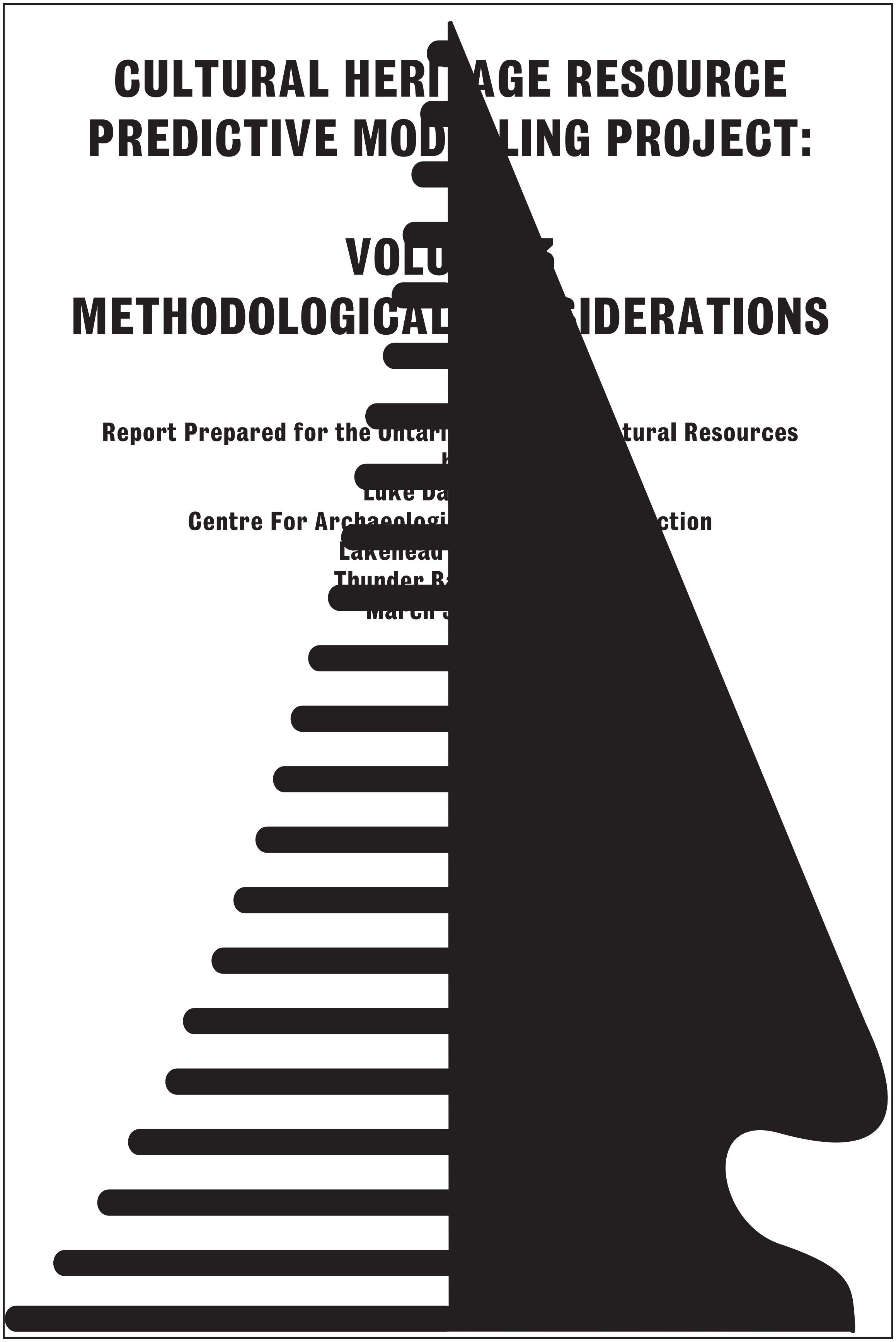
CULTURAL HERI AGE RESOURCE PREDICTIVE MOD LING PROJECT:

VOLU
METHODOLOGICA SIDERATIONS

Report Prepared for the Ontari tural Resources
b
Luke Da
Centre For Archaeologi ction
Lakehead
Thunder B
March



EXECUTIVE SUMMARY

The theoretical and applied aspects of conducting archaeological predictive modelling are a relatively new field within archaeology. It has its basis in studies conducted during the 1950s and 1960s but gained prominence during the late 1970s and 1980s and coincided with a surge in cultural resource management in the United States. During the 1980s the development of geographic information system (GIS) technology resulted in its integration in archaeological predictive modelling. Predictive models developed to date are either inductively or deductively derived. Inductively-derived models are dependent upon a database from which to generate models and thus, are subject to any biases existing in the database. Deductively-derived models begin with theories predicting human behaviour. While deductive models better encompass the range of human behaviour, they suffer from changing interpretations and theoretical viewpoints. Two main directions are taken in the development of a predictive model: the numerical approach and the weighted value approach. The numerical approach makes use of statistical methodology to discover associations among archaeological sites and characteristics of the physical environment. Within these paramters, models are physically generated by either an intersection or weighted value method. The intersection method begins with the basic assumption that all variables used in the generation of a predictive model contribute equally to the determination of site location potential. Calculating high, medium, low potential areas is simply a process of determining where the the greatest number of variables that converge in a given location. The weighted value method begins with the basic assumption that each variable contributes differently to the final determination of site location potential. This is accomplished by developing and applying a weighting scale which effectively ranks variables numrically. Site potential is determined by the arithmetic addition of all variables. Areas of high potential will have the largest numeric values and areas of low potential will have smallest numeric values. During the development of a predictive model, a number of issues must be considered. These include the representativeness of the variables to that being modelled, the quality of databases consulted, the scale at which modelled should take place and the manner in which potential is presented. Predictive modelling is presented as a three stage process. Primary stage predictive modelling includes hypothesis development, organization and data collection. Secondary stage modelling includes initial model development and testing and is the stage where most predictive models stop. Tertiary stage modelling includes continued application of the model and ongoing refinement. Ideally, tertiary stage modelling is a never ending process whereby lessons learned from previous model applications are incorporated into new and future applications maintaining or increasing the predictive robustness of the model.

The introduction of geographic information systems into archaeological research had two profound results. The first was that the application of research approaches such as predictive modelling could now be effected over relatively large areas. Secondly, the use of geographic information systems allowed for the uniform analysis of large areas. Concurrently, the use of GIS introduced a range of considerations not traditionally a part of archaeological research. Issues surrounding digital data, cartographic theory, and general data integrity became an integral part of research design and strategy.

ACKNOWLEDGEMENTS

On behalf of the Centre for Archaeological Resource Prediction, I would like to acknowledge the support of the following agencies without whom this research could not have been conducted:

The Ontario Ministry of Natural Resources
The Ontario Ministry of Culture, Tourism and Recreation
The Ontario Ministry of Northern Development and Mines
The Ontario Heritage Foundation
Lakehead University

A large number of people have worked in the office and in the field over the last two and half years. While they are too numerous to list here, their assistance and contributions are appreciated.

We also received assistance and advice from Canadian Pacific Forest Products, specifically from Bill Murphy, Phil Keenan, Derek Dool and Bill Roll. Bill Smith and Mike Lipa from Abitibi-Price, Inc., were also quite helpful and all have earned our gratitude.

