10. Geographic Information Systems and archaeology

Kenneth L. Kvamme
Arizona State Museum and Department of Anthropology, University of Arizona, Tucson, AZ 85721 U.S.A.

10.1 Introduction

The purpose of this exposition is to introduce some of the capabilities of Geographic Information Systems (GIS) technology that may be of interest to archaeologists. Reviews of hardware and software features, or even of basic concepts, are not undertaken. These topics have been presented in the archaeological literature many times before (e.g. Wansleeben 1988; Kvamme 1989; Allen *et al.* 1990) and it is time to move beyond mere fundamental descriptions. My focus is on the computational power and analytical potential of GIS, which is best carried out through the examination of several applications of the technology to archaeological domains of inquiry. My goal is to demonstrate not just the utility of GIS in archaeological applications, but that GIS is a necessity to a number of pursuits and areas of research.

10.2 Spatial data management

Spatial data base management represents, and will continue to represent, the primary use of GIS in archaeology. This application is, of course, extremely important to archaeologists working in various cultural resource management contexts. From an analytical stand-point, however, most data management efforts are not very interesting or exciting. After all, they typically consist merely of rote data queries, searches, and the production of maps of various retrieved data elements. Nevertheless, there is a large potential stemming from such fully automated cartographic data bases that only recently has been realized.

Considering the vast and diverse bodies of information that archaeologists collect and use in any region, a cohesive means to integrate these data has been sorely needed for decades. GIS certainly meets this need. Such data as archaeological site locations, site features, artifact lists, temporal and cultural affiliation, field surveyed regions, land ownership, environmental conditions like soils, elevation, and distance to water, roads, towns, and much more, can be encoded in one place, within a single data base. Better still, it is the rapid retrieval of the data and the fact that it is organized by spatial location that makes GIS so useful. Maps and various other cartographic products (e.g. orthographic views) can be quickly produced, offering an important potential. Because multiple data types from a region can be simultaneously displayed so easily, the discovery of spatial patterns and relationships in a large body of data is greatly facilitated (Farley *et al.* 1990). This constitutes what might be referred to as a spatial exploratory data analysis potential, or spatial EDA (SEDA?). Merely by viewing a prehistoric distribution of sites with respect to soils, proximity to prehistoric roads, or other kinds of information that might exist in a comprehensive data base, one might be able to realize spatial relationships or other patterns that might exist. Such results could lead to additional, and more formal, analyses. Hence, the use of spatial data bases in GIS contexts could constitute an important first step in regional investigations.

10.3 Computational power

Although spatial data management applications offer great potential in regional studies, they do not illustrate the computational power of GIS. A good illustration of this power lies in the derivation of visibility data, or what is termed 'viewshed analysis'. Based on a computer representation of a landscape (in computer parlance a digital elevation model or DEM), a viewshed includes those areas that are visible from a specified point on the landscape. Algorithms for computing a viewshed are not that complicated; they basically involve the determination of whether an unblocked line-of-sight exists between the point of interest and every other point on the landscape (usually to some specified resolution, e.g. 20m). Their application represents a considerable computational task, however.

A viewshed analysis conducted in Tonto National Monument, Arizona, U.S.A., provides an example. Upper Tonto Ruin, a fourteenth century pueblo located in a large rockshelter overhang, is the Park's major attraction. With encroaching development in this desert region and near the National Monument's boundary, there is much concern to preserve the natural view from the ruin. In order to confront this issue, Thomas Potter, a graduate student at the University of Arizona, conducted a GIS-based viewshed analysis from the Upper Tonto Ruin. The result, when mapped, indicated a substantial region that could be seen from the ruin, much of which was outside the Park's boundary (Fig. 10.1). This analysis has contributed to Tonto National Monument's consideration of a management plan that includes acquisition of those lands to ensure that development will not occur on them, thus preserving the natural scenery and view of the surrounding desert.

10.4 Derivation of new data

One of the most valuable features of GIS is their ability to generate or derive new information from other data that may be encoded within the system. Frequently the new information is more useful for analysis or other purposes than the original data. A good example is the interpolation of elevation surfaces and the derivation of terrain form measures from digital elevation models (DEM), raster matrices of elevations with regular grid spacing. Based on digitized elevation contour lines obtained from topographic map sheets, DEM can be created by interpolating elevations at grid points based on their spatial relationships with nearby contour lines of known elevation (see Kvamme 1990a for more details). The DEM, in turn, can be employed to determine such terrain features as ground steepness (slope), direction of ground facing (aspect), or even drainage courses through algorithms that consider interrelationships between adjacent or nearby elevations in the matrix. Some of these procedures and their resulting products are illustrated in a later chapter of this volume (see Kvamme, Chapter 16).

