Using multivariate statistics and fuzzy logic system to analyse settlement preferences in lowland areas of the temperate zone: an example from the Polish Lowlands

Jasiewicz Jarosław a,b,*, Iwona Hildebrandt-Radke a

a Adam Mickiewicz University, Department of Paleogeography and Quaternary Geology, Geocology and Geoinformation Institute, Dzięgielowa 27, 60-680 Poznan, Poland

Article info
Article history:
Received 26 November 2008
Received in revised form
13 March 2009
Accepted 10 June 2009

Keywords:
Predictive modelling
Linguistic variables
Geomorphometry
Great Polish Lowland
Settlement preferences
Bootstrap simulation

Abstract
The influence of relief, as well as other environmental features such as soils or the hydrological regime, on prehistoric settlement patterns is commonly accepted. Settlement patterns and preferences changed over time in different areas, so it is difficult to formulate general rules of human behaviour throughout history. In lowland areas of the temperate zone, which lack radical relief changes and have widespread access to water resources, specific settlement preferences and significant changes in these preferences are difficult to interpret.

New methods developed to address the problem of the settlement transformations as a result of prehistoric settlement and economic processes using multivariate statistics have been adapted for the middle part of the Great Poland Lowland. In this region, data from the Polish Archaeological Record Project database have been used.

Models of the dependence between geomorphological features and settlement intensity form the basis for estimating changes in prehistoric settlements. The degree of preference (or avoidance) is a nonparametric function of the number of archaeological sites observed in the area relative to the number of sites expected from a completely random pattern. The most crucial factors for settlement are distance to plateau edges, distance to water bodies, and the wetness index. The results show that the entire investigated area is more or less suitable for settlement, but the most occupied terrains are the areas near plateau edges.

1. Introduction
There is no doubt that the distribution of settlement sites in all morphoclimatic zones is not random. Numerous studies of settlement processes, conducted both before and after the introduction of GIS techniques to archaeology, have indicated that apart socio-economical conditions there are relationships between features such as topographic relief, the distance from water bodies, the soil cover type, and the locations of archaeological sites (Willey, 1953; Williams, 1956; Williams et al., 1973; Warren, 1990; Kvamme 1992; Stančič and Kvarme, 1999; Vanacker et al., 2001; Duke, 2003; Bauer et al., 2004; Fletcher, 2008). However, in spite of the intensive settlement of lowland areas in the temperate zone since the Neolithic period, it is difficult to unambiguously distinguish areas that were preferred or avoided in the prehistoric settlement process in these territories. The relatively small diversification of the individual environmental features over distances of several dozen kilometres, as well as the large number of archaeological sites found in almost all types of environments, are the main problems in such studies of this region.

For that reason, developing a method for studying settlement in areas of low natural diversity is necessary. Researchers attempting to analyse settlement preferences in temperate zone lowlands must consider three important questions: 1) To what extent is landscape variety of archaeological sites an effect of real preferences, and to what extent does it reflect the accessibility of the specific areas or different social or economical conditions? 2) Which environmental variables, and to what extent, have a real affect on settlement patterns? 3) How can boundaries between preferred and avoided areas be defined when landscape diversity is small and landscapes lack clearly outlined settlement zones?

The first problem can be solved by comparing the distributions of individual environmental features over the entire studied area with the distributions of archaeological sites, and then examining the significance of the differences using statistical procedures...
variables absolves the researcher of making arbitrary decisions about a given area classification, replacing unambiguous notions with terms that define membership to a given set in an intuitive but not unambiguous way.

This study presents a method of examining settlement preferences with regard to selected hydro-geomorphological features using multivariate statistics and fuzzy logic systems. The study was conducted using a data set from the Polish National Record of Archaeological Sites program (PNRAS) (Jaskanis, 1992; Prinke, 1992, 1999, 2002), described also as the Polish Archaeological Record Project (PARP) (Barford et al., 2000), which is a variant of the Site and Monument Record (SMR) program (Wheatley, Garcia Sanjuán, 2002). The database covers all of Poland (312,685 km²), and currently contains more than 450,000 inventoried archaeological sites (Prinke, 2002), which makes this program one of the most advanced in the world.

