Applications of geographical information systems in archaeological research

M. Wansleeben

31.1 Introduction

The application of computers in research and analysis has spread beyond the exact sciences a long time ago. In almost all scientific fields automation is occurring. The computer, especially the personal computer, is here to stay. Nowadays in most archaeological institutes computers are present, too. Apart from word processing, a routine application, the computer is increasingly being used for storing and analysing databases. Good database management programmes permit the handling of quite extensive databases using a simple PC. What used to take a mainframe can now be done with a desktop computer.

Throughout the sixties in geography almost all information was recorded on maps. Geographical maps contain a lot of information, but advances in computer technology now allow the electronic storage of maps, and the use of electronic map-storage allows any particular geographical aspect can be highlighted. The making of thematic maps for special purposes has been simplified enormously in this way; several thematic maps may be combined into a new thematic map. The computer technology that allows this sort of application is collectively called Geographical Information Systems.

In archaeology relatively little use is made of this technology. In this paper some possibilities of a Geographical Information System will be described and illustrated. Various analysing techniques will be used on a single archaeological database; results will be compared and a preliminary interpretation will be presented.

31.2 Geographical Information Systems

A Geographical Information System (GIS) is: 'a computerized set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world for a particular set of purposes' (Burrough 1986, p. 6) A GIS is a piece of software consisting of two major components. One component stores the spatial information (database management), whereas the other serves to manipulate and display these data. The data, transformed or not, are usually displayed cartographically. Both the...
differences in the storable shapes of the topological units according to 'vector-based' and 'raster-based' Geographical Information Systems database management capacity and the possibilities for transformation and display are crucial to the performance of a GIS. Both major components merit a closer look.

Let us start with the first major component: storage of the spatial information. Any GIS database consists of so-called topological units: a point, a line or an area. For each topological unit as a rule two particulars are recorded:

1. the geographical position;
2. an intrinsic attribute.

The geographical position is usually recorded as (a series of) X- and Y-coordinates. Any kind of property can be designated an intrinsic attribute and recorded as such. In geography, attributes may e.g. be soil texture, ground water level, pH, height above sea level and vegetation type. The description of the topological units of one intrinsic attribute can be considered a thematic map. In this context this is called an overlay. So an overlay consists of a series of localised topological units, each with a label specifying the value of the intrinsic attribute (Fig. 31.1). Several overlays together constitute a geographical database (Fig. 31.2).

Some geographical databases can handle lines and areas of varying sizes and shapes. These are recorded by means of vectors, hence the name 'vector-based' GIS. Storage of this type of graphic information is quite complex, inhibiting the use of this technique without professional software. Other databases can only store data about points. If these points are located in a fixed grid and each point represents a square area (a cell), this is called a 'grid-based' GIS. This type of GIS is relatively simple, allowing even the non-professional software engineers to make this kind of computer program.
Figure 31.2: In the archaeological research in the south of the Netherlands 6 geographical and 8 archaeological overlays were used (left). The way a geographical map is deformed in order to enable storage in a ‘raster-based’ Geographical Information System is shown (right). (EM = early Mesolithic, LM = late Mesolithic, LBK = Bandkeramic culture, Rō = Rōssen culture, MK = Michelsberg culture, WSV = Wartberg-Stein-Vlaardingen culture, PFB = Protruding Foot Beaker, BB = Bell Beaker culture) (after Harris 1986)
In both types of GIS the topological unit should be consistent. Each attribute has only a single value that is common to whole unit, e.g. in the entire topological unit the soil texture is sand or the height above sea level is 30 m. + N.A.P. (= Dutch Ordinance Datum). A vector-GIS causes no special problems in this respect, since each unit is defined in such a way as to meet the above mentioned requirement of a consistent topological unit. However, this requirement does have its repercussions in the case of a grid-GIS. Here the terrain must be deformed into homogeneous square units. The degree of deformation depends on the size of the grid cell. Small cells offer a more exact picture of the real world than larger cells. However, the number of small cells that constitute an area is considerably higher than the number of large cells for the same area. So the greater accuracy is outweighed by a strong increase in the number of observations, which could cause problems in storing the data. Although a grid-GIS is easy to use, a major drawback is the less exact image of reality in the geographic database. By choosing a particular size for the cells, depending on the archaeological problem and the size of the area under investigation, the disadvantages can however be minimised.