The Steering Committee which oversaw the progress of this project also deserve thanks for the advice they offered and the support given at crucial times. The Steering Committee is Dr. Jean-Luc Pilon, Mike Opper, Bill Ross, Del Parker, Peter Carruthers, Dr. Patrick Julig, Robert Cormier, John Osborn and Dave Gordon.

Finally, special thanks go to Dave Gordon, Bill Ross and Del Parker. These gentlemen were involved in this project at a detailed level. Their advice, patience, humour and ability to come up with crucial resources at critical times were integral to the success of this project.

TABLE OF CONTENTS

REVIEW OF PREDICTIVE MODELLING AND ARCHAEOLOGICAL INFERENCE

GEOGRAPHIC INFORMATION SYSTEMS AND ARCHAEOLOGICAL RESEARCH

GIS AND ARCHAEOLOGY: PREDICTIVE MODELLING RESEARCH DESIGN CONSIDERATIONS

LIST OF FIGURES

LIST OF TABLES

REVIEW OF PREDICTIVE MODELLING AND ARCHAEOLOGICAL INFERENCE

1.1 INTRODUCTION

> Predictive capacity without explanatory capacity is worthless. Mere clairvoyance irrespective of its sharpness, does not have scientific standing. Only predictive capacity that arises out of having coherent and communicable explanations has scientific standing. The power to predict is subsidiary to the power to explain (Liebenstein 1976:13 in Kohler and Parker 1986:397).

Predictive modelling is an avenue of research within archaeology that has gained prominence over the past two decades. Predictive modelling for archaeology is defined as a

> ... simplified set of testable hypotheses, based either on behavioral assumptions or on empirical correlations, which at a minimum attempts to predict the loci of past human activities resulting in the deposition of artifacts or alteration of the landscape (Kohler 1988:33).

Parker (1985) sees predictive modelling as a natural outgrowth of the theories and methodologies of spatial archaeology and predictive modelling has become the focus of a number of archaeological studies (e.g., Allen et. al. 1990; Brown and Stone 1982; Judge and Sebastien 1988; Kvamme 1992).

1.2 SETTLEMENT STUDIES

Predictive modelling has its basis in the settlement studies first carried out in the 1950s and 1960s by Gordon Willey and other archaeologists. Willey (1953) intended to examine archaeological data on a regional level in an effort to understand the processes inherent in settlement systems in the Virú Valley in Peru. On the whole, Willey was successful at developing settlement pattern archaeology, and his work provided the stimulus for other settlement studies to be conducted. Resulting publications by Willey (1956), Willey et. al. (1965), Chang (1968) and Adams and Nissen (1972) established studies of settlement patterns as a valued research methodology within archaeology. Settlement pattern studies were further refined and used by archaeologists to conduct catchment analyses (Vita-Finzi and Higgs 1970), interpret social and technological change at the regional level (Adams and Nissen 1972), while others focused on the environmental determinants of settlement location (Haury 1956; Heizer and Baumhoff 1956; Williams 1956).

Throughout much of the 1950s and 1960s, archaeologists operated within an inductive framework where research into settlement patterns was based upon little or no theory. Haggett et. al. (1965) provided a more solid grounding for locational theory to archaeologists by introducing many relevant concepts into the discipline from geography. He influenced a generation of archaeologists by outlining theories of settlement hierarchies, sampling procedures and hexagonal lattices (Haggett et. al. 1965). Trigger (1968) outlined more clearly the various aspects of settlement patterns and offered some determinants of settlement location. Concurrent research in other fields of archaeology was beginning to emphasize the importance of ecological variables in understanding settlement variability (e.g. Flannery 1968).

In the decade that followed the 1960s, the manner in which archaeological data was handled changed considerably. Many archaeologists adopted more systematic approaches to collecting and analysing data. The use of computers allowed for the manipulation of greater amounts of data, the generation of more detailed analyses and more generally, for a greater variety of questions to be asked of the data by archaeologists at the time. Studies ranged from examinations of minute differences in artifact types, to macroscopic studies of ceramic variability, to studies of prehistoric culture change (e.g. Flannery 1976). These studies contributed to further refinement of the level of detail in which settlement variability was presented by archaeologists.

As a result of settlement pattern studies, the research emphasis of some archaeologists was shifting from the study of single sites to the study of regions and their archaeological contents. For example, following closely from the settlement studies discussed above, the Southwestern Anthropological Research Group (SARG) set out to determine “why prehistoric populations locate sites where they did” (Plog and Hill 1971:8). Clearly stated in this research goal was the delineation of the “formal variability in sites, variability in temporal loci of sites, and variability in the spatial loci of sites” (Plog and Hill 1971:8). Indeed, settlement pattern research had turned, at least in print, from the elementary description of archaeological remains to the recognition of site distribution patterning.

SARG presented a detailed research design for the study of human settlement systems. They recognized that a regional approach to studying variation in human settlement patterns was absolutely necessary to understand settlement systems. Previous research in the American southwest concentrated primarily upon a few core areas and these interpretations were then generalized for the entire region. The need for more detailed and standardized investigations prompted the formation of SARG. Foremost among the goals outlined by the project leaders was the explanation of:

> variability in the distribution of prehistoric sites - settlement and limited activity sites… Why do we want to explain site location or settlement system patterning?… The most important reason for explaining settlement locations is that we hope to arrive at tested and useful laws that can be used by social scientists to *predict site locations anywhere at any time, including the present and the future* (Plog and Hill 1971: 10-11, orig. emphasis).

The majority of settlement studies carried out in the Americas contained more description of settlement locations than explanations for their specific existence. Plog and Hill (1971) recognized the need for explanation of the ‘system behind the settlement pattern’ but strove to arrive at explanation via other avenues. Realizing that the explanation of settlement systems derives from an understanding of their mechanics, SARG sought to predict unknown site locations from the principles of the known settlement systems. Thus, SARG’s goals anticipated those of many archaeologists by several years. At the same time, much of the discipline was embroiled in a methodological debate concerning paradigms and polemics, but several research projects that complimented the directions and goals set out by SARG eventually emerged.

Plog and Hill (1971) were not the only archaeologists intent upon predicting site locations. Although not an explicitly stated research strategy, prediction as a subset of settlement pattern analysis was making its way into the archaeological literature. Perhaps the first settlement pattern study designed to identify sites using prediction was that carried out in the Reese River Valley in the Great Basin of the American Southwest (Williams et. al. 1973). The authors carried out a settlement pattern study in central Nevada focusing on winter village placement. They stated that “given the proper set of environmental conditions, [they] could successfully predict presence/absence of archaeological sites" (Williams et. al. 1973:215). Wanting to confirm their intuition about 'where sites could be found', they developed hypotheses based on those intuitions and measured their soundness. Variables, definitions and criteria used to develop their predictions were carefully outlined as follows (Williams et. al. 1973:227):

> 1)The locus should be on a ridge or saddle.
> 2)The ground should be relatively flat (relatively flat <5% slope).
> 3)The locus should be in the low foothills (low foothills <250 meters above the valley floor).
> 4)The locus should be within the modern piñon-juniper ecotone.
> 5)The locus should be near the modern piñon-juniper ecotone (near <1000 meters).
> 6)The locus should be near a semi-permanent water source (near <1000 meters).
> 7)The locus should be some minimal distance from this source (some minimal distance >100 meters).