Application of data from the PNRAS, rather than from detailed excavation studies, to examine settlement preferences has both advantages and disadvantages (Garcia Sanjuán, 2005). An unquestionable advantage of the surface research is the relatively even and representative reconnaissance of the study area (the representativeness of an excavation survey is greatly reduced in surveys made in areas of planned construction and development) as well as the large number of recognized sites. However, the simplified chronology (most often limited to a culture, sometimes to a period only), as well as the simplified method of classification of the type of archaeological sites based just on the number of finds, is a major limitation of the PNRAS data application. The PNRAS program is not carried out in forested areas, and to a limited degree in urbanized areas. The precision of the archaeological site location is also a difficult problem to solve (Staničić and Kvamme, 1999, Vanacker et al., 2001). As a rule, a site is marked on maps in the form of a point (traces and settlement points) or a spot (settlement). Unfortunately, the quantity of the analysed material, as well as the methodology of the surface fieldwork, often does not allow for field verification of the analysed material. In opinion of authors, the use of data from surface surveys is possible on the assumptions that: 1) there is a random distribution of the location errors; and 2) the accuracy of the data locations is about 200–220 m (1 cm on a map at a scale of 1:25000). This level of accuracy limits the analysis of the preferences to a scale of geomorphological forms of sizes less than 200 m.

Spatial models based on the analysis of settlement material may be divided into descriptive and predictive models (Kvamme, 2005). Predictive models are aimed at predicting the decision making for the locations of sites in areas not studied before (Fry et al., 2004). In contrast, descriptive models are used for presenting general tendencies and patterns in data; they are aimed at understanding the relationships between the natural environment and the settlement process. The research material and methodology that have been applied in this study classify the presented model in the descriptive group; however, the lack of PNRAS fieldwork in forested areas allows us to predict the extent of settlements in these areas where there are no archaeological data.

2. Study area

When choosing the study area, the following criteria were taken into consideration in order to make the results as credible as possible: the percentage of forested and urbanized areas should be the lowest possible (below 10% in the examined area compared with around 30% in Poland) in order to minimize the area excluded from the PNRAS research; arable farming ought to be the prevailing type of land use; surface surveys carried on the freshly ploughed areas provide the most surface material; an area should be well recognized by means of the surface survey (an average density of
the poly- and monocultural sites about 3.5 per km$^2$, compared with 1.43 per km$^2$ for all of Poland (Prinke, 2002); the relief of the area ought to be diverse so as to cover the wide spectrum of morphological forms typical of the temperate lowland landscape.

The selected area, located in the southern part of the Great Poland Lowland (Nizina Wielkopolska) (Fig. 2) is typical of the Middle-European Lowland landscape. The area consists of the boggy Warta-Odra Pradolina in the north, and the flat lakeless Kościan Plain (Równina Kościańska) formed from the morainic plateau and cut with glacial channels to the south, where it turns into the Śląsko and Krzywiń Lakelands (Pojezierze Sławskie i Krzywińskie), as well as into the most morphometrically diverse morainic Leszno Plateau (Wysoczyzna Leszczyńska). The morphometrical characteristics of the study area are presented in Table 1.

3. Material and methods

A multivariate statistical analysis incorporating standard GIS procedures based on the continuous reclassification of relief in derivative maps was used. This resulted in partial maps of settlement intensity to which a weighted geometrical overlay procedure was applied in order to draw maps of settlement intensity. The summary flowchart (Fig. 3) shows the successive steps of analysis.

3.1. Selecting and computing the appropriate hydro-geomorphological variables

A Digital Terrain Model (DTM) was obtained by manual digitization of the contour lines from a map of a scale of 1:50,000. These data were used to develop the DTM using the r.surf.rst algorithm (Mitasa and Mitas, 1993) with the default settings at the resolution of 30 m/cell. The vertical error of the model is less than 1.4 m.

The value of an analysis depends mainly on the choice of the geomorphological variables used to perform the analysis. The basic criterion chosen was their potential influence on the location of the settlement sites in the past. Eight geomorphological variables were selected for analysis. All are uncorrelated except pair LS=SLOPE, which are moderately correlated. The absence of purely socio-economic variables has a methodological basis, since the uncertainty of variables based on superficial materials would be significantly more than the uncertainty of natural variables. Table 2 shows and describes the variables with its socio-economical interpretation. The independence of the separate variables is a condition of the analysis correctness (Kvamme, 1990). If the variables were correlated, their influence on the final result would be overestimated. Values of the correlation coefficients between the individual variables are presented in Table 3.