The second major component of a GIS contains the possibilities to change/transform and retrieve/display data. Three kinds of transformations exist: 'point', 'legend-unit' and 'neighbourhood' transformations (Burrough 1986). In point transformations the value of a cell on a newly created overlay is determined by the values of cells with the same geographical position on other overlays. This operation is repeated for each single cell. An example of this type is ADD: add the value of cell i on overlay A to the value of cell i on overlay B and integrate into cell i on overlay C (Fig. 31.3). In legend-unit transformations all cells on an overlay with equal value (belonging to the same legend-unit) are reworked. Using the RENUMBER-routine all cells with e.g. value 3 will be renumbered to value 5. In neighbourhood transformations a cell will be assigned a new value depending on the values of the neighbouring cells (e.g. SPREAD).

The results of transformations of spatial data can be displayed both as tables and graphically as new overlays.

31.3 Archaeological applications

31.3.1 Site location analysis

So much for the theoretical background and major principles of the various Geographical Information Systems. We shall now discuss the archaeological relevance of these techniques.

The Institute for Prehistory of Leiden University and the Museum of Antiquities in Leiden (the Netherlands) have been engaged since 1986 in a regional research programme into the Mesolithic and Neolithic inhabitation of the south of the Netherlands (province of Limburg and parts of the provinces of Noord-Brabant and Gelderland). It is an archaeological investigation of an area of about 4500 km². Up to now 3700 sites have been recorded, 600 of which were determined to date from the Mesolithic and 2300 from the Neolithic. The number of sites is still increasing. The purpose of this investigation is to be able to reconstruct the (food)economy in the period of ca. 8300 to ca. 1700 B.C. The assumption underlying the investigation is that the nature, distribution and position in the terrain of sites from a particular archaeological period...
Figure 31.3: Example of a 'point' transformation. The value of a cell on overlay C is determined by summation of the values on the overlays A and B at the same geographical location.
reflect the economic behaviour of its inhabitants. Expressed in a more abstract way, the assumption is:

site pattern = (food)economy.

The nature (the artefact composition) and the spatial distribution of sites have been disturbed to varying degrees by post-depositional processes and research factors. So only the environmental position of the sites is left as a source of information. Site location analysis is therefore one of the most important methods of analysis in this investigation. Traditionally in a site location analysis the surrounding terrain is characterised for each site. So for each site a number of environmental variables have to be determined. Instead of determining the environment for each site, this can also be done for each square kilometer. In that case each site is assigned the environmental attributes of the square kilometer in which it is located. This method was employed *inter alia* by Hamond (Hamond 1978) in the site location analysis of Bandceramic sites in the lower Rhine floodplain (Germany). For the investigation in the south of the Netherlands this method was chosen, too: instead of 3700 site determinations, 4500 determinations for square kilometers were executed. This has two advantages:

- for a good evaluation of site location analyses it is important to know the 'natural' distribution of the geographical map-units over the study area. If many sites occur on coversands this does not automatically imply a clear preference for locations on coversands; it may simply mean that these coversands are the largest geographical unit in the area. The number of sites in a particular geographic unit should always be compared to the size of that unit. Recording the environment by square kilometers also records the natural distribution

- the environmental variables of newly discovered sites are known at once.

Collecting environmental data by square kilometer is identical to the way data are compiled for a grid-GIS. Applying GIS in this archaeological investigation was therefore a logical step.