These criteria were not revolutionary by any means. In fact, they appear to be criteria quite obviously related to site location. What was new about these criteria was their clear definition and implementation in the overall research strategy. If any five of the seven criteria were met, "the locus was recorded as an area of potential habitation, whether or not cultural material was found" (Williams et. al. 1973:231).

The results of the research were positive. The variables outlined above were shown to be present at 97% of the sites in the study area while 85% of the potential loci contained sites (Williams et. al. 1973:233). Although the authors acknowledged that refinements could be made to the prediction criteria, on the whole, they were successful. The authors showed that no one variable determined the location of a prehistoric habitation in this area.

> In spite of the fact that a single locational criterion would not significantly restrict the spatial distributions of sites, combinations of two or more mildly restrictive criteria would quickly reduce the number of possible locations that will fit the specified criteria (Williams et. al. 1973:234).

On a more general level, the authors confirmed the suspicion held by many archaeologists concerned with location of sites: something acknowledged as a 'feel' or 'insight' gained from intimate familiarity with the data (Williams et. al. 1973:217). Indeed, new archaeological insight was gained into the prehistoric inhabitants of the Reese River Valley regarding their choice of activity loci. In this example, it was successfully demonstrated that prediction could provide insight into the explanation of the settlement system.

At approximately the same time, another settlement pattern study was carried out in the British Honduras (now Belize) by Green (1973). Drawing her methodology primarily from Haggett et. al. (1965), Green mirrored the questions posed by SARG: "the analysis is aimed at answering the question: why did the ancient inhabitants settle where they did?" (Green 1973:279). Although the author's primary goal is to explain the variability in settlement locations:

> a corollary goal of the analysis is to predict the location of sites in portions of the region which have not yet been explored archaeologically. Prediction, in this case, is based on determining the correlation between sites and environmental features in the known region and projecting this knowledge to environmentally similar areas. The method can also suggest locations within the study area

> which should be rechecked for the presence of undiscovered sites (Green 1973:279).

Central to Green's analysis was the proposition that sites were located in order to minimize the effort expended in acquiring critical resources (1973:279). Green worked with a partial sample of the entire archaeological database. The results showed a strong association between site locations, soil types and vegetation (Green 1973:287). Also apparent from the study was the importance of proximity to navigable water. In fact, the author concludes that the location of every site in the sample can be explained by association with these three variables (Green 1973:289). Green's attempt to predict locations of undiscovered sites based upon the above criteria met with less success than did the Reese River Valley study discussed earlier. High measures of variability were generated from statistical tests. Areas that were predicted to have potential for site location were very large and impractical for efficient survey due to accessibility and the nature of the physical landscape. Overall, despite some of the questions raised by Green's conclusions and the pioneering nature of her study, the results of her attempts to use prediction to help explain the settlement pattern were promising.

While some archaeologists were utilizing more sophisticated analytical techniques in performing regional archaeological analyses, many more archaeologists used analytical techniques scarcely more advanced than Willey's (1953) work. Peregrin's description of work conducted earlier in his career exemplifies this point:

> We began by laying out the Rosario phase 1:20,000 map. One inch colored beads were used to designate sites by level of population, mounded architecture, specialized activities, pottery characteristics, etc. *By standing up on stools we could get a visual impression of settlement patterns and other prominent aspects of the regional system* (emphasis added) (1988:875).

In summary, predictive modelling developed from studies of settlement patterns. Settlement studies often provided data on site location and their distribution and it was with this information that researchers attempted to predict other, unknown, site locations. The first predictive models attempted to turn an understanding of specific settlement systems into predictions of site location which would hopefully contribute to explaining the settlement pattern. The above examples of prediction within settlement studies are representative of the directions and results of research in this field of archaeology. By the 1980s, researchers built upon the base provided by settlement studies. Predictive modelling became the subject of much research, but its role in settlement pattern research diminished as it was applied more and more as a cultural resource management tool.

1.3 PREDICTIVE MODELLING

The literature concerning predictive modelling has become more extensive throughout the 1980s. For the most part, it has reflected the fact that developing a predictive model is not as elementary as that outlined by Williams et. al. (1973). Although much of the literature and examples of predictive models is buried in government files and consulting reports to business, some academic archaeologists doubt the efficacy and value of prediction in archaeology (Kohler and Parker 1986:396). It is seen as an expensive exercise to discover the obvious, regarded as suspect or unreliable or being limited in value (Kohler and Parker 1986:398). The concerns of cultural resource managers, contract archaeologists and academic archaeologists have resulted in a body of literature that begins to address some of the issues relevant to developing a predictive model (Brown 1981; Carr 1985; Kohler and Parker 1986; Limp and Carr 1985; Ebert and Kohler 1988; Judge and Sebastien 1988; Kohler 1988; Kvamme 1988a, 1988b, 1989, 1990; Warren 1990). Kohler has contributed extensively to the literature of predictive modelling, both in published and unpublished (contract/government) areas and sees predictive modelling developing in two directions: that is, inductive versus deductive modelling (Kohler and Parker 1986:399; Kohler 1988:37), elsewhere called the behavioral approach (Hay et. al. 1982:14).

1.3.1 Inductive Models

The roots of the inductive model can be traced to the research conducted by Steward (1938) and Willey (1953). These archaeologists focused their analysis at the regional level rather than at the level of the site itself. These pioneering investigations, coupled with the increasing archaeological insistence upon representative sampling "has set the tone for two decades of work in CRM and in non-CRM work" (Kohler and Parker 1986:399).

An inductive model usually begins with data and then builds its conclusions based upon all the biases inherent in the original data set.

> They begin with survey data… and then they estimate the spatial distribution of the population of archaeological materials from which the sample was drawn… Any inferential locational model predicts only what would have been found had the population of space from which the sample was drawn been surveyed in the same manner as was the sample, using the same rules for attribute coding, site recognition and data analysis. Such inferential models predict neither the systemic interaction between a cultural system and a landscape nor the archaeological context resulting from it; rather they predict what we will find and how we will interpret it if we consistently follow a particular set of rules (Kohler 1988:37).

Inductive models form the basis for a large percentage of predictive models developed to date. Since for many areas of North America there already exist large site databases, their examination could provide tremendous amounts of site-related information. In fact, these data are readily integrated into many predictive models.

> [F]or this particular exercise the computerized database (AZITE computer database) faithfully represents our current knowledge of site location… [and] contains a variety of descriptive information pertaining to the environment, location, cultural affiliation, site function and temporal components… (Altschul 1990:228).

While existing databases contain a wealth of invaluable information, these data are not without error and bias. For example, site locations may be incorrectly recorded, environmental information may be recorded in too little detail, data may be missing from some records, or information gathered by previous researchers may differ in quality compared to the standards of present-day archaeologists. More seriously, systematic biases may exist in the current site inventory. As this biased inventory will form the basis of the search for landscape correlates of site distribution, the original biases will be perpetuated into the resultant predictive model. That is not to say that this information should not be used; rather, it should be used carefully only after evaluation of its integrity as a complete database.

1.3.2 Deductive Models

Deductive models are seen by Kohler to begin with a theory predicting human behaviour.

