much about the actual topography. In this context a classification of the landscape in characteristic landforms ap-
pears reasonable. For this purpose a raster based algorithm for identifying slope positions and characterization of the landscape was used [11]. The algorithm enables to emerge 10 different classes of landform, from which 5 classes can be found in our study area. Analysis of the find distribution in comparison with these classes shows that nearly all settlement finds are located in the two landform classes ‘wide open valleys’ and ‘elevated plains’.

- **Distance to rivers**

To model this influence for each raster cell the distance (unit: minutes of walking time to the next river) was calculated by use of an Anisotropic Cost-Distance analysis, which incorporates Tobler’s hiking function [12] for modelling a slope dependant hiking velocity. Belief in the hypothesis \{Presence of site\} decreases with increasing distance to the river network and is modelled by use of a sigmoidal monotonically decreasing Fuzzy membership function.

### 3.2.2. Variables, which support the hypothesis \{Absence of site\}

- **Geology**

In conjunction with evaluation of natural resources for agricultural use geologic substrate plays an important role as basic material for pedogenesis. Some substrates indicate poor resources, which are reflected in nutrient poorness or loamy material and which also appear in today’s utilization as predominately forest area. Moreover holocene floodplains should have been avoided as settlement places due to temporary flooding. For these reasons some areas can be identified as unfavourable areas for formation of settlements thus supporting the hypothesis \{Absence of site\}.

- **Distance to road system**

In Celtic as well as in Roman times the formation of settlements and their associated burial grounds are reflected in rule based distances concerning the relative position to the road system. In many cases settlements were established at a minimum distance of 100 m to a main road whereas burial grounds are located in the immediate vicinity of the main roads. A Least Cost Path analysis, which is a common GIS analysis technique, was used for calculating an ideal road network for our area of study. For this purpose connections between known burial grounds were calculated by taking into account that the course of Roman and Celtic roads principally followed the mountain ridges. The result was associated with a historical map and a map of the known Roman road network [13] and allowed in that way to create a reconstruction of the former road network. For calculating the BPA for this variable a buffer of 100 m around the roads was built and assigned a low probability for finding a settlement. Furthermore a sigmoidal monotonically increasing Fuzzy membership function in a range of 100 m – 800 m represent the increasing probability for finding a settlement with cumulative distance to roads. For distances greater than 800 m to the road network also a low probability was assigned.

- **Distance to graves**

The spatial relation between burial grounds and their surrounding settlements shows consistency for both epochs and is reflected in a minimum distance of 150 m between burial grounds and the nearest settlement. In context with modelling based on Dempster-Shafer’s theory the hypothesis \{Absence of site\} is supported for a distance < 150 m between burial ground and related settlement. Quantification in terms of belief is expressed with an assignment of a BPA of 0.8 for all raster cells with a distance < 150 m and a BPA of 0.2 for those raster cells which are in farther regions.
4. Result and outlook

Dempster’s Rule of aggregation was used to combine the single BPA’s to a final result for representing the total belief for the different hypothesis. Figure 4 shows the total belief for the hypothesis \{Presence of site\} in conjunction with all known settlement finds in the study area.

The result was grouped into three equidistant classes, which in their entirety represent the suitability of a certain location for settlement by attaching a gradient value between 0 and 1 to it. It can be pointed out, that 18.75 % of the finds are located in areas of low or middle suitability for settlement and 62.5 %, respectively, can be found in areas of high potential for settlement. Inconsistencies, which are reflected in assignment of real finds to areas of calculated low potential, suggest inaccurate source information concerning the find position. Hence intensive field work should consider the results of modelling and, vice versa, should contribute to the refinement of the modelling process by improving the model’s parameterization.

Further support for targeted prospection campaigns and their success rate should be achieved by consideration of the results in context with erosion processes. From an archaeological point of view erosion causes unearthing of finds and effects in that way destruction of archaeological remains by weather influences, whereas deposition of soil material contributes to conservation of archaeological remains by covering the remains. Mitas and Mitasova [14] developed an algorithm, which allows for calculating the spatial distribution of erosion and deposition rates of a specific area and which was applied for the study area. The result of the model enables excluding specific areas with high erosion- or deposition rates and allows in that way a more differentiate evaluation of the result of the predictive model with regard to targeted prospections.
References