Although all hydro-geomorphological variables are broadly discussed in numerous papers (see: Wilson and Gallant, 2000 for details), we will present summarised characteristics of relevant variables. The first group of variables consists of factors connected with the relief: RELEV – relative elevation, and EDGE – distance from the plateau edge. Obtaining these variables required designating the surfaces for which the relative values of elevation have been defined. For RELEV, the nonparametric trend surface (Cleveland, 1979) of the initial points of the drainage system was fitted.

![Fig. 2. The topographic relief of the research area with the locations of the archaeological sites.](image-url)
This surface allows the determination of the local elevations and depressions in terms of the local hydrological trend. To define EDGE, the trend of the nonparametric surface formed by the second and third stream order watercourse contact points, according to the Strahler’s (1952) classification, was selected. It was observed that second order streams determine the run-off routes from the plateau areas, whereas the third order streams develop in the local valleys and intramorainic channels. This is why the surface fitted to these points will be similar to the surface separating the morainic areas from the valley tracts. This procedure was based on the ISOBASE concept derived for tectonic landscapes (Filosofov, 1960) and adapted to lowland areas. The detailed description of the construction of both surfaces, as well as their geomorphometrical significance, are both beyond the scope of this work and shall be the theme of a separate publication.

The second group includes the hydro-geomorphological variables: WATER, DRAIN, and WET. The WATER variable was determined on the basis of a pattern of main natural contemporary rivers and lakes, while the DRAIN variable was defined on the basis of the whole drainage network (including dry valleys), which was used before to determine the RELEV parameter. A simple buffering operation was applied to establish the distance zones. All these variables require prior calculation of the hydrological topology as defined by the flow paths. Flow directions were computed using a conventional D8 algorithm using the freeware GIS software: Terrain Analysis System (Lindsay, 2005). The terrain model without depressions is an essential part of the drainage network setting procedure. Depressions, which at the resolutions considered here are artefacts, had been removed from the digital model before calculating the flow directions. Afterwards, drainage in flat areas was enforced. The drainage network was determined by the popular algorithm of O’Callaghan and Mark (1984), which sets the drainage network on the basis of the threshold value of the specific catchment area. Only the WET (wetness index; Moore et al., 1991) was determined on the basis of the multiple flow direction algorithm DInf (Tarboton, 1997).

The third group of variables is connected with the slope inclination and exposure. LS, RAD, and SLOPE were computed on the basis of the terrain model parameters using the standard procedures available in the GRASS GIS and TAS software. The RAD variable was calculated for the shortest day in a year because this day demonstrated the biggest spatial variation of RAD.

To distribute the uncertainty error (Foody, 2003) of the site locations and also to take into consideration the sites surfaces, which usually exceed the area of an individual cell, each layer of the hydro-geomorphological variables was blurred using a local average filter with a 5 x 5 cell moving window. On the basis of the hydro-geomorphological variables, a second set was created in which the forested areas and large water basins were excluded from the analysis.

3.2. Creating the archaeological data set

Data regarding the locations of the archaeological sites were obtained from the manual digitization of 25 PNRAS map sheets. Each sheet covers an area of about 5 x 7 km and is a part of the topographic map of Poland in the PUWG 1965 reference system (datum: Pulkowo 1942, EPSG code: 2174). Each sheet was scanned and georeferenced on the basis of the original 1:25000 scale map. Next, the vector layer was created, and all sites marked on the map were added to this layer. The accuracy of the site location is about 150 m due to georeferencing errors and symbol sizes on the original map. This seemingly small precision is nevertheless comparable to the average area of settlements in the past (Stanić and Kvamme, 1999). Since different types of sites are subject to different natural restrictions (Hatzinikolaou et al., 2003), each archaeological site was classified as either permanently or temporarily settled on the basis of archival descriptions. Old ramparts, settlements, burial grounds, and farm objects were included among the permanent
Table 2
Characteristics of hydro-geomorphological variables used for the analysis.