> The challenge for deductive models is to build the bridge to the analytic context from the systemic context, which is where the outputs of the

> system can be observed. This bridge-building… is called explanation (Kohler 1988:37).

He, with Parker, sees deductive models as encompassing three considerations.

> A deductive model must:
>
> 1)consider how humans make choices concerning location ... This requires considering: (a) a *mechanism* for decision making; and (b) an *end* for decision making- what is the goal?
>
> 2)specify the variables affecting location decisions for each significant chronological or functional subset of sites;
>
> 3)be capable of operationalization; it must propose a means for measuring each of the relevant variables and must allow for a set of predictions that can be compared with the archaeological data (Kohler and Parker 1986:432).

A number of interesting points can be raised when considering deductive models. For example, environmental variables are often considered by archaeologists to be important in conditioning the choice of activity location by prehistoric people. Most predictive models make the fundamental assumption that “settlement choices made by prehistoric peoples were strongly influenced by characteristics of the natural environment” (Warren 1990:202). This assumption figures prominently in determining which environmental characteristics or variables are used in the modelling process. An examination of the literature reveals some of the most basic environmental variables used in predictive models: elevation, slope, aspect, and distance to water (Kvamme 1985; Parker 1985; Altschul 1990; Carmichael 1990; Warren 1990). However, most researchers recognize that a wide range of environmental considerations are important including vegetation changes over time as well as the use of various plants for medicinal purposes.

> From the standpoint of human adaptation, patterns of local vegetation are of crucial concern. Many plants serve as primary food and technological resources as well as secondary resources which attract economically important animals. The distribution of non-food resources, especially water and fuel, can be equally important to settlement decisions. Diversity is also beneficial when considering non-food resources. In addition to fuel, a variety of trees provide the raw materials for tools, utensils, shelter, and weapons, pitch for sealing seams, and fibres from the inner bark for cordage, bags, and nets. A variety of plants can be used to make dyes, reeds can be woven into mats, and clay from local stream banks can be made into pottery. Evaluations of topography, water, soils, vegetation, precipitation, temperature, and availability of rock outcrops or glacial till exposures are all important in decisions about the adequacy of shelter and the availability of economic resources (Schermer and Tiffany 1985:220).

Dean (1983:11) has pointed out that people may look for only a few clues in their surroundings when identifying and selecting activity locations, rather than processing the entire range of environmental “cues” available. It may only be these basic variables that really have any association with archaeological sites. This raises interesting questions about the analysts’ choice of the proper environmental variables for inclusion in the modelling process.

> Perhaps in building predictive models we are too ready to make the assumption that only a complex multivariate model can adequately

> account for human locational behaviour, when in fact, a few (proxy?) variables, observed in the highly correlated data base that is our environment, may be sufficient for forming locational decisions (Kohler and Parker 1986:433).

Support for this position lies with the fact that archaeologists have presented successful predictive models using very few variables. For example, Altschul (1990) developed a predictive model for the 9,000 acre Mount Trumbell area of Arizona. There were 228 known sites in the study area that had been sampled by various agencies in the past. Three environmental variables were identified which account for the majority of site locations that include elevation, slope and aspect (Altschul 1990:229-230). Altschul concluded that in this area "over 70 per cent of all component locations can be predicted with just three variables" (1990:234). However, Altschul does not discuss what his three variables are measuring. What are they 'proxy' variables for? Without this information, we are unable to discuss why sites are being found where they are nor are we able to offer explanations for settlement systems in the area.

Another point involves the consideration of land use choice derived from 'habitual behaviour' derived from cultural norms, traditions and spiritual proscriptions, rather than an overriding consideration of the economic attractiveness of a specific locality (Kohler and Parker 1986:435 citing Wright and Dirks 1983). Factors related to actions having little archaeological visibility, such as spiritual influences, may have resulted in activities being located in less 'typical' locations. Choice of activity location may also be the result of historical events that override environmental considerations. Other criteria have been recognized by archaeologists to be important in choosing activity location. Flannery (1976) and Reynolds (1976) discuss social factors that condition site placement. Jochim (1976:12) details criteria of economic relevance and assumes that "the determination of resource use tends to precede and condition the site placements and demographic arrangements of a hunter-gatherer group". A predictive model may take into account distance to resources and activities carried out at a location. Wood (1978:161-162) offers the following criteria for different site types:

> 1)Limited activity sites will be located so that the distance between a site and the resource indicated by the activity will be minimal;
>
> 2)Multiple activity sites with dominant subsets of activities will be located so that the distances between a site and the matching resources indicated by the dominant subsets are minimal;
>
> 3)Multiple activity sites will be located so that the acreage distance to all of the critical resources is minimal.

Kohler and Parker (1986:438) point out that although Wood's propositions model, in a more realistic manner, the possible decisions made at those locations, applying them in a predictive sense could prove to be more difficult.

1.3.3 The Numerical Approach

In addition to the inductive and deductive theoretical frameworks, the methodological approaches employed in predictive modelling may be separated into two different groups. The first is described as the numerical approach, and the second as the graphical approach. The numerical approach may be considered a direct outgrowth of the emphasis placed on the statistical analysis of archaeological data since the early 1970s. Predictive models using the numerical approach employ multivariate statistics as a *discovery* technique to identify associations among variables which ultimately lead to predictions of areas with archaeological resources.

This approach makes a number of primary assumptions that are crucial to the validity of the model. The first relates to the nature of the sample. Because statistical methodology discovers meaningful associations among variables from known site information, it is important that the known site information is representative of the actual sites that exist.

> Probabilistic designs are of little use if the population sample is not the same as the population across which predictions are to be made (the target populations)… As one practitioner remarks, '[We] cannot make inferences about the archaeology of verdant grasslands with good intermittent and permanent streams from a sample restricted to scoria ridge tops, badlands and breaks' (Peebles 1983:8) (Parker 1985:406).

Roper echoes this view in her comments on a predictive model developed for the Vermilion River/Embarass River region of Illinois:

> Methodologically, multiple regression should eventually be a valuable predictive tool but its use with the poor data available for east central Illinois is unwarranted. The discriminant function analysis at the end of the report is an interesting idea, but I wonder if it is really describing where sites are located or where people have intuitively felt they should be and have therefore looked for them (Roper 1981:149).

Thus, users of predictive models derived using the numerical approach must carefully evaluate the nature of the existing database. In addition to a very careful examination of the representativeness of these data, an assessment must be made as to whether known site locations reflect the actual distribution of archaeological sites, or simply reflect where archaeologists have conducted their surveys (e.g., Acheson and French 1992).

It is also important to recognize that the physical and cultural environment has changed over time, and these changes may have affected the choice of activity location through time. Kohler and Parker state that

> ... despite numerous studies in diverse areas indicating change in site location through time in response to changes in adaptation type, and despite evidence that within any adaptation type, functional subsets of sites may have differing environmental determinants, most empirical correlative models aggregate sites of all types and ages together for prediction (1986:408).

Models developed using the numerical approach rarely address temporal considerations. Some researchers opt to avoid the issue of 'time' and develop a generalized model,such as that generated by Lewis and Murphy (1981). Other researchers do not avoid 'time' as a variable, rather it is suggested that discernible patterns of human behaviour cross-cut considerations of time. This perspective is discussed by Kvamme (1992:23).