A GIS allows comparison of two different spatial phenomena. For example, the simultaneous occurrence of sand and a high ground water level can be expressed in a table or displayed on a map. The distribution of archaeological sites is a spatial phenomenon as well; the number of sites from a particular cultural phase which occur in each cell is recorded and it is possible to correlate several archaeological distribution maps or compare them to one or more environmental maps. A comparison between two geographic phenomena is based on the cell. A conclusion could be: 'in 20% of all cells sand and a high ground water level occur simultaneously.' When comparing a geographical overlay to an archaeological overlay the cell is the basic unit as well; a conclusion might then be 'in 20% of all cells sand and two archaeological sites occur simultaneously.' In archaeology however conclusions are often phrased somewhat like '60% of the sites lies on sand'. This appears to be a matter of semantics, but there is a fundamental difference. With a GIS the unit of observation is the cell, whereas in archaeological research it is the site. So a cell comprising 3 sites, counts only for one in the GIS percentage calculations! Geographically this method is correct, but archaeologically this is unsatisfactory. Three sites in a single cell signify a distinct preference for that type of environment, which should be reflected in the percentages.
31. GEOGRAPHICAL INFORMATION SYSTEMS IN ARCHAEOLOGICAL RESEARCH

as well. The value of a cell in an archaeological overlay must be taken into account, as opposed to the value of a cell in a geographical overlay. Adapting the GIS-principles to archaeological needs is therefore necessary. A new point operation, allowing for this type of archaeological analysis, has been added to the GIS.

In the investigation in Limburg six geographical variables have been recorded, viz. height above sea level, soil texture, ground water level, relief, geological homogeneity and the presence of open water. The topographical map (1:25,000) and soil map of the Netherlands (1:50,000) provided these data. In the case of the first three variables, the largest unit per square kilometer was considered to typify that cell. In the case of relief and geological homogeneity the number of contour lines (interval: 2.5 m) and the number of texture units respectively were considered characteristic. A river or brook anywhere in a square kilometer comes under the heading open water present. Apart from the geographical information the following eight archaeological overlays were made: early Mesolithic, late Mesolithic, Bandceramic, Rössen, Michelsberg, Wartberg-Stein-Vlaardingen (Louwe Kooijmans 1983), Protruding Foot Beaker and Bell Beaker culture.

These overlays were used for a site location analysis, some results from an area in Middle Limburg will be presented here. The distribution of Mesolithic, early and middle Neolithic sites over the geographical units has been compared to the ‘natural’ distribution of the geographical units themselves. Two remarkable facts emerge from the results. On the one hand there is a clear trend, extending from the early Mesolithic to the Michelsberg culture, to place the sites increasingly less often near open water and in areas with a higher ground water level (Fig. 31.4). Probably as a result of prolonged occupation (more permanent settlements) the sites were no longer situated in areas that might experience periodical high ground water levels.

On the other hand the site preference of the Bandceramic and Rössen sites is widely divergent from that of the Mesolithic and Michelsberg sites. Even accounting from the natural distribution, relatively many early Neolithic sites are located in (cover)sand areas and far from other geological units (homogeneous environment). These sites are both in size and in number of artefacts relatively small and no features were found. They appear to have been camps, strongly different from the villages with wood houses so characteristic for this period. These villages are almost always located on löss. The part of the research area analysed here, on the other hand, consists mostly of coversands. We presume that the löss-villages and the coversand-camps are part of the same exploitation system. The coversand-camps would be used for special purposes by the people having their agricultural settlements in the löss area. The camps might have been used for an ‘agricultural’ reconnaissance of the sands and/or transhumance.

The sites in the succeeding cultural phase (Michelsberg culture) exhibit a location preference strongly reminiscent of the late Mesolithic sites. Research in the delta area in the west of the Netherlands however proved that the economy of this period was based in part on agricultural products. The strong resemblance to site patterns from the late Mesolithic might indicate the importance of hunting and gathering, apart from food production. So on the sands in the south of the Netherlands the introduction of a ‘Neolithic’ economy will have been very gradual. Around 3400 B.C. it gains a foothold, but the percentage of agricultural products is still very low.
M. WANSLEEBEN