<table>
<thead>
<tr>
<th>Abbreviation of geomorphological variable</th>
<th>Geomorphological variable</th>
<th>Description of variable</th>
<th>Reference for method of calculation</th>
<th>Socio-economic interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RELEV</td>
<td>Relative elevation</td>
<td>An elevation related to the hypothetical surface of initial points of the drainage network</td>
<td>Filosofov (1960) (modified)</td>
<td>Observation points, defence</td>
</tr>
<tr>
<td>EDGES</td>
<td>Distance to plateau edges</td>
<td>Distance to the (lower) plateau edge. The edge was defined as the zone where the hypothetical trend surface of the contact of 2nd and 3rd order streams intersects the recent terrain surface.</td>
<td>Strahler’s (1952)</td>
<td>Distance to valleys and water</td>
</tr>
<tr>
<td>WATER</td>
<td>Distance to main streams and lakes</td>
<td>Distance from the main rivers and water bodies</td>
<td></td>
<td>Distance to water bodies</td>
</tr>
<tr>
<td>DRAIN</td>
<td>Distance to recent drainage network</td>
<td>Distance from all streams and drainage channels</td>
<td>Hofierka (1997)</td>
<td>Solar energy, farming</td>
</tr>
<tr>
<td>RAD</td>
<td>Diffuse irradiance</td>
<td>The volume of received diffuse irradiance [W.m⁻²] (winter)</td>
<td>Wood (1996)</td>
<td>Buildings, farming</td>
</tr>
<tr>
<td>SLOPE</td>
<td>Slope gradient</td>
<td>Slope inclination</td>
<td>Moore et al. (1991)</td>
<td>Erosion intensity, ground stability</td>
</tr>
<tr>
<td>WET</td>
<td>Wetness index</td>
<td>Topographic index (or Wetness index)CTI – ln(1/SCA/tan θ)</td>
<td>Moore and Burch (1986)</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>Transport capacity index</td>
<td>Transport capacity index (length of slope inclination)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Pearson correlation coefficient of hydro-geomorphological variables.

<table>
<thead>
<tr>
<th>RELEV</th>
<th>EDGES</th>
<th>WATER</th>
<th>DRAIN</th>
<th>RAD</th>
<th>SLOPE</th>
<th>WET</th>
<th>LS</th>
<th>RELEV</th>
<th>EDGES</th>
<th>WATER</th>
<th>DRAIN</th>
<th>RAD</th>
<th>SLOPE</th>
<th>WET</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.39</td>
<td>0.47</td>
<td>0.52</td>
<td>0.01</td>
<td>0.06</td>
<td>0.27</td>
<td>0</td>
<td>0.39</td>
<td>0.47</td>
<td>0.52</td>
<td>0.01</td>
<td>0.06</td>
<td>0.27</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.39</td>
<td>1</td>
<td>0.11</td>
<td>0.27</td>
<td>0.03</td>
<td>0.09</td>
<td>0.1</td>
<td>0</td>
<td>0.39</td>
<td>0.11</td>
<td>0.27</td>
<td>0.03</td>
<td>0.09</td>
<td>0.1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.47</td>
<td>0.11</td>
<td>1</td>
<td>0.46</td>
<td>0.01</td>
<td>0.01</td>
<td>0.28</td>
<td>0.08</td>
<td>0.47</td>
<td>0.11</td>
<td>1</td>
<td>0.46</td>
<td>0.01</td>
<td>0.01</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>0.52</td>
<td>0.27</td>
<td>0.46</td>
<td>1</td>
<td>0.02</td>
<td>0.04</td>
<td>0.14</td>
<td>0.48</td>
<td>0.52</td>
<td>0.27</td>
<td>0.46</td>
<td>1</td>
<td>0.02</td>
<td>0.04</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>–0.01</td>
<td>0.03</td>
<td>–0.01</td>
<td>–0.02</td>
<td>1</td>
<td>0.15</td>
<td>0.04</td>
<td>0.08</td>
<td>–0.01</td>
<td>0.03</td>
<td>–0.01</td>
<td>–0.02</td>
<td>1</td>
<td>0.15</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td>–0.09</td>
<td>0.01</td>
<td>–0.04</td>
<td>0.15</td>
<td>0.48</td>
<td>0.64</td>
<td>0.64</td>
<td>0.06</td>
<td>–0.09</td>
<td>0.01</td>
<td>–0.04</td>
<td>0.15</td>
<td>0.48</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>–0.27</td>
<td>–0.1</td>
<td>–0.28</td>
<td>0.14</td>
<td>0.04</td>
<td>0.48</td>
<td>1</td>
<td>0.33</td>
<td>–0.27</td>
<td>–0.1</td>
<td>–0.28</td>
<td>0.14</td>
<td>0.04</td>
<td>0.48</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>–0.08</td>
<td>–0.05</td>
<td>0.08</td>
<td>0.64</td>
<td>–0.33</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>–0.08</td>
<td>–0.05</td>
<td>0.08</td>
<td>0.64</td>
<td>–0.33</td>
<td></td>
</tr>
</tbody>
</table>