> By associating sites representing many different functional, chronological, and cultural types into a single open-air class, a great deal of locational variability is introduced to the modeling problem, thereby reducing the potential power of the result. Nevertheless, it is believed, and it has been elsewhere shown (e.g., Kvamme 1985, Kvamme and Jochim 1989), that there are common locational tendencies that may cross-cut functional categories, such as preferences for level ground or proximity to water.

Few researchers have developed models applicable to specific time periods (e.g. Lewis and Murphy 1981). The reasons for this are not clearly presented in the literature. In fact, there seems to be a fixation to create **one model** to explain everything as if one magic set of variables could predict all site types in all time periods. At one level, it is recognized that many different factors influenced site location through time just as different factors influenced the location of different site types. Perhaps controlling for many factors including changes in physical geography, climate, flora and fauna, cultural groups and technology proves too formidable a task for archaeologists working under tight budgets and/or strict mandates. Whatever the reason, the majority have developed predictive models that encompass all prehistoric time periods and all site types.

Another consideration relates to the choice of variables, and the detail with which information will be selected and manipulated. The choice of variables is determined by the nature of the predictive modelling project, the type of data available, the nature of the study area, and other considerations. Parker (1985) describes two characteristics of variables used in predictive models. The first, site-focused data, all require measurement at the site level. Examples of site-focused data include distance to water, vegetation and slope. The second characteristic, and the one Parker suggests is commonly employed, is quadrant data. These are data that are generalized from survey quadrants. In some cases, where a high resolution model is being developed, quadrant data may closely resemble and augment site-focused data. In other cases, where coarse resolution models are developed, the quadrant data may generalize the study area to the point where the data are less meaningful than is preferred (Kohler and Parker 1986:408).

An example of research using the numerical approach is Sandra Parker's Sparta Mine predictive model. In this study, Parker aims to

> …develop an explanatory model relating site locations in an area to the biophysical characteristics of that area. To perform the desired functions, such a model must allow one to state the probability that a particular geographic unit in the area would have been selected for the location of a site. Such a model may be in the form of a prediction equation in which the dependent variable is site presence/absence and the independent or predictor variables are the biophysical variables (Parker 1985:176).

Parker's primary means of discovering associations between variables and site locations is multivariate statistics. Two basic data collection methods were employed. First, biophysical data were collected from United States Geological Survey (USGS) 7.5 minute topographic maps to provide the independent or predictor variables for the entire Sparta area. Secondly, a field survey was conducted to provide data about site presence/absence, the dependent variable (Parker 1985:182). The model was evaluated using a number of different tests: observed vs. predicted site frequencies, cross-validation tests, and field tests. Overall, Parker (1985:198) demonstrates by these tests her "confidence in the validity of the model".

The numerical approach is certainly a valuable method which can lead to the discovery of significant associations between site locations and variables. However, it is an approach which requires a high degree of statistical training and competence in order to develop the model, interpret the results and validate/replicate the results. Invariably, with the results of the model presented in a numerical table outlining the associations between variables and sites, a great deal of interpretation is required to relate the results to 'on-the-ground' locations. Roper, commenting on a specific predictive model, and summarizes some deficiencies of the numerical approach.

> While the authors make a reasonably good start at such, they fail to produce a satisfactory end product be-

> cause of naive use of statistics. They begin with cross tabulations of variables… cross tabulation of each variable with each other variable is not an efficient use of statistics, and does not discern those variables that do or do not have predictive power. Further, this report declines to summarize those statistics into a meaningful interpretation (i.e., predictive model) of site location patterns; rather, it assumes that the tables will speak for themselves. The text reflects neither a good understanding of statistical analysis nor an ability to employ statistics in interpretation of archaeological data (Roper 1981:150-151).

Roper's criticism of the above model does not invalidate the use of statistics as a means for developing predictive models. In fact, the development of predictive modelling is coincident with the use of advanced statistical techniques. However, despite the sustained use of statistics in archaeology, there are still those, including some developers of predictive models, who are familiar with only basic statistical procedures and tests. The use of multivariate logistical regressions requires an advanced level of understanding of statistical theory and techniques. Accepting the validity of a model like Parker's requires a tacit acceptance of the calculations and results presented. Verification and/or duplication of the methodology might prove daunting to some archaeologists who may ultimately accept the results primarily on faith. This may contribute to the statistical approach not being the choice of some developers of predictive models. Thus, while the statistical approach is still used to generate valid predictive models, other models utilize different approaches.

1.3.4 The Graphical Approach

The second methodological approach used in predictive modelling research developed as a consequence of technological changes that have taken place since the early 1980s. This approach also involves the development of predictive models using environmental variables, but differs from other approaches in using a graphical methodology using map overlay techniques. This technique is primarily achieved using computer software such as geographic information systems (GIS). The different variables are represented on separate computer map layers, and these map layers are combined in ways that identify areas spatially associated with valued landscape characteristics. Different combinations of variables give rise to various stages of the predictive model. Finally, apparent associations between variables and sites are evaluated using statistical techniques. With this approach, statistics are not used as a means of *discovering* associations during model development. Instead, statistical techniques are used as a means of *evaluating* the strength of association between variables and sites after the model is applied.

An example of this graphical approach is the model developed for the Mt. Trumbell area of northern Arizona. Jeffrey Altschul began the study by asking cultural resource managers if modelling was a useful tool for their specific needs. He discovered that what

> …managers need to know is where the 'red flags' are…what is needed are not models predicting the unknown, but rather models that bring some order and direction to the huge databases that have been, and are continuing to be, amassed (Altschul 1990:227).

In other words, in some jurisdictions data has been amassed at such a rate that information managers cannot adequately cope with it. Predictive modelling becomes useful as a means of aiding the identification of landscape variables that are consistently correlated with known site distributions. These correlates offer a means of organizing the existing database, and identifying presently uninvestigated localities which have a high probability of containing sites similar to the presently known sites. In essence, this is not investigating the unknown, rather it is merely investigating more of the same, only focussing on areas not yet field surveyed. This can be viewed as modelling existing assumptions and expectations. To predict the unknown, implies and requires that archaeologists step outside what is 'expected' and employ a modelling rationale that does not build exclusivity into its results. From its very inception, this type of modelling approach necessarily will consider all possible options allowing areas to be excluded because of the manner in which variables interact.

With these red flag models, Altschul takes an entirely different approach from previous orientations. While valuable, this approach has weaknesses. The predictions generated tend to perpetuate the 'presently known' site distribution, and enable the prediction of 'average', repeatedly used, site locales that cluster around stable and important environmental variables. However, these models, no matter how sophisticated, are not discovering anything more than new sites that conform to presently known site distribution patterns. However, using the graphical approach, one may identify landscape characteristics that are associated with a significant proportion of the site inventory, and thereby highlighting a minority of sites that clearly are different. This smaller subset can then be subjected to another round of analysis to determine patterns of association with different landscape variables. This second round of analysis may lead to significant new insight.

A predictive model that seeks site localities that do not conform with the known pattern offer advantages in that it offers new information of a different order. That is, types of sites, land use strategies, and idiosyncratic behaviour that are presently unknown. In the event that the model identifies sites that do not conform to the expected site distribution, a resource manager is in the position to re-focus research to identify new sorts of ancient land use that are not immediately apparent in the current heritage resource inventory. This will result in an ongoing refinement of what appears to be of 'low potential' for containing archaelogical sites.

