Archeologia e Calcolatori
7, 1996, 129-135

UNBIASING THE ARCHAEOLOGICAL RECORD

1. Introduction

Regional GIS projects have been a "promising" archaeological activity since the technology became available in Europe in the late '80s. However, it appears that this promise is not being delivered on, and from the literature (see, for instance, Allen *et al.* 1990; Lock, Stančič 1995) a worrying picture emerges in which GIS for archaeological management purposes has acquired a certain foothold but in which its research potential is limited.

What is causing this? There are obvious technological limitations to the capability of today's GIS to handle archaeological data satisfactorily. I need only mention the issues of 3D (spatial) and 4D (spatio-temporal) GIS, the problems of dealing with 'fuzzy' data and 'fuzzy' logic, etc. But in this paper I will argue that a much more fundamental problem to the application of GIS technology in archaeology is the low quality of the data that are available to us. I will illustrate this thesis, and outline some of the strategies adopted to try and deal with the problem, using current research within the Wroxeter Hinterland Project.

2. The Wroxeter Hinterland Project

The Wroxeter Hinterland Project is a three year regional research project, based at the University of Birmingham Field Archaeology Unit, that aims to reconstruct the landscape and settlement history of a 38 by 31 km area in south Shropshire, centred on *Viroconium Cornoviorum*, the capital of the Romano-British *Cornovii* tribe (see also R. White, this volume). Largely unoccupied since the early Middle Ages, this town and its surrounding hinterland present us with one of the best-preserved archaeological landscapes in which to study urban-rural dynamics from the Later pre-Roman Iron Age (LPRIA) to the sub-Roman period (*circa* 100 BC – *circa* 600 AD) in Britain. The project's ultimate goal is to test and refine current thinking about the processes of Romanisation in Britain (see, for instance, Millett 1990). GIS is used to store and analyse existing and newly collected archaeological and geographic data sets, to manage the project's extensive field work programme, and to build and test models of social and economic processes within the study area (for an introduction to the project, see Gaffney, van Leusen 1996).

The project is a typical regional GIS study in that it has started out in the autumn of 1994 to acquire a comprehensive set of 'base maps' for the area. These consist of both archaeological records and environmental maps. Among the archaeological databases acquired were those held by bodies rang-

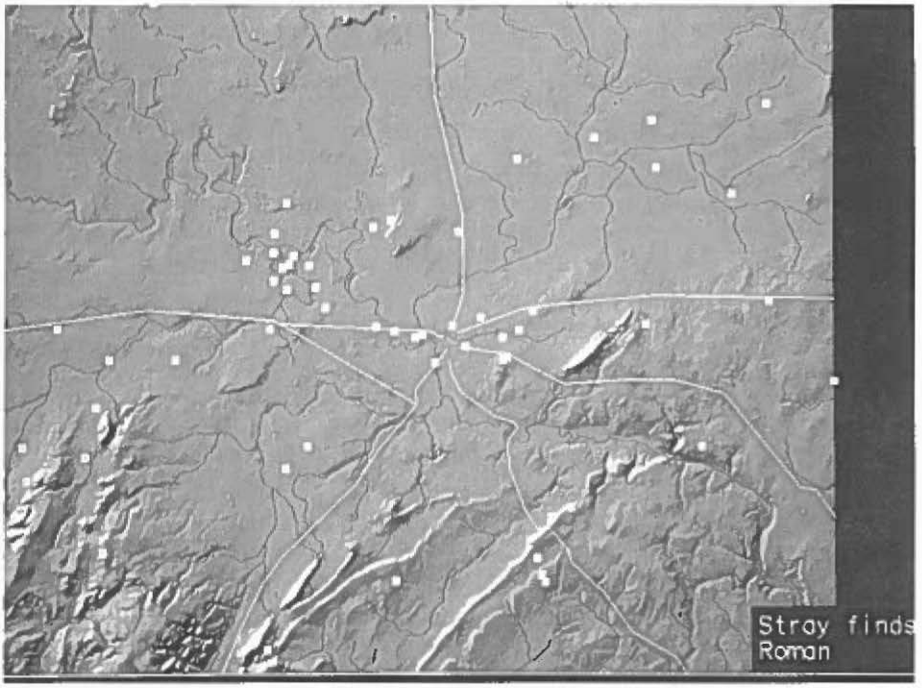

Fig. 1 – Distribution of 'stray finds' from the Roman period (source: Shropshire Sites & Monuments Records) against a background of the presumed contemporary infrastructure – roads and rivers.