4. Selecting and applying an appropriate analysis methodology

4.1. Univariate statistical analysis

Determining the influence of the particular hydro-geomorphological variables on the development of prehistoric settlement patterns requires the comparison of the given variable distribution over the whole area (Iₑ) with the given variable density distribution in the appropriate sets of settlement points (Iₚ). A visual comparison of the density plot (Fig. 4) allows one to detect the differences between both distributions for all periods and the environmental variables, and points out what values of the variable were preferred and what values were avoided by the prehistoric population. Examining the significance of these differences requires an explanation of whether the pattern could have originated as a result of a random point process. The most commonly used statistical method for univariate analysis is the chi-square test. The method...
has an important limitation: a significant disproportion between the raster cells in the morphometrical maps (833,876) and the number of archaeological sites for the individual types and archaeological periods (50–1500, see Table 4). Therefore, we tried to develop the alternative method that does not require assumptions about the distribution and the proper number of archaeological sites. This method is based on the bootstrap simulation (Diaconis and Efron, 1983), and can be directly applied to the percentage of sites located within a defined interval of map values.

To perform the simulation, the value range of the individual hydro-geomorphological variables was divided into 14 classes. The number of classes was determined using Strugess’s rule (1926; after Armstrong et al., 2003) assuming the number of all archaeological data as a base. Because the percentage of extreme classes for each variable was very small, the partition based on Fisher’s (1958) method was used. This method creates classes of various widths (Fig. 5) so as to make the variance inside the individual classes the smallest possible. The classification was done using the classInt package of the R software (Bivand, 2008).

For the separate hydro-geomorphological variables and for all types of archaeological sites and periods in each class, the number of existing archaeological sites was compared to the number of sites expected if the distribution was random and reproduced the given hydro-geomorphological variable distribution. Altogether, 168 comparisons were made, representing 7 archaeological periods, 3 types of sites, and 8 hydro-geomorphological variables. To determine the interval of values for which the distribution should be recognized as random, for each of the 168 comparisons the simulations/random samplings were carried out from the population of the raster cell of r_raster set (Table 5) using the number of cells number of a given set that the archaeological sites had. For each of the combinations, 100 runs were carried out. Vanacker et al. (2001) showed that the quantile values in this kind of simulation stabilize at around 80 runs. Similarly, Baddeley and Turner (2005) applied 100 Monte Carlo simulations to determine the envelope of complete spatial randomness of point patterns at a 99% level of confidence. In order to determine the distribution envelope for each class, the maximum and minimum values were chosen. In order to conduct the differences significance test, the null hypothesis was assumed that the values of the hydro-geomorphological variable distribution within the archaeological sites maybe explained by the random arrangement of points. The test rejects the null hypothesis if the graph of the observed sites (Ip) lies outside the envelope at any class. For all examined comparisons, with the exception of the Neolithic period, the null hypothesis was rejected.

### Table 5

Arrangement of data used in analysis.