Sites in settings presumed to be anomalous according to conventional wisdom, by definition, must be the result of behaviour that does not fit current explanations of why prehistoric inhabitants settled where they did. Under any definition, these sites must be significant, for they more than any others have the potential of telling us something new about prehistory (Altschul 1990:228). Additionally, as more anomalies are identified, patterns may emerge, become predictable and therefore are no longer anomalous. Those sites whose locations remain anomalous become the target of further study for it is these sites that will give us greater insight into the past (Altschul 1990:228).

Altschul developed a predictive model for the 9000 acre Mount Trumbell area of Arizona. There were 228 known sites in the study area that were sampled by various agencies in the past. Three environmental variables were defined to account for site location: elevation, slope and aspect. Altschul outlines four steps in the development of his predictive model (1990:230-232):

> <u>Step 1: Data exploration</u>
> Descriptive data are compiled on spatial location, temporal affiliation, and site function. The relationship of these variables with each environmental variable is confirmed through the use of simple associational statistics.
>
> <u>Step 2: Confidence and independence</u>
> The second step involves assessing the environmental variables in terms of confidence and generalization. Altschul wants to determine whether survey coverage of each environmental zone is adequate in order to have

> any faith in the resulting distributions. He also wants to determine the degree of statistical independence between the three environmental zones.
>
> Step 3: The favourability map
> The third step is to create a map showing the relative probability of each map unit to contain a site.
>
> Step 4: The red flags
> Once the favourability map is created, site locations are cross tabulated with favourability zones. Statistics are once again used to confirm associational significance of environmental variables and site locations.

Altschul's approach differs from that of other researchers in that he is focusing on predicting anomalies rather than predicting the already known. He is trying to identify why sites are located in unexpected places; in other words, why did we not expect them to be there? Rather than "viewing models as end products, we view them as analytical tools" (1990:237). Altschul's graphical approach is more useful to cultural resource managers because patterns or non-patterns are more readily apparent. While statistics are used for validation and confirmation, the results need not be translated from statistical tables to archaeologically-meaningful statements. Using the graphical approach

> ...at a glance, managers can determine the likelihood of determining sites on a particular development project. For archaeologists, [a map such as this] represents a compilation of the relationship between environmental variables and site location (Altschul 1990:233).

1.4 MODELLING PROCEDURES

There are two primary modelling procedures that can be extracted from the existing literature. These procedures refer to the manner in which different variables are manipulated to produce a predictive model. These may be called the intersection method and the weighted value method.

1.4.1 The Intersection Method

The intersection method is perhaps the most commonly used methodology for developing predictive models. The intersection method begins with the basic assumption that all variables used in the development of a predictive model contribute equally to the determination of site location potential. Calculating high, medium or low potential areas is simply a process of determining the number of variables that converge in a given location. Areas where the highest number of variables occur can be labelled 'high potential' while areas where the lowest number of variables occur can be labelled 'low potential' (Figure 1.1).

The assumption that all variables contribute equally to the determination of the predictive model is one that does not accurately reflect the complexity of human land use decision-making. For example, if modelling the location of prehistoric fishing camps, variables such as 'proximity to water' are of greater significance than such variables as 'vegetation zone'. Thus, while the intersection method is employed in a number of predictive models, it does not result in a model which faithfully reflects the true range of prehistoric decisions employed in determining the location of activities.

1.4.2 The Weighted Value Method

The weighted value method is perhaps the least used methodology in the development of predictive models. The weighted value method begins with the basic assumption that each landscape variable contributed differentially to ancient land use decision-making. To account for this, each landscape variable is given a different numeric weight to reflect its deemed importance. For example, this arbitrary weighting scale, might offer a range from 0 to 3 whereby 0 = poor, 1 = fair, 2 =

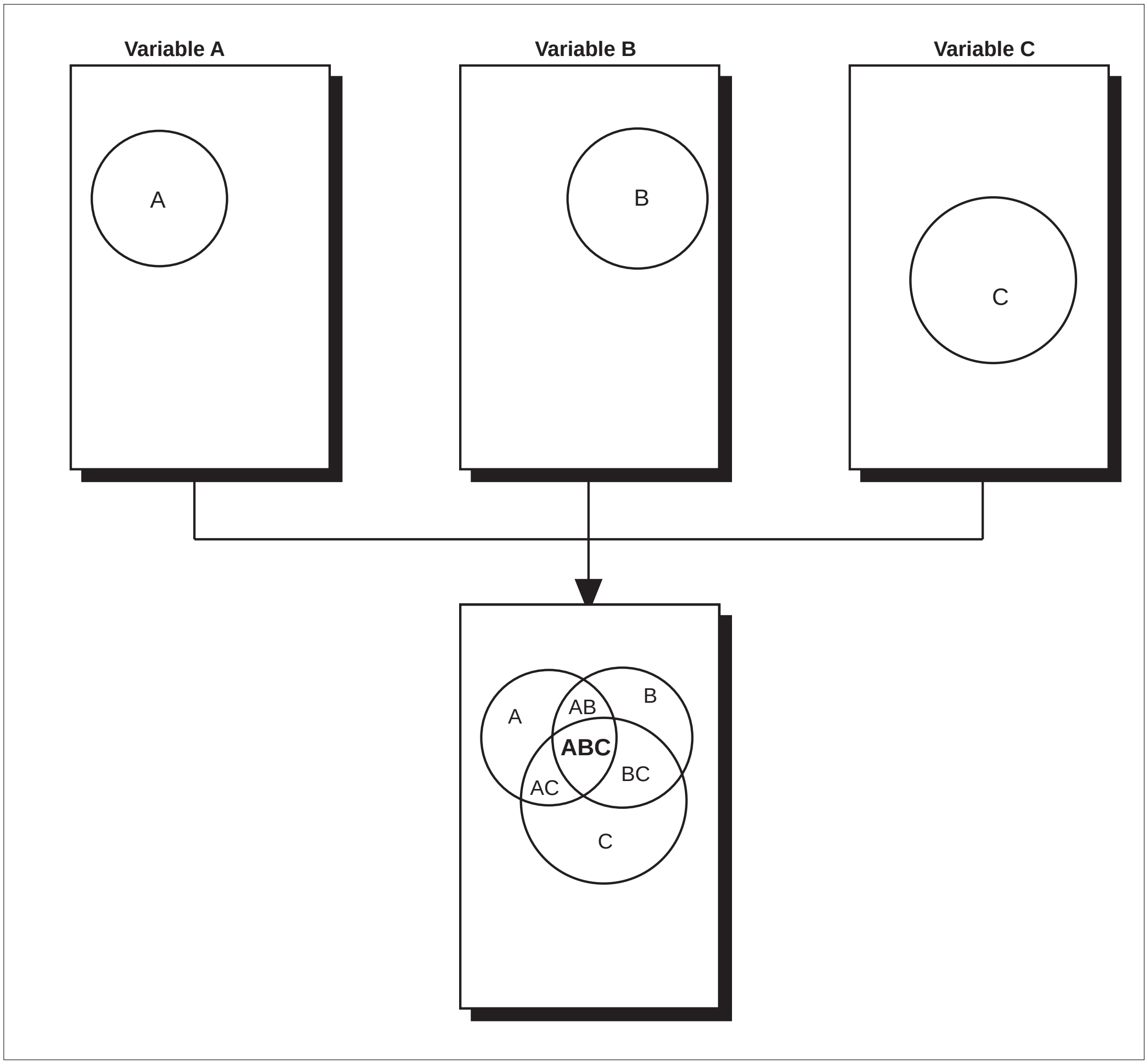


Figure 1.1. Schematic Diagram Illustrating The Principle Of The Intersection Method Where The Area Identified As "ABC" Would Be Considered High Potential, "AB, AC, BC" Medium Potential and "A,B,C" Would Be Considered Low Potential