Other examples of the derivation of new information include the generation of distance-to-water data based on

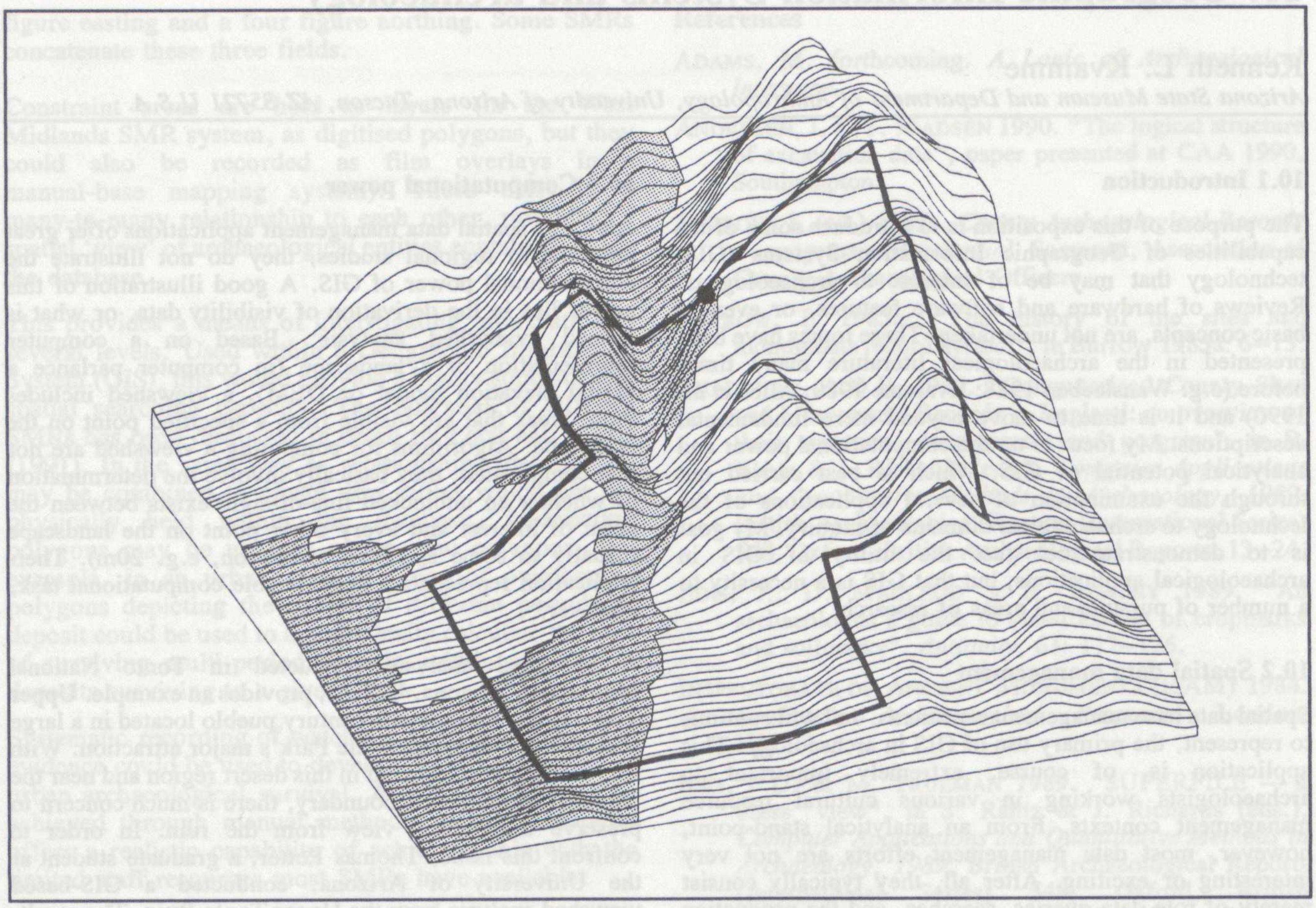

Figure 10.1: GIS-computed viewshed from Upper Tonto Ruin, Tonto National Monument, Arizona, U.S.A. The solid circle represents the ruin, the shaded area the viewshed, and the bold line the park boundary.

the locations of river and stream courses or, as illustrated in the next section, the computation of a distance-to-road surface from digitized road locations.

10.5 Locational analysis

Archaeologists often analyze archaeological distributions with respect to some background feature of interest. For example, analyzing site locations in terms of water distances, south or north facing slopes, or specific soil classes has been undertaken for some time (e.g. Plog & Hill 1971; Shermer & Tiffany 1985). Such studies require measurements at each archaeological site location and frequently measurements from the background environment as well. With traditional manual techniques obtaining numerous measurements on map sheets can be a laborious and time-consuming undertaking.

A study by Hodder and Orton (1976:226–229) provides a good illustration. They examined the distribution of 173 Iron Age inscribed or 'dynastic' coin finds with respect to Roman road locations in central and southern England. Statistical analysis revealed a significant association between the coin distribution and the Roman road network, and several hypotheses were advanced to explain this observed relationship.

The manual analysis required a considerable investment of labour (Orton pers comm. 1991). In addition to measurements of the distance to the nearest Roman road at each of the 173 coin locations, Hodder and Orton (1976) also had to obtain measurements to characterize the background distribution of Roman road distances throughout the entire 23,000 square kilometre study area. For the latter task they settled on three distance categories that were drawn on a base map of the study area. The proportion of the total study area within each distance category was determined by superimposing a fine-mesh grid, counting the number of grid units that fell in each class, and computing their proportion relative to the total number of grid units in the entire study area. A graph of the cumulative distribution of Roman road distances for the study area as a whole was obtained by plotting these known values cumulatively, and interpolating the remainder of the distribution. The cumulative coin distribution was then statistically compared with this background distribution using the Kolmogorov-Smirnov one-sample test (Hodder & Orton 1976:226; see also Kvamme, Chapter 16 of this volume, for further discussion of this analysis). Orton (pers comm. 1991) reports that this analysis was tremendously time-consuming involving days of work.