Figure 31.4: Stacked percentage histogram of the 'natural' distribution of a geographical attribute (left) and location of Mesolithic, early and middle Neolithic sites (right). The location with respect to the presence of open water (above) and soil texture (below) is shown. (EM = early Mesolithic, LM = late Mesolithic, LBK = Bandceramic culture, Rö = Rössen culture, MK = Michelsberg culture)
31.3.2 Site pattern predictions

In the process it became clear that the data stored in the GIS database could also be used in several other ways to serve the archaeological goal. Based on the various geographical overlays it is possible to predict a specific site pattern. The environment has to meet specific conditions for particular economic activities. Hunting camps will not be located where no game is present. There will be no agriculture on steep slopes. We assume that the number of sites will be large in areas where various geographical attributes favour a particular economic activity. In less favourable areas there will be less sites. Dependent on the 'natural' distribution and the presumed economy a site pattern can be predicted in this way. If the predicted site pattern corresponds to an observed site pattern, the economy will be assumed to have been based on that economic activity. If the prediction is false, a different economy will have to be assumed for that period and a new prediction made, which in turn will be compared to reality. This method of prediction and comparison was called archaeological land evaluation by Kamermans (Kamermans et al. 1985). A similar line of reasoning and environmental evaluation were used inter alia by Foley 1977, Jochim 1976, Paludan-Müller 1978.

As an example of this method we shall make a prediction about the pattern of hunting camps in a part of the research area. It is based on the assumption that in areas where various ecosystems meet (heterogeneous areas) the plant and wildlife will be more varied and therefore richer than in homogeneous areas. Intuitively we used the following evaluation scheme:

<table>
<thead>
<tr>
<th>geological heterogeneity</th>
<th>hunt evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of textural units</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2, 3, 4, 5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>relief</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of contour lines</td>
</tr>
<tr>
<td>1, 2, 3</td>
</tr>
<tr>
<td>4 to 13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>open water</th>
</tr>
</thead>
<tbody>
<tr>
<td>open water</td>
</tr>
<tr>
<td>absent</td>
</tr>
<tr>
<td>present</td>
</tr>
</tbody>
</table>

Using the RENUMBER-transformation all cells with a favourable geological heterogeneity can be selected. The same goes for the favourable relief cells. The two renumbered overlays can be combined using boolean operators. All cells meeting both (AND) or either of the two criteria (OR) are selected. A combination of AND and OR is possible too: cells meeting both criteria are valued higher than cells meeting one or none of the criteria. This generates a kind of order or rank (Fig. 31.5). We used this last method in our research. In the hunting prediction three overlays are important. The order in the evaluation for hunting is 0, 4, 7 and 9 for cells meeting 0, 1, 2 or 3 respectively of the criteria posed (Fig. 31.6).
Figure 31.5: Different ways to combine two overlays to obtain an archaeological land evaluation. All cells meeting both (AND) or either of the two criteria (OR) are selected. The cells meeting both criteria can be valued higher (darker color) than the cells meeting one criteria (RANK).
Figure 31.6: The landscape is evaluated for hunting (above) and arable farming (below). A darker color indicates a more favourable environment for that particular economic activity.
Table 31.1: Percentage distribution of Mesolithic, early and middle Neolithic sites over the hunting and arable farming evaluation overlays in the middle Limburg

<table>
<thead>
<tr>
<th>Hunting evaluation class</th>
<th>0</th>
<th>4</th>
<th>7</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Mesolithic</td>
<td>0</td>
<td>26</td>
<td>44</td>
<td>30</td>
</tr>
<tr>
<td>Late Mesolithic</td>
<td>11</td>
<td>39</td>
<td>32</td>
<td>18</td>
</tr>
<tr>
<td>Bandceramic culture</td>
<td>25</td>
<td>31</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Rössen culture</td>
<td>71</td>
<td>0</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Michelsberg culture</td>
<td>26</td>
<td>40</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>natural distribution</td>
<td>15</td>
<td>32</td>
<td>43</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arable farming evaluation class</th>
<th>0</th>
<th>4</th>
<th>7</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Mesolithic</td>
<td>35</td>
<td>4</td>
<td>22</td>
<td>39</td>
</tr>
<tr>
<td>Late Mesolithic</td>
<td>39</td>
<td>2</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td>Bandceramic culture</td>
<td>19</td>
<td>6</td>
<td>12</td>
<td>63</td>
</tr>
<tr>
<td>Rössen culture</td>
<td>14</td>
<td>14</td>
<td>15</td>
<td>57</td>
</tr>
<tr>
<td>Michelsberg culture</td>
<td>41</td>
<td>0</td>
<td>21</td>
<td>38</td>
</tr>
<tr>
<td>natural distribution</td>
<td>33</td>
<td>4</td>
<td>34</td>
<td>32</td>
</tr>
</tbody>
</table>