<table>
<thead>
<tr>
<th>Sets of raster data</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>all_sites</td>
<td>all_sites include 3 data sets: all_sites, perm_sites, and nperm_sites. Values of hydro-geomorphological variables come from sampling of r_raster set in locations of archaeological sites.</td>
</tr>
<tr>
<td>all_perm</td>
<td>The sets consist of subsets appropriate for given archaeological periods. Each subset is a set of eight layers of hydro-geomorphological variables.</td>
</tr>
<tr>
<td>all_nperm</td>
<td>The size of each subset is presented in Table 4</td>
</tr>
</tbody>
</table>

#### 4.2. Potential settlement intensity maps and its statistical weights

Having rejected the hypothesis about the random points distribution regarding the individual hydro-geomorphological variables, an index of the relative settlement intensity was determined for each comparison of the area population to the points population. This index is a quotient of the stated number of settlement points in a given class of the variable value to the number of the points if their distribution was independent of a given variable:

\[ I = \frac{I_p}{I_e} \]

[1]

Without doubt an index value of 1 indicates the lack of settlement preferences. A value less than 1 means that in the area of a given variable, fewer points were found in comparison with a random distribution, while a value greater than 1 indicates the opposite situation: the preference of the feature given values. Additionally, the nonlinear character of the relationship between selected hydro-geomorphological (Fig. 6) variables prevented us from conducting the standard procedure of reclassifying as a fuzzy set of the values from 0 to 1 (Zadeh, 1965; Klir & Yuan, 1995; Demicco and Klir, 2004). Thus the centre of linguistic variable “indifferent areas” can be specified, but boundaries of “preferred” and “avoided” variables cannot be defined. Therefore has been assumed that all ranges of indicators belong to the linguistic “indifferent area” variable (Fig. 7) which couples with “avoided” (left) and “preferred” (right), but has uncertain boundary
definitions (Foody, 2003). Because of this, the original values of the settlement intensity index were used in the analysis. The example maps of the index values for the whole sites population are presented in Fig. 9, and confirm the initially assumed thesis that the whole area to a smaller or greater extent is predisposed to settlement, while the differences are related only to the intensity. Since they were attributed to the classes of the precisely defined intervals, and also because some classes diverged strongly from the surroundings (e.g., Fig. 6, the RELEV variable), the precise classification was exchanged with the smoothed model. It was difficult to fit any parametric model to the data (Wood, 1971); this was the reason why the nonparametric model, the Local Polynomial Regression Fitting using least-square method (LOESS), was decided on (Cleveland, 1979). This is a model of regression methods designed to address situations in which the classical procedures do not perform well or cannot be effectively applied. The neighbourhood parameter $\theta = 0.4$ was assumed for the analysed combinations.

Because the models (Fig. 6) were used to define the relationship between the hydro-geomorphological variable value and the settlement index, the outlining of the settlement intensity maps required only applying the model to the variables map (using the predict() function). As a result, 186 maps of the relative settlement intensities (Fig. 8) were obtained. These maps were then subjected to the overlay procedure. For each archaeological type and period, the individual partial maps were multiplied by themselves, achieving 21 resultant maps of the potential settlement intensity (three for each period). Additionally, to weaken the significance of the less essential variables and to strengthen the more essential ones, the weights in the interval from 0 to 1 were applied, computed on the basis of the number of settlement points located the areas where the index value was greater than the average index value plus the standard deviation (Table 6, column 6).

Since the index values indicate the settlement intensity regarding the random distribution of points and determine how many times more often the sites for given variables occur in comparison with the random distribution, and thereby have a log-normal distribution, the overlaying operation was realized using the weighted geometrical average according to the formula presented in the supplementary data.

The resultant settlement maps of the sites are presented in Fig. 9.

### 5. Model evaluation

Defining the model accuracy requires taking two sources of uncertainty into consideration: uncertainty connected with the quality of the data and the accuracy of georeferencing, and uncertainty connected with model assumptions and factors other than those considered in the model (Foody, 2003).

In the case of the first uncertainty, estimation of the error value is not possible because it is dependent on the accuracy of the archaeological site location, the precision of its dating, and on the correct classification of its function, which is not always possible.