The same analysis in a GIS setting can be a nearly trivial exercise. To illustrate, the Roman roads, major water bodies, and the 173 Iron Age coin locations were digitized directly from the map shown in Hodder and Orton (1976:227). This task required about an hour's work (Fig. 10.2a). The next step required the computation of the distance to the nearest Roman road on a systematic basis (i.e. at a regular interval or grid spacing) over the entire study area, in order to provide

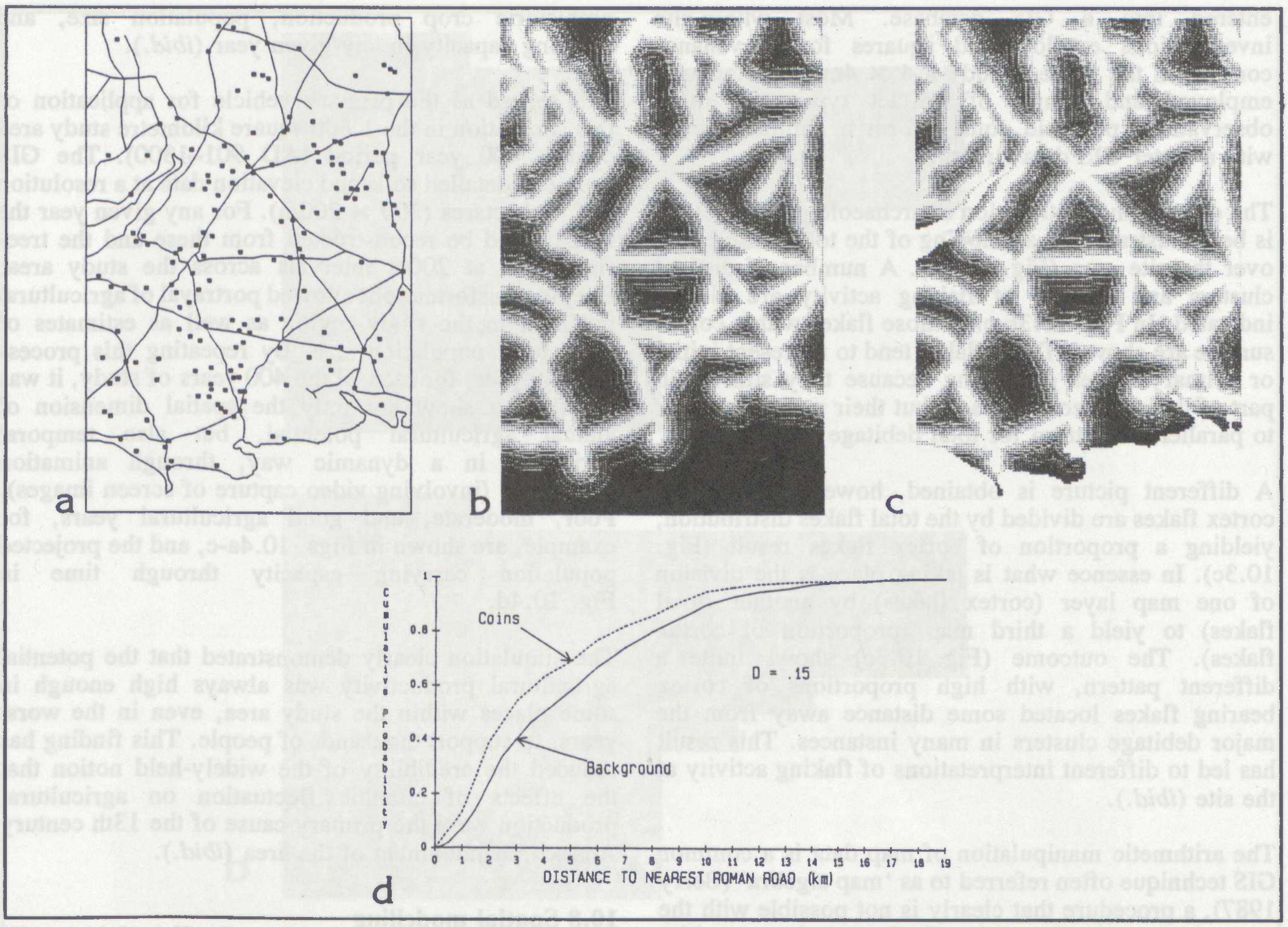

Figure 10.2: a) The Roman road network and Iron Age coin distribution in central and southern England (after Hodder and Orton 1976:227). b) The GIS-produced distance-to-Roman-road surface. c) The distance-to-Roman-road surface minus major water bodies. d) Kolmogorov one-sample test for the Iron Age coin distribution.

a data set that could be employed to characterize the background distance-to-road distribution. A grid spacing or resolution of one square kilometre was deemed more than sufficient to accurately approximate the true background distribution (Fig. 10.2b). The rather large distances that occurred in the English Channel (Fig. 10.2b) had to be eliminated next; this was accomplished by subtracting off those grid cells located within water bodies to yield the final analysis matrix containing approximately 23,000 distances (Fig. 10.2c). These tasks required approximately 10 minutes. Finally, the Roman road distances at each of the 173 coin locations were extracted from this data base, their cumulative distribution was obtained and plotted against the cumulative background distribution based on the 23,000 distances, and the Kolmogorov test was recomputed (Fig. 10.2d). This took another 10 minutes.