In the same way, a new overlay with an order for arable farming is made. Favourable criteria for arable farming are: at most 7.5 m difference in height, low ground water level, geologically homogeneous and löss or sand.

The site pattern for several archaeological period has been compared to both evaluation overlays (Table 31.1). Many Mesolithic sites are located in cells suitable for hunting (evaluation classes 9 and 6) and few in unsuitable terrain (class 0). For the early and middle Neolithic sites the opposite is true: many in class 0. During the Michelsberg phase once again areas were used that were suitable for hunting too (class 9). In the arable farming evaluation the high percentage of Bandceramic and Rössen sites occurring in areas suitable for arable farming (class 9) is particularly remarkable. During the Michelsberg, like in the Mesolithic, areas were used extensive that were not suitable for arable farming (class 0).

Although there may be minor differences, these results match quite well the site location analyses. Since both analyses are more or less similar, this is not really remarkable. In the archaeological land evaluation, however, a combination of several selected geographical attributes is considered, whereas in the site location analysis the position with regard to just a single environmental attribute is determined.

31.3.3 Pattern reconstruction

The third application of GIS refers to attempting to create a site pattern that is untouched by post-depositional processes and research factors. The archaeological distribution map always presents a more or less disturbed image of the original material legacy of prehistoric man. Changes may have been effected by geology (erosion, sedimentation), modern-day land use (urban area, wood, heath and pastures) and search
activities by (amateur) archaeologists. Because of all these factors, only a small portion of any research area will actually be investigated. We have to content ourselves with a small sample of the original population of sites. We would like to dispose of a map offering a reasonable approximation of the original site pattern. Using a GIS this was attempted in the following way.

Let us assume that a homogeneous geographical unit has actually been investigated for 50% and 2 sites from a particular period have been discovered. Since our sample in this case covers 50%, it seems justified to assume that in that particular geographical unit originally at least 4 sites must have been located. This assumption will of course have to be verified in the future. In a regional investigation so many factors are involved that it is still quite impossible to decide how typical the known sites are of the nature and distribution of the original habitation pattern. This is one of the major problems in regional archaeological research.

Each cell in a GIS is by definition a homogeneous geographical unit. So for each cell the number of sites can be revised using the method mentioned above. What it comes down to in practice is a combination of an overlay with the number of sites and an overlay with the size of the sample (%). The formula for the original number of sites in a cell is:

\[
\frac{100}{\text{sample size}} \times \text{number of known sites}
\]

To cells containing no sites that have been investigated, are attributed zero sites using this method, and justifiably so. However, problems occur when cells have not been investigated: the number of sites would always remain zero. The corrected site pattern is only valid for areas that have been investigated (Fig. 31.7). So it seems impossible to ever obtain a completely revised archaeological distribution map.