In the case of the second uncertainty, the location of some points may be connected with other factors not taken into consideration in the model, such as soils and lithology, location with respect to the socio-economic restriction like ancient transport routes, or the appearance of new settlements near older ones as a result of the population migration. These features cannot be examined using the methods proposed in this paper require a different approach.
Validation of the model was carried out by comparing the density plot of the resultant settlement intensity index for the entire area with the distribution of this factor in the individual subpopulations of the archaeological sites. Comparisons carried out for all of prehistoric time as well as the Middle Ages, presented in Fig. 10, demonstrate the essential differences between the density plots for the whole area as well as the individual sites. Moreover, the distinct differences between the distributions for the subpopulation of the permanent and non-permanent sites show that the settlement requirements for the permanent sites are stricter than for the temporary sites. Between 50 and 60% of the sites, depending on the archaeological period, occur in the indifferent and preferred areas, whereas up to 40% occur in the avoided areas; this indicates that the model uncertainty explains from 50 to 60% of the settlement preferences. The remaining 40% is caused by other factors not taken into consideration in the analysis, or is random variability.

6. Discussion

The fundamental value of the presented method is its ability to determine settlement preferences on the basis of a combination of several factors simultaneously by individually determining the weight of each factor. It is difficult to determine which factors have
a real impact on settlement process and which have little or no
significance. With the use of partial maps of settlement intensity, as
well as weights attributed to them (Ergin et al., 2003), we can define
at each point the hydro-geomorphological factors that had a real
impact on the preference index value without evaluating the role of
particular factors as proposed by Ergin et al. (2003). Furthermore,
the analysis of weights, as described above, can identify the actual
role of particular hydro-geomorphological variables.

In the studied area, both temporal and spatial variability of the
settlement preferences is observed. The most preferred areas are
the slope and near-edge zones of the river valleys and small valleys
(Williams et al., 1973). This is confirmed by the weight analysis for
the individual periods and types of sites (Tab. 6). WATER, WET, and
SLOPE are also important factors. Detailed analysis of the models
(Fig. 6) shows slightly more complicated relationships. The EDGE
variable indicates that the near-edge areas are very popular, both
towards the valley and the plateau. The diversification begins
outside the edge zones. Valley tracts, apart from the edge zones,
are characterized as having the least interest, whereas the decline
in the interest is smaller in the plateau zone. The significantly less
interest in the valley zones might be explained by the increased
flood probability, as well as by the wetness of the valley tracts
(Rotnicki, 1991), which is demonstrated by the WET variable. This
conclusion is correlated with the model for the WATER variable,
where the decrease in preferences in the zones in the vicinity of
water is also noticeable. The greatest interest is manifested in the
areas 150–250 m away from large water bodies. The relationship
between the distance from water and the intensity of the settle-
ment in the section more than 250 m far from water is linear and
decreases with the distance from water. Interesting relations are
demonstrated by the SLOPE variable. It turns out that the least
interest is manifested in the flat areas with slopes less than 1
degree, whereas the greatest interest is in areas with slopes of
1–3 degrees. This is caused most probably by the fact that the flat
areas are the bottoms of glacial valleys, which in this region do not
receive much interest. The WET variable has a similar meaning. The
relationship is linear, and the greatest interest is manifested in the
areas of the least potential wetness. The weights of the remaining
factors are decidedly smaller. Additionally, the RELEV variable is of
almost no significance for the settlement pattern (weight = 0.07).
This probably results from the lack of distinct elevations in the
examined area, including natural defensive and observation points,
the presence of which could have changed the pattern. The pref-
ferences peaks noticeable on the model curve for the RELEV vari-
able are connected with the settlement concentration in the edge
zones.

Fig. 9. Example of maps of the potential settlement intensity (the set all sites).
The proposed method also allows the detection of the temporal changes of the settlement processes. Unfortunately, the low precision of the cultural ages of the sites in the PARP database makes it impossible to date the settlements more precisely than within a culture or period. Only a dozen or so percent of the material is defined to a higher level of precision. Because of this, the authors decided to present more generalized conclusions without the “artificial accuracy” of the settlement transformation pattern.