Even though the same conclusion was reached, that the coins tend to lie in a significantly closer proximity to Roman roads than dictated by chance (Fig. 10.2d), a number of comments are warranted. The results presented here are much more accurate than those obtained by Hodder and Orton (1976) because instead of using only three distance categories and error-prone manual methods to approximate the background distance-to-road distribution, 23,000 distance measurements were employed with a one kilometre resolution. This more accurate approximation yielded a conclusion somewhat less statistically significant than suggested by Hodder and Orton (1976:228), with a maximum separation between the cumulative distributions of only $D=.15$ (Fig. 10.2d), compared to their $D=.19$ (both, however, exceed the Kolmogorov .01 critical value of $d = 1.63/\sqrt{n} = .12$, where $n = 173$). Finally, this analysis required only an hour and a half to complete (as opposed to days in the original investigation), but I had to wait nearly 15 years for the technology!

10.6 Map algebra and within-site analysis

The graphical output of GIS is nothing new in archaeology. Archaeologists have employed computer graphics for a long time to produce cartographic products. Various CAD (Computer-Aided Design) and drawing programs remain a popular means for producing maps today. While many of these programs can produce images that are equal to, or even exceed, the quality obtained by most GIS, the latter retain a distinct advantage because information remains linked with the images, allowing much greater flexibility.

Over the past several years I have investigated the surface archaeology of a highly deflated and exposed ridge in a remote region of western Colorado, U.S.A. (Kvamme & Larson 1991). More than 25,000 artifacts have thus far been mapped over a 360 x 220m region. Nearly 50 types of artifacts have been recognized and a variety of measurements and observations have been entered into a GIS database. Most within-site investigations employ grid squares for provenance control; in the present study a 4 × 4m grid square is employed and counts of artifact types and other observations per grid square form a natural linkage with a raster GIS data structure.

The density and distribution of archaeological materials is best illustrated by a mapping of the total flake count over the site area (Fig. 10.3a). A number of distinct clusters and regions of flaking activity are clearly indicated. In Fig. 10.3b only those flakes with a cortex surface are shown. These flakes tend to represent initial or primary stages of flaking because they still retain part of the outer core surface, but their mapping seems to parallel the pattern for total debitage in general.

A different picture is obtained, however, when the cortex flakes are divided by the total flakes distribution, yielding a proportion of cortex flakes result (Fig. 10.3c). In essence what is taking place is the division of one map layer (cortex flakes) by another (total flakes) to yield a third map (proportion of cortex flakes). The outcome (Fig. 10.3c) shows quite a different pattern, with high proportions of cortex bearing flakes located some distance away from the major debitage clusters in many instances. This result has led to different interpretations of flaking activity at the site (*ibid.*).

The arithmetic manipulation of map data is a common GIS technique often referred to as 'map algebra' (Berry 1987), a procedure that clearly is not possible with the static images produced by CAD or other graphics systems. It is only because GIS retains the information linked with each image that such an undertaking is possible. Other examples include adding one map to another (e.g. early projectile points + late projectile points = total projectile points) or multiplying a map by a constant value (e.g. elevation in feet × .3048 = elevation in meters).

10.7 Simulation of spatial processes

Recently, the power of GIS for simulating spatial and temporal processes has been demonstrated by Van West (1990). A major purpose of that study was to examine whether a series of droughts were responsible for the abandonment of one portion of the American Southwest (an area in southwestern Colorado, U.S.A.) occupied by a prehistoric farming culture, the Anasazi, during the late thirteenth century A.D. This task was accomplished largely through the reconstruction of annual soil moisture conditions, as measured by the Palmer Drought Severity Index (PDSI), and tree-ring widths which are indicators of climatic fluctuation. Contemporary temperature and precipitation data recorded during the past century were used to calculate PDSI values for the multiple soil classes and elevation ranges in the study area. Next, historic tree-ring width data were correlated with these PDSI values. It then became possible to retrodict the PDSI for all soil groups and elevation ranges backward in time to the prehistoric period on the basis of the prehistoric tree-ring record. Finally, agricultural yield studies and the reconstructed PDSI were employed to estimate prehistoric crop production, population size, and carrying capacity in any given year (*ibid.*).

GIS served as the primary vehicle for application of this simulation in the 1,800 square kilometre study area over a 400 year period (AD 901–1300). The GIS provided detailed soils and elevation data at a resolution of four hectares (200 × 200m). For any given year the PDSI could be reconstructed from these and the tree-ring data, at 200m intervals across the study area. Simple transformations allowed portrayal of agricultural potential in the study region as well as estimates of prehistoric population size. By repeating this process year-by-year, for each of the 400 years of study, it was possible to show not only the spatial dimension of annual agricultural potential, but also temporal variations in a dynamic way, through animation techniques (involving video capture of screen images). Poor, moderate, and good agricultural years, for example, are shown in Figs. 10.4a-c, and the projected population carrying capacity through time in Fig. 10.4d.

The simulation clearly demonstrated that the potential agricultural productivity was always high enough in some places within the study area, even in the worst years, to support thousands of people. This finding has reduced the credibility of the widely-held notion that the effects of climatic fluctuation on agricultural production were the primary cause of the 13th century Anasazi abandonment of the area (*ibid.*).

10.8 Spatial modelling

In recent years archaeologists have shown tremendous interest in developing models of archaeological distributions over broad areas (Kohler & Parker 1986; Judge & Sebastian 1988). A large part of this interest originally stemmed from cultural resource management projects conducted primarily in the western U.S. On many large tracts of government controlled land various agencies required models of archaeological site location to be developed based on knowledge gained from sites discovered in sample surveys. In other words, spatial models were generated from sample data and employed to indicate likely locations of archaeological sites in unsurveyed regions (Judge & Sebastian 1988). These models have been used for the discovery of new sites, much like a 'prospecting' tool, but more importantly they have provided a useful planning mechanism for management agencies by their ability to indicate archaeologically sensitive regions where land disturbance or development should be avoided (Parker 1986). Good locational models also serve another useful purpose: they can provide considerable insight into patterns of prehistoric land use by removing statistical noise and portraying only the principle trend derived from a body of multivariate data. The essence of spatial patterning thus can become clearer offering great heuristic potential in regional studies (e.g. see Kvamme & Jochim 1989).