ing from the Royal Commission for the Historic Monuments of England to Shropshire County Council, the National Trust, and other research projects operating in the same area. Among the environmental data acquired were maps that could be digitised from published map sheets (e.g., streams, land use, drift geology) or unpublished manuscripts (soils), and maps that were bought in digital form (DTM, Landsat TM). From these, we expect to derive yet other maps using well-known GIS techniques (e.g., local relief, slope).

At this stage of the project it was possible to produce some 'pretty pictures' (Tav. VII, a, showing the locations of archaeological site records against the background of the DTM). Preliminary selections and combinations of the base data were made and some simple spatial analysis was performed (Fig. 1, for instance, shows the distribution of stray finds from the Roman period against a background of the presumed contemporary infrastructure). We could then have gone wild and started modelling, say, the distribution of bronze age barrows in terms of the distance to the nearest watershed, amount of exposition and underlying geology. Fortunately, the project design had anticipated this and directed our research toward more fruitful, if less immediately rewarding, problems.

3. Garbage in, garbage out

One of the unexpected benefits of the use of GIS in archaeological research has been the increased visibility of gaps, biases, errors and uncertainties in the base data. Although all of this has been known for decades to archaeologists who have used maps in their work, the formalisation of map analysis caused by the advent of GIS technology has brought a new awareness of just how unreliable our current font of knowledge about past (or even present) landscapes really is. And if our current records of past human behaviour are so scanty and biased by differential preservation, visibility and reporting of finds, by the vagaries of staffing, funding and research interests, how can we trust that they reflect the past landscape in any sense (Mills 1985)? How can we trust the archaeological models – either 'theoretical' or GIS-derived – that, in one way or another, are founded upon it?

Many authors have encountered this problem in one guise or another. Chartrand and Miller (1994), for instance, highlight their difficulties in dealing with inconsistent and disparate archaeological databases in the York Environs Project. In my own work modelling Linear Bandkeramic settlement in the southern Netherlands, I found that the strongest correlation between these early Neolithic sites and 'environmental' base maps was a highly significant χ^2 of 13.6 – with built-up areas in the *modern* land use map (van Leusen 1993)! More examples can easily be found elsewhere in this volume. They all go to show the shakiness of what we like to think of as our 'knowledge' of past spatial behaviour, which in effect rarely goes much beyond conjecture. When one thinks of all the depositional and post-depositional processes, the accidents of visibility, reporting, and research which are filtering our vision of the past, this can hardly be surprising.

4. Unbiasing the archaeological record

So, archaeological records, even if they are diligently collected and carefully screened, are too heavily biased to be used directly for GIS modelling of past spatial behaviour. The models we build would probably bear little relation to past reality. An important area of research within the Wroxeter Hinterland Project is therefore to develop strategies to 'unbias' the base data. In this paper I will detail two of the many ways in which the Project team tries to do this. The first of these is what I call *bias modelling*, the second is the more traditional collection and use of control samples through field work.

5. Bias modelling

Many authors have identified factors that tend to bias the archaeological record. Examples are geological erosion and deposition (Allen 1991), land use and land cover (VAN LEUSEN 1993), and coverage and technique of archaeological surveying. By mapping such 'bias factors' and modelling their effects on both pre-existing archaeological records and the project team's newly collected data we can correct for the bias.