The tendency of the settlement changes is not significantly different from the pattern that is known from other settlement studies, as well as from palaeogeographical studies (Rotnicki, 1991). The results presented in Figs. 8 and 9 indicate two tendencies of the spatial distribution of settlements. In the Neolithic, relatively less than average interest in the river valley bottoms is registered; however, there is an increase of interest in the near-edge parts of the plateau. The Lusatian Culture (Late Bronze/Early Iron Age) is a period of settlement distribution with an increased share in the valley tracts. The Przeworsk culture (Iron Age – the La Tène period – the Roman Iron Age) is characterized by the decline in interest both in the river valley areas and in the terrains of the plateau, as well as in other settlement concentrations in the near-edge areas. A similar process of settlement movement from the low areas/terrains to the edges of the plateau can be observed at the turn of the Early and Late Middle Ages, where the tendencies of the settlement changes are even more distinct (Fig. 11). Those settlement transformations are connected with the hydroclimatic changes that occurred in the Late Holocene. The wet climate, with high flood probability, of the Late Atlantic period did not favour settlements in the valley bottoms. The drier Subboreal period (of the lower hydrological regime) favoured directing the settlements towards the river valleys. The wetter climate at the turn of the Subboreal and Sub-Atlantic, and the process of rising water levels connected with it, caused settlements to migrate towards the plateau. The same phenomenon is clearly visible at the turn of the Early and Late Middle Ages.

Comparison of the settlement intensity maps between the subpopulations of the permanent and non-permanent sites indicates that the criteria of the settlement preferences for the permanent sites are more unambiguous than for the non-permanent sites, whose locations are much more random.

Fig. 10. Validation of the models: the comparison of the value distribution of the potential settlement intensity index to the value distribution of the index where the archaeological sites occur: (A) all sites of all periods; (B) permanent sites of all periods; (C) non-permanent sites of all periods.

Fig. 11. The settlement pattern transformations between the Early and Late Middle Ages (permanent sites) in comparison with the relief. Warm colours indicate an increase of the value index, while cold colours indicate a decrease.
7. Conclusions and further development

Here we have presented the multivariate statistical method for determining the impact of various natural factors on prehistoric settlement preferences. Our method is data- and area–independent, with all procedures written as a collection of R-language scripts which can be applied to any set of archaeological data anywhere, and can also be applied to any natural or socio-economic variables. The most important advantages of the method are:

The method was developed for and tested on lowland areas of the temperate zone, where low contrasts of most natural factors require a more sensitive tool than Boolean operations. The challenge of applying Boolean methods to studies of prehistoric settlements in the lowlands of the temperate zone results from taking into account estimates of settlement preferences.

The novelty of the method is that the notion of settlement preferences have been defined as the ratio of the number of sites in a given class in an area to the percentage of sites of a given class in the entire studied area. To verify the hypothesis about the random distribution of sites regarding the individual geomorphological features, a bootstrap simulation was applied. This simulation, in contrast with traditional nonparametric tests, is more sensitive and provides more information about the scale of the differences among the distributions.

Assumption that the entire area belongs to the fuzzy set on both sides and defined by the linguistic variable “indifferent area”, eliminated the inconveniences connected with the very subtle differences between the individual areas defined as “preferred” and “avoided”. This is in contrast with precisely defined sets, where decisions like these, even based on the statistical criteria, would leave very large uncertainties on the sets boundaries (Ergin et al., 2003; Foody, 2003).

Constructing a model based on the weighted overlay of geomorphological features allowed us to determine the character of the settlement preferences on the basis of the combination of several features at the same time. Determining the role of each hydro-geomorphological components was possible in any place and in any period of the studied area because of the use of the partial maps and its weights combined with the partial models. The use of weights allows us to overcome problems with redundant variables. The method has shown that not all variables have the same impact on settlement pattern. Distance to water bodies and plateau edges, wetness index, and slope inclination are more than other factors. Furthermore, the role of various factors changes over time. This problem will be presented in a comprehensive regional study.

Evaluation of the model shows that no more than 50–60 percent of settlement variability can be explained by hydro-geomorphological factors. Instead, environmental factors constrain decisions by eliminating areas unsuitable to settlement based on hydro-climatic conditions, but they do not fully determine human spatial behaviour (Barceló et al., 2002).

Acknowledgments

There are many people who helped prepare this paper. In particular we would like to thank Wojewódzki Konserwator Zabytków in Leszno, Jacek Nowakowski for provision of PNRAS data; Robert Graś for digitising the archaeological data; Prof. Karol Rotnicki for his care on Robert Graś’ Master’s thesis. We would also like to thank anonymous reviewers for their valuable comments on the first version of this manuscript.

The work was partially supported by of Polish Ministry of Scientific Research and Information Technology, Grant of State Committee for Scientific Research, No 2P04E01630.

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.jas.2009.06.004.

References