While numerous methods and approaches for archaeological modelling have been pursued (see Kvamme 1990b for a recent overview), GIS probably constitute the single major factor that has promoted the growth of interest and applications in this topic. The

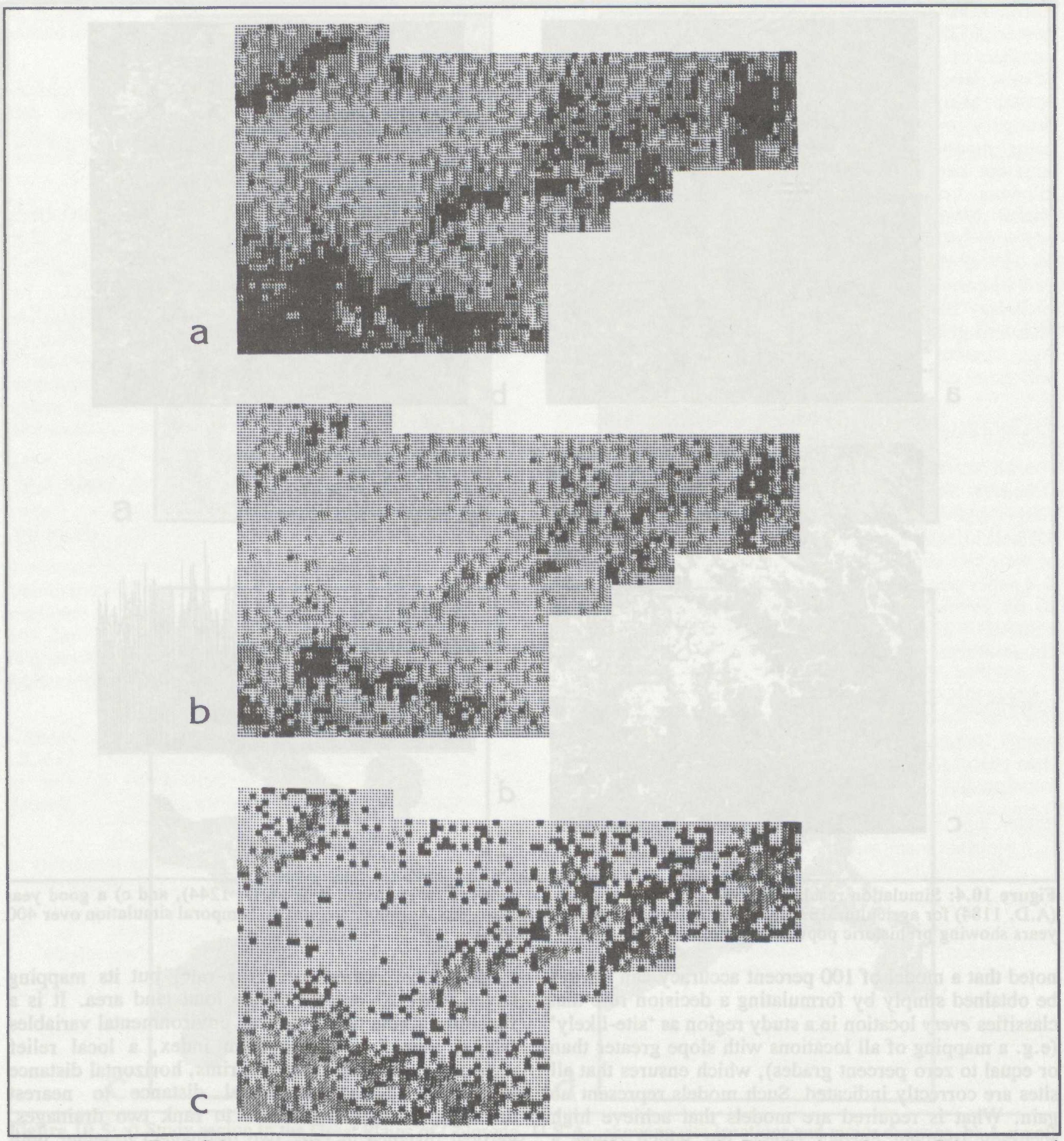

Figure 10.3: a) Total flake count distribution at a western Colorado, U.S.A., site. b) Distribution of cortex flake counts. c) Distribution of the proportion of cortex flakes. In all maps dark shading represents large values.

reason for this is the vast number of measurements and calculations that are required in order to do modelling on a regional scale, regardless of approach. To illustrate, Fig. 10.5a portrays an 8 × 5km study region in the Great Plains of southern Colorado, U.S.A., in which 95 lithic scatters are known to occur. This figure actually shows the slope surface, containing about 19,000 gradient values, where the steep-sided walls of several canyons that dissect the plain are indicated with darkest gray tones. The sites appear to exhibit a locational preference for proximity to the canyon rims, but on relatively level surfaces.