Taking land use in our study area as an example, we decided to map both the current (1992) land use by classifying a Landsat TM image of the area, and the land use before the industrialisation of agriculture (mapped in the 1930s by the Ordnance Survey). On the latter, we see (Tav. VII, b) that the existing recorded archaeological sites have a marked 'preference' for arable land. If we break the records down into groups of similar sites, this preference appears to be largely caused by a subset of enclosures and field systems that were identified from air photographs, a technique which is known to introduce a strong bias in favour of arable land – either freshly ploughed or under mature grain crop. Rather than interpreting the patterning we have discovered in our data as a reflection of the *original* distribution of this type of site, we would suspect it to be caused perhaps by especially disruptive agricultural practices in these areas, or by a heightened soil and crop response which makes certain archaeological features show up better in air photographs. Either way, recent and modern land use is heavily implicated in the formation of the pattern.

Having identified land use as a factor contributing bias to our recorded site distributions, it is now possible to use GIS to, firstly, quantify that bias and, secondly, compensate for it. Various ideas and GIS techniques for doing this have been proposed (VAN LEUSEN 1996, in press); they are generally referred to as *weighting* or *evaluation schemes*. Quantifying the bias in a distribution of archaeological sites could be done by directly assigning weights or values to each of the land use categories on the basis of its presumed effect on the visibility of sites. For example, high weights (i.e. low visibility) are assigned to built-up and forested areas and water bodies, intermediate weights to heath and grassland, and low weights (high visibility) to arable land. Weights or values could also be derived automatically by assuming that the known distribution of archaeological sites directly reflects differential visibility – in this case, one could, for instance, use the χ^2 or P values for random distribution of sites as the weight. It is important to remember, however, that these are just the extremes of a whole range of possible methods for creating a map quantifying a particular bias.

In the next step, the bias map is combined with the original site distribution to produce an unbiased map. This is most easily done if archaeological sites are represented as a density distribution, i.e. a continuous raster surface. The Project team will be developing methods to do this in early 1996. It should therefore be stressed that the examples presented here are for purposes of illustration only.

6. Field work

Traditionally, archaeologists have dealt with bad or missing data by collecting their own data *de novo*, using survey and excavation designs that are calculated to either "fill the gaps" in the existing archaeological record or to provide independent control samples from the same underlying population, and taking care that the areas covered and the numbers of artefacts collected are sufficiently large to allow statistical inferences to be made. The Wroxeter Hinterland Project, in addition to implementing the bias models described above, also collects such independent data, using systematic field walking and airborne remote sensing techniques.

A three year programme of extensive surface collection, carried out during the autumn and spring seasons and using our 200-strong local volunteer force, provides the Project with detailed information about artefact densities in a number of carefully selected areas. These areas, which will eventually total some 20 km^2, are spread about along a Y-shaped field work transect which cuts across the main geological structure of the study area and follows the important Severn river valley out from Wroxeter (Tav. VIII, a). The total area of the transect itself is 225 km^2, or some 18% of the study area. Positioning the transect in this manner allows us to not only look for variations that are distance dependent with respect to Wroxeter itself, but also to separate that effect from variations due to the nature of the drift geology and soils.

Finds from this survey, when plotted as point on a map, already give some idea of the intensity and distribution of background land utilisation, but by further processing these distribution data into continuous raster maps of the fieldwalked areas (Tav. VIII, b), and eventually – by interpolation – of the whole of the transect arms, the Project will have a full set of background find density maps for all periods under study at its disposal. By comparing the pre-existing archaeological records in detail to this fresh dataset, we can trace both qualitative and quantitative biases. An example of the former is the existence of unenclosed settlements in the study area, which are not picked up from air photography, but which do produce assemblages of surface finds that will be picked up during field surveys. An example of the latter is the preferential reporting of Roman pottery over less attractive Iron Age and Medieval pottery, which skews existing records whereas controlled field walking does not.