However, for this purpose the spatial information stored in the GIS can be used as well. On the basis of the geographical overlays the landscape has been classified. In this case 4 geographical attributes were taken into account, viz. height above sea level, soil texture, ground water level and the presence of open water. The landscape could be divided into 9 so called ecotopes. An ecotope is the smallest homogeneous unit in a terrain, defined as: 'a recognizable, unique, and interdependent combination of the environmental characteristics of landform, geology, soil, vegetation and/or water' (Burrough 1986, p. 4)

Techniques for distinguishing ecotopes come from the field of Physical Geography assume an ecotope consists of 30 cells. Six of these have been partly or completely investigated. After the sample correction 12 sites from a particular period are located in these 6 cells. The average number of sites in the part of the ecotope that was investigated is 2 sites per cell. This value is extrapolated to the 24 cells of the ecotope that were not investigated. In this way, an estimate can be made of the average number of sites for each ecotope. These results could be displayed graphically in a distribution map, which would be an approximation of the original site pattern. Shifts in this site pattern from one archaeological period to another might indicate changing economic activities. The nature of these changes can be derived directly from the changes in the average number of sites per ecotope.

For a small part of the coversand area data were available about the degree to which it had been investigated archaeologically. For this small area site patterns could be reconstructed. The revised Bandceramic pattern shows a relatively high number of
Figure 31.7: Attempt to reconstruct the original site pattern from the actual site pattern. The corrected pattern is obtained by a correction for the intensity of archaeological
sites in ecotope 2 and lower averages in ecotopes 3 and 4 (Table 31.2). Ecotope 2 are low lying wet sands with open water. Considering the small number of sites (7) from the Rössen cultural phase, data were not revised for this period between the Bandceramic and Michelsberg cultures. The Michelsberg pattern shows relatively high averages for ecotopes 4 and 8 and lower averages for ecotopes 3, 6 and 7. Ecotope 4 are low lying dry sands without open water and ecotope 8 are low lying dry sandy clays without open water. The main differences between the Bandceramic and Michelsberg patterns seem to be the higher ground water level and the presence of open water in the vicinity of the site of the former phase. As earlier, this could be interpreted as a result of a prolonged site occupation.

The results of this analysis do not differ to a high degree from those of the two previous analyses. In the previous analyses disturbing factors were weighed indirectly as well, since undiscovered sites are not included in a site location analysis. This is true whether you are calculating the percentage in relation to actual geographical units or calculating it in relation to derived factors like a suitability for hunting. The last analysis has undergone a correction for the disturbances, resulting in a more pronounced site pattern. In the land evaluation several geographical attributes are combined on the basis of a particular economic strategy. However, in the ecotope analysis several attributes are combined on the basis of their simultaneous occurrence in the terrain. This contrast can be considered as a ‘social’ versus a ‘physical’ classification of the terrain.
31.4 Conclusions

As demonstrated by these applications, Geographical Information Systems appear to be an important resource in archaeological research, especially in regional investigations. Now it becomes possible to manipulate large amounts of geographical and/or archaeological spatial information. It is emphatically not true that everything will be better, faster or simpler, but the available data can now be moulded to fit the typically archaeological problems and methods. We have not yet realised all the potential and implications of the use of a GIS. As far as that is concerned, we are still in an experimental stage. The archaeological analysis and interpretations should therefore be considered preliminary.

One of the major drawbacks to a good GIS is the extremely high price both for hard- and software. Many archaeological institutes cannot afford acquisitions of that magnitude. At the Institute for Prehistory, an example of a 'low-budget' institute, attempts were made to provide students and scholars with better access to automated analysis of spatial data, using existing hard- and software. A number of compromises had to be struck:

- only grid-based GIS
- only spatial databases of limited size
- only point and legend-unit transformations
- relatively slow transformation of data
- only simple graphic output

For the database management component of a GIS, within these limits, dBase III proves to be quite sufficient. The data transformation too can be handled to a large degree by dBase. For the analysis and representation of the spatial data some programs were written in BASIC and Turbo-Pascal, based on the GIS-program MAP2 (van den Berg et al. 1985). All software was designed for DOS-personal computers with 512 K memory and a Hercules graphics board.

These applications of GIS in archaeological research demonstrate that despite financial restrictions it is possible to use one of the latest developments in geography, planning and civil engineering. The advantages seem to outweigh the drawbacks. With some small modifications Geographical Information Systems can be very useful and within reach for a large group of archaeologists.

References