We can make use of the latter observation to produce a simple site location model. Through a standard GIS *reclassification* procedure our model will include all those locations with a slope value less than or equal to a ten percent grade. The mapping, shown in Fig. 10.5b, illustrates principally the exclusion of the steep cliff faces along the major canyons. This 'model' represents a high degree of accuracy, since about 91 percent of the known sites are correctly classified, but it offers little gain or power. The latter stems from the fact that the model mapping covers about 96 percent of the total land area, representing no improvement over chance. This can be put into better perspective if one considers that a model covering 96 percent of the landscape should include, by chance, about 96 percent of the sites. That only 91 percent are included is not a significant achievement. Furthermore, it should be

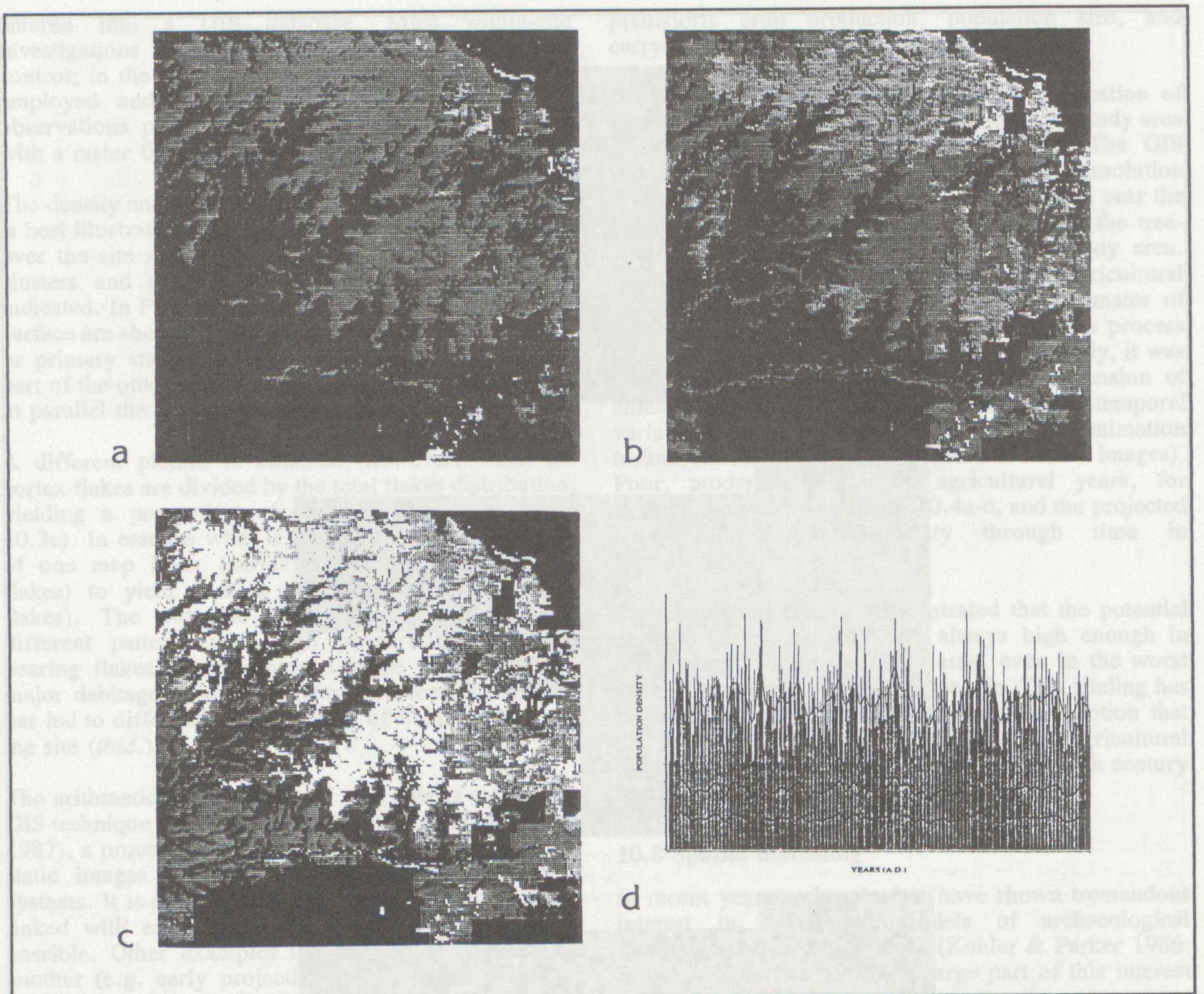

Figure 10.4: Simulation results showing a) a poor year (A.D. 1150), b) a moderate year (A.D. 1244), and c) a good year (A.D. 1184) for agricultural productivity in a southwestern Colorado, U.S.A., study region. d) Temporal simulation over 400 years showing prehistoric population potential (after Van West 1990).

noted that a model of 100 percent accuracy can always be obtained simply by formulating a decision rule that classifies *every* location in a study region as 'site-likely' (e.g. a mapping of all locations with slope greater than or equal to zero percent grades), which ensures that all sites are correctly indicated. Such models represent no gain. What is required are models that achieve high accuracy (percent correct rates), but which cover a minimum area when mapped.

This reduction-of-model-area concept is better illustrated by considering a second variable, proximity to secure water. The mapping in Fig. 10.5c represents all locations in the study region that meet the previous level ground condition *and* all locations within 400m of rank two water courses. This model offers about the same level of accuracy (89 percent correct), but is much more powerful because its mapping covers only 33 percent of the study area, representing a considerable improvement or gain over chance.