good, and 3 = excellent. Additionally, the variables to be modelled are divided into categories and subcategories. Categories encompass broadly defined divisions such as *Proximity to Water* or *Soil Drainage*. Subcategories encompass detailed subdivisions of categories. For example, if the category was *Proximity to Water*, the subcategories might be *Major Waterways, Minor Rivers, Minor Lakes*. Variables include even finer divisions of categories. Thus, for example, if the subcategory was *Major Waterways*, then the variables might be *0-100m from water, 100-250m from water*, and *250-500m from water* (Table 1.1).

A value (V) is applied by the researcher to the category to reflect its importance and contribution in the modelling process. In addition, variables are assigned weights (W) to reflect differences within categories in their contribution to the modelling process. For example, the category *Proximity to Water* might be given a value of 3 to reflect its importance in determining site location. The variable *0-100 metres from water* might be given a weight of 3, *101-200 metres* a weight of 2, *201-400 metres* a weight of 1, and *401+ metres* a weight of 0. By multiplying the category value by the weight of the variable (W x V), a weighted value is defined for each

Table 1.1. An Example Of Assigning Weights And Values To Categories, Subcategories And Variables

CATEGORY (W)	**SUBCATEGORY**	**VARIABLE**	**VALUE** (V)	**WEIGHTED VALUE** (WxV)
Proximity To Water (W=3)	4-5 Order Water	0-100m	3	9
	3 Order Water	101-250m	2	6
	1-2 Order Water	251+	1	3
Slope (W=2)	Slope	0-5°	3	6
		6°+	1	2
Soils (W=3)	Soils	Dry	3	9
		Mixed	2	6
		Wet	1	3

variable used in the modelling process (Table 1.1). The determination of the numerical weight or value is researcher specific. There must be some basis upon which the researcher makes these numerical assignments. Reference may be made to previous archaeological work which has identified characteristics of the landscape presumed to be associated with archaeological sites. Ethnographic, ethnological, historic or ethnoarchaeological studies may also be sources upon which the basis for weighting of variables is based. The experience of the archaeologist and colleagues also working in the area may also contribute to determining a weighting scheme. Additionally, the nature of the project itself may have some bearing on the weighting applied to variables. For example, a researcher applying a predictive model within a given theoretical framework may give more importance to economically-related variables than some geographic variables. In another instance, a researcher may combine his or her own experience with data obtained from the ethnographic literature and derive weights and values accordingly. In conclusion, the manner in which weights and values are applied is subjective yet it is based upon data obtained and evaluated by the researcher from a variety of sources and applied within project specific frames of reference.

Because the weighted value method allows certain variables to have more ‘predictive strength’ than other variables, it results in a model that better reflects the decisions made by prehistoric people when choosing their activity locations. In addition, because it is imperative that the manner in which the categories and variables are weighted is clearly outlined, the contribution of each variable to the final model is also clearly established. This final point is the most important point of all. For any model to be valid, it must be reproducible and defensible. With the weighting factor of each variable clearly defined, discussions can occur concerning the weights of individual variables and, the effects of changing weights can be tested. The results of these tests can then be evaluated. In the end, one is left with a model for prehistoric activity location which is clearly defined, testable and reproducible.

1.5 METHODOLOGICAL ISSUES

There are a number of methodological issues concerning predictive modelling as it is represented in the literature. The first of these relates to all of the data sources used to develop predictive models. Despite the theoretical framework or the methodological approach used in the development of predictive models, all make use of a limited number of primary variables. These are slope, aspect, elevation and distance to water, vegetation zones, and in some cases, soil characteristics. These variables are ones that are regularly found on site record forms. Additionally, these variables can be traced to three main sources: topographic maps (either paper or digital), existing archaeological data, or aerial photographs. For example, Parker's Sparta Mine predictive model lists fifteen variables used to predict site location. Seven of those fifteen relate to streams, one is soil moisture and another is depth to the water table. Thus nine of the fifteen variables are expressions of a water variable. While the impression is given of a complex interplay of variables, it is simply that more emphasis is being given to water.

Also, while the choice of variables has been limited and the source for these variables even more so, there seems to be an absence of culturally relevant variables. Very few, if any, predictive models incorporate variables derived from Native land use studies, ethnographic data or local informant interviews. While the utility of incorporating these kinds of data has yet to be demonstrated, logically, there is considerable value to incorporating them into predictive models of prehistoric activity locations (Dalla Bona and Larcombe 1992).

A second concern relates to the quality of some primary data sources. For example, Altschul states that the primary source for his three variables is a USGS digital terrain model. Kvamme (1990:114) has evaluated USGS digital terrain data and concludes:

> although there is a general correspondence, in the (purchased USGS data) (1) many small ridges and hills are absent, (2) minor drainages are missing, (3) large features are greatly smoothed, and (4) there is a major error in the form of a 120 ft. high cliff face which would surely make a spectacular waterfall on the Colorado River (which flows down the central valley) if it really existed!

Archaeologists generally consider terraces, hills and small creeks to be extremely important for regional settlement analysis. It is difficult to perform an analytical study using criteria such as terraces and small hills, when the map from which these criteria are drawn does not represent them accurately.

In addition to digital data of questionable quality is the issue of appropriate scale for modelling. For example, let us consider a predictive model developed for a large region at an effective scale of 1:50,000. However, much of the data may be derived from maps published at a scale of 1:125,000 and 1:250,000. The high level of data generalization on these maps relative to the 1:50,000 map forces a significant degradation of the quality of the final model. Additional potential difficulties can develop if some data variables are derived from large scale sources such as 1:15,000 aerial photographs. In this circumstance it may be necessary to reduce the precision and detail of variables derived from large scale sources (air photos) to match the scale of the 1:50,000 predictive model. The implications of variable scales of primary data are addressed more fully in a later chapter.

Thirdly, those developing predictive models tend to present their results as statements of high/medium/low potential areas, or areas of favourability/non-favourability. However, the means by which these terms were defined is seldom clearly expressed. The reader is rarely informed of the means by which determination of categories of potential is made. The cutoff points between high and medium, and medium and low potential is rarely if ever discussed. Clearly, this is an issue that is of importance to cultural resource managers and archaeological researchers alike. The modelling approach developed in this research does not categorize potential, rather it presents a scale of potential where zones of high/medium/low can be determined more clearly and the rationale for that determination is openly presented for further discussion.