7. Future work

Although not the subject of the present paper, base 'environmental' data sets suffer from problems very similar to those discussed above. They reflect recent rather than past environmental characteristics; map printing requirements limit the type and amount of information represented, etcetera. The Wroxeter Hinterland Project team will be studying ways to improve the representation of such data in digital form, ways of quantifying their inherent biases and uncertainties, and ways of using them for the making and testing of archaeological predictions.

In addition to surface-collected data, the Wroxeter Hinterland Project uses various air photographic and remotely sensed data sets collected via the Natural Environment Research Council's airborne remote sensing platform. Multiband Thematic Mapper images and colour stereophotographs, sampling about 20 percent of the transect area, are currently being acquired to study the presence of biases in the existing air photographic records. After pre-processing of these GPS controlled data we expect to begin studying them in the autumn of 1996.

The problem of bias in base data is one that not only the Wroxeter Hinterland Project has had to deal with. Increased processing speed and power, and increasingly sophisticated GIS software, are drawing many archaeologists into the study and management of extended archaeological 'landscapes', and it is to be hoped that eventually robust methods of unbiasing base data will be developed.

MARTIJN VAN LEUSEN
Field Archaeology Unit
University of Birmingham

BIBLIOGRAPHY

ALLEN K.M.S. *et al.* (eds.) 1990, *Interpreting Space: GIS and Archaeology*, London, Taylor & Francis.

ALLEN M.J. 1991, *Analysing the landscape: a geographical approach to archaeological problems*, in A.J. SCHOFIELD (ed.), *Interpreting Artefact Scatters: contributions to ploughzone archaeology*, Oxbow Monograph 4, 39-58.

LOCK G., STANČIČ Z. (eds.) 1995, *Archaeology and Geographic Information Systems*, London, Taylor & Francis.

CHARTRAND J., MILLER P. 1994, *Concordance in rural and urban database structure: the York experience*, «Archeologia e Calcolatori», 5, 203-217.

GAFFNEY V., VAN LEUSEN P.M. 1996, *Extending GIS Methods for Regional Archaeology: the Wroxeter Hinterland Project*, in H. KAMERMANS *et al.* (eds.), *Computer Applications in Archaeology 1995*, Analecta Praehistorica Leidensia 28, Leiden, Leiden University Press, (in press).

MILLETT M. 1990, *The Romanisation of Britain*, Cambridge, Cambridge University Press.

MILLS N. 1985, *Sample Bias, Regional Analysis and Fieldwalking in British Archaeology*, in C. HASELGROVE *et al.* (eds.), *Archaeology from the Ploughsoil: studies in the collection and interpretation of field survey data*, Sheffield, J.R. Collis, 39-50.

Ordnance Survey of England and Wales, n.d. (approx. 1938), *Land Utilisation Survey of Britain*, map sheets 60: Shrewsbury and Welshpool, 61: Wolverhampton, and 70: Bishop's Castle.

VAN LEUSEN P.M. 1993, *Cartographic Modelling in a Cell-based GIS*, in J. ANDRESEN *et al.* (eds.), *Computing the Past: Proceedings of the 1992 Computer Applications in*

Archaeology Conference, Aarhus, Aarhus University Press, 105-124.

van Leusen P.M. 1996, *CRM and Academia in the Netherlands: Assessing Current Approaches to Locational Modelling in Archaeology*, in H. Maschner (ed.), *New Methods, Old Problems: Geographic Information Systems in Modern Archaeological Research*, Carbondale, Center for Archaeological Investigations (in press).

ABSTRACT

The Wroxeter Hinterland Project is a three-year regional research project employing GIS technology in the planning, management, research and publication of a study of the Iron Age and Roman landscape around the Roman city of *Viroconium* (Wroxeter) in Shropshire. The use of GIS technology in projects such as this has again highlighted the problems associated with scientific interpretation of the known archaeological record, built up mainly over the past century. In particular, the capability of GIS to recognise, extract, and extrapolate patterns in multivariate map data has meant that archaeologists must now resist the temptation to interpret such patterns as anything more than relatively recently introduce biases of site preservation and visibility. This paper explores ways in which GIS can be used to discover, and correct, such biases.