The previous models are rather simplistic decision rules. A more complex and powerful logistic regression model is illustrated in Fig. 10.5d. This model maintains a 95 percent correct accuracy rate, but its mapping covers only 20 percent of the total land area. It is a function of nine GIS-produced environmental variables (slope, aspect, a terrain form index, a local relief measure, proximity to canyon rims, horizontal distance to nearest drainage, vertical distance to nearest drainage, horizontal distance to rank two drainages, vertical distance to rank two drainages) together with the fourth-order polynomial terms of a trend-surface obtained at the archaeological sample points and a sample of control points taken randomly from the background environment (Kvamme 1988).

The models in Fig. 10.5 can be regarded as being largely descriptive of the site locational tendencies in this region of study. At the same time they also can be considered predictive tools when their mapping is applied to regions not yet surveyed by archaeologists to indicate likelihoods of site presence. In other words, if the sites in Fig. 10.5 constitute only a sample of sites from that region, then the models represent predictions of site location in unsurveyed areas.

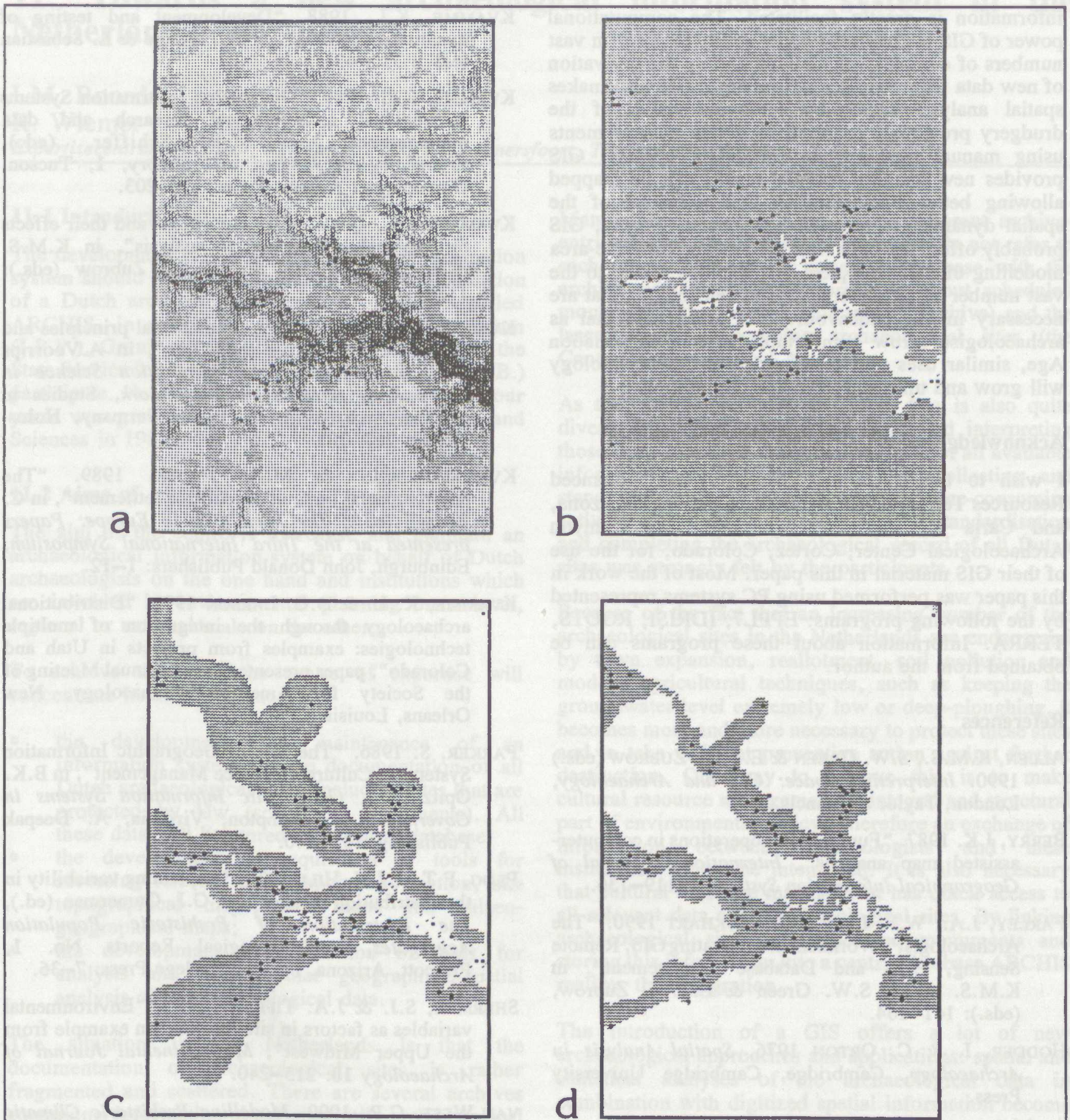

Figure 10.5: a) Study region in the Great Plains of Colorado, U.S.A., showing site locations and major canyons. b) A simple site location model produced by mapping all locations with slope less than or equal to a ten percent grade. c) A more complicated site location model consisting of locations with level ground (as in b) within 400m of secure water. d) A complex logistic regression model of site location that is a function of nine environmental variables.

These examples of modelling amply illustrate both the power and complexity of results that can be obtained, as well as the absolute necessity of GIS. Archaeological models generally are functions of environmental data that must be applied systematically across landscapes. Consequently, systematic measurements of the necessary environmental data are required over the entire area of study. In the foregoing, raster GIS provided the ideal vehicle for this endeavour because each environmental data theme needed by a model contained about 19,000 measurements at a regular grid interval of 50m. Besides providing the necessary measurements for model building and application, GIS also produced the cartographic results together with the important model performance statistics (i.e., percent correct sites and model areas). Thus, GIS clearly offers to archaeology a comprehensive means for spatial modelling.

10.9 Summary

In the foregoing sections I have presented a number of archaeological uses and applications of GIS technology. From an analytical stand-point GIS has great potential in the domain of spatial database management because the discovery of relationships in large bodies of information is greatly facilitated. The computational power of GIS is illustrated by its ability to perform vast numbers of complex calculations and by the derivation of new data from existing information. GIS also makes spatial analysis easier by removing much of the drudgery previously required to obtain measurements using manual methods. In simulation contexts GIS provides new insights because results can be mapped allowing better understanding and portrayal of the spatial dynamics of simulated processes. Finally, GIS probably offers the only workable means for wide-area modelling of archaeological distributions owing to the vast number of measurements and calculations that are necessary in this endeavour. It is anticipated that as archaeologists grow comfortable in this Information Age, similar uses and applications of GIS technology will grow and ultimately become commonplace.

Acknowledgements

I wish to thank Thomas Potter, of the Advanced Resources Technology Program, University of Arizona, and Carla R. Van West, of the Crow Canyon Archaeological Center, Cortez, Colorado, for the use of their GIS material in this paper. Most of the work in this paper was performed using PC systems represented by the following programs: EPPL7, IDRISI, ROOTS, TERRA. Information about these programs can be obtained from the author.

References

ALLEN, K.M.S., S.W. GREEN & E.B.W. ZUBROW (eds.) 1990. *Interpreting Space: GIS and Archaeology*, London, Taylor & Francis.

BERRY, J.K. 1987. "Fundamental operations in computer-assisted map analysis", *International Journal of Geographical Information Systems* **1**: 119–136.

FARLEY, J.A., W.F. LIMP & J. LOCKHART 1990. "The Archaeologist's Workbench: integrating GIS, Remote Sensing, EDA and Database Management", in K.M.S. Allen, S.W. Green & E.B.W. Zubrow, (eds.): 141–164.

HODDER, I. & C. ORTON 1976. *Spatial Analysis in Archaeology*, Cambridge, Cambridge University Press.

JUDGE, W.J. & L. SEBASTIAN (eds.) 1988. *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modelling*, Washington, D.C., U.S. Bureau of Land Management, Department of Interior, U.S. Government Printing Office.

KOHLER, T.A. & S. PARKER 1986. "Predictive models for archaeological resource location", in M.B. Schiffer, (ed.), *Advances in Archaeological Method and Theory*, **9**, New York, Academic Press: 397–452.

KVAMME, K.L. 1988. "Development and testing of quantitative models", in W.J. Judge & L. Sebastian (eds.): 325–428.

KVAMME, K.L. 1989. "Geographic Information Systems in regional archaeological research and data management", in M.B. Schiffer, (ed.), *Archaeological Method and Theory*, **1**, Tucson, University of Arizona Press: 139–203.

KVAMME, K.L. 1990a. "GIS algorithms and their effects on regional archaeological analysis", in K.M.S. Allen, S.W. Green & E.B.W. Zubrow (eds.): 112–126.

KVAMME, K.L. 1990b. "The fundamental principles and practice of predictive modelling", in A.Voorrips (ed.), *Mathematics and Information Science in Archaeology: A Flexible Framework*, Studies in Modern Archaeology, **3**, Bonn, Germany, Holos-Verlag: 257–295.

KVAMME, K.L. & M.A. JOCHIM 1989. "The environmental basis of Mesolithic settlement", in C. Bonsall (ed.), *The Mesolithic in Europe: Papers presented at the Third International Symposium*, Edinburgh, John Donald Publishers: 1–12.

KVAMME, K. L. & D.O. LARSON 1991. "Distributional archaeology through the integration of multiple technologies: examples from projects in Utah and Colorado", paper presented at the Annual Meeting of the Society for American Archaeology, New Orleans, Louisiana.

PARKER, S. 1986. "The role of Geographic Information Systems in Cultural Resource Management", in B.K. Opitz (ed.), *Geographic Information Systems in Government*, **I**, Hampton, Virginia, A. Deepak Publishing: 133–140.

PLOG, F.T. & J.N. HILL 1971. "Explaining variability in the distribution of sites", in G.J. Gumerman (ed.), *The Distribution of Prehistoric Population Aggregates*, Anthropological Reports No. 1, Prescott, Arizona, Prescott College Press: 7–36.

SHERMER, S.J. & J.A. TIFFANY 1985. "Environmental variables as factors in site location: an example from the Upper Midwest", *Midcontinental Journal of Archaeology* **10**: 215–240.

VAN WEST, C.R. 1990. *Modelling Prehistoric Climatic Variability and Agricultural Production in Southwestern Colorado: A GIS Approach*, unpublished doctoral dissertation, Department of Anthropology, Washington State University, University Microfilms, Ann Arbor.

WANSLEEBEN, M. 1988. "Applications of Geographical Information Systems in archaeological research", in S.P.Q. Rahtz (ed.), *Computer and Quantitative Methods in Archaeology 1988*, British Archaeological Reports (International Series) **446**(ii), Oxford, British Archaeological Reports: 435–451.