1.6 ADVANTAGES OF PREDICTIVE MODELLING

With respect to the kinds of research and applications being made by cultural resource managers, predictive modelling holds considerable promise as a planning tool. A predictive model can offer explicit measures of the likelihood of cultural resources in specific localities. Generating such information can greatly increase the timeframe within which cultural resource managers may plan survey, confirmation and mitigation - well ahead of development activities. This could result in the avoidance of conflicts between land "values", and allow land developers and cultural resource managers to plan land use in fashions that minimize deleterious impacts. In addition, a credible predictive model can help focus archaeological reconnaissance, and direct research to areas within a region that hold some cultural significance. Using such "pointing tools" offers tremendous savings in time and money, and can be highly effective as a means of "stratifying" conventional random sample surveys over large areas. Finally, a good predictive model can be of use to both the cultural resource manager and the archaeologist interested in academic research applications. A predictive model may provide insight into the dynamics of the settlement system within a given region. By explicitly outlining the variables associated with specific types of sites, a predictive model may actually indicate some probable choices made by prehistoric people in developing their land use strategy.

1.7 ON DEVELOPING A PREDICTIVE MODEL

As discussed in the sections above, there are many different approaches to the development of predictive models. These approaches make use of different theoretical and methodological frameworks. The methodology outlined in Volume 4 of this report series crosscuts a number of theoretical and methodological constraints. It can be applied to academic research applications or for 'applied' cultural resource management purposes. It can be used from a deductive or inductive perspective, or employ elements of both. It can be used to develop a numerical and/or graphical model, and it can manipulate variables using the intersection and/or the weighted value approach. In summary, this methodology is relevant to many different archaeological applications with limits imposed only by the creativeness of individual researchers.

As stated in Kohler's definition (1988:33), a predictive model is comprised of a set of testable hypotheses. To arrive at testable hypotheses, a model must be explicit in the variables that are used, and the manner in which those variables are manipulated. This includes clearly identifying and outlining the variables included in the model, the manner in which these variables interact, and any weighting placed upon the variables. Ideally, a flowchart-like diagram outlining the various processes involved in developing the model should be available. Such a diagram would graphically illustrate what variables are used, and how they interact to produce the final result. One of the major stumbling blocks and criticisms of all predictive models is the subjective input of the researcher's own knowledge. All archaeologists acknowledge that this information is important and should not be ignored. However, to be really useful, it should be made clear *what* knowledge is being applied to the development of the model as well as *how* it is being used. The methodology employed here makes that explicit - indeed the researcher is forced to be explicit.

There are a number of assumptions that one works under when developing a predictive model. The first involves the assumption that choices of activity locations made by prehistoric people were influenced by elements of the natural and physical environment. The researcher also assumes these environmental variables have survived, and can be represented by presently available data. These data may be in the form of maps, monographs or may still remain to be collected in the field. The third assumption asserts that correlations between archaeological sites and the natural/physical environment observed by modern researchers reflect land use choices made by prehistoric decision makers. That is, the correlation is assumed to be not due to chance, or reflecting the affect of another, presently undocumented, independent variable. These assumptions may be strengthened or confirmed by repeated testing or application of a model, but the true nature of prehistoric human action can never be fully known.

As predictive models attempt to codify aspects of human behaviour, one cannot expect a model to be simplistic in its makeup, or to be developed in a single effort. The development period of a predictive model is not finite. Altschul calls this a

> ...dynamic modelling approach. Once anomalies...are identified, they become the subject of additional research. As patterns are found, many anomalies become predictable. Those sites whose locations remain anomalous grow in importance (1990:228).

Modelling should be seen and conducted as a dynamic process whereby data collected from any source, at any time, can be incorporated into the modelling process to increase its integrity, accuracy and scope. As such, predictive modelling may be seen as involving three stages: (1) primary stage predictive modelling involving data collection and organization; (2) secondary stage predictive modelling in which an initial model is developed and tested, and; (3) tertiary stage predictive modelling in which the model is subjected to an infinite number of applications and refinements. This process is summarized in Table 1.2.

Table 1.2. Summary Of The Three Stage Predictive Modelling Process

PRIMARY STAGE • hypothesis building - data collection strategies • initial data collection • field reconnaissance, collection of baseline data
SECONDARY STAGE • deductive phase of modelling • association between environmental variables and sites • literature review and integration into model • development or application of initial model • testing of model on previously surveyed areas
TERTIARY STAGE • continued application of model • continued refinement of model • new data continuously incorporated into process • new sites discovered are interpreted and incorporated into existing model

1.7.1 Primary Stage Modelling: Organization and Data Collection

The development of the primary stage predictive model involves three activities: hypthesis building and data collection strategies, initial data collection, and field reconnaissance. Hypothesis building and data collection strategies are the crucial first step. Hyptheses must be generated about the people and activities being modelled. These hypotheses will in large part dictate the important variables to be modelled, the manner in which those variables will contribute to site potential and the data that will be collected.. Initial data collection is conducted within the parameters of hypotheses generated and can be viewed as taking place within both deductive and inductive theoretical frameworks. Existing archaeological site inventories are often a primary source of such information. Predictive models presented in the literature can be reviewed for pertinent data and analytically useful variables. In addition, information gathered from other sources such as ethnographic or land use studies can be evaluated and incorporated into the model. These data are important in developing a theoretical framework in which to interpret the results of the model as well as to guide the data to be collected.

> To have confidence in any models which emerge, we need to know *why* the behavior we predict patterns as it does (Tainter 1983:7).

It is important to note that the researcher must start somewhere and existing data and successful examples of other predictive models offer an acceptable base, subject to a careful evaluation of their relevance and completeness.

The primary stage may be understood as the organizational stage of the modelling process. The researcher must make numerous decisions including:

a) the scale at which modelling will take place;
b) the spatial boundaries within which the model is applicable;
c) the temporal scope of the model, and;
d) the functional scope of the model, i.e. does it apply to all or selected activity types.

Many issues may be predetermined and a function of the project proposal or terms of reference. This is particularly the case with predictive models developed for cultural resource management purposes. It is during the primary stage that an archaeological field survey may be conducted. Usually this involves inductive data collection from portions of the “research universe” that are unrepresented in the existing heritage resource inventory. While it may be that some archaeological information already exists in the form of a site database, it may be subject to a number of biases beyond the control of the researcher. Thus, the collection of new baseline archaeological data provides the researcher with a more complete and representative database with which to build the model. The field program should include as complete an areal survey as possible. The size of the survey area need not be exceedingly large but should represent the study area as a whole. It is also important that a range of environmental characteristics, that are deemed to be the independent variables, be known and mapped within the survey area. The intention of the survey is to understand the distribution, frequency, and component parts of all the sites in the survey area. With the completion of the initial round of data collection and archaeological reconnaissance, primary stage predictive modelling is complete.

1.7.2 Secondary Stage Modelling: Initial Model Development and Testing

A secondary stage predictive model can be said to begin when the requirements of primary stage modelling have been fulfilled. Once this has been achieved, the researcher enters into a deductive phase and can begin to incorporate this data into the second stage of the model. The degree of correlation between the sites discovered during the field survey and the defined environmental variables is measured and ranked. Existing variables derived from the literature can then be evaluated on the basis of the strength of their correlation with this expanded site database. Cultural variables such as plant gathering or species-specific hunting activities can be incorporated into the variables to be modelled.

The researcher may now develop an initial predictive model and test it using the area surveyed in the primary stage. While it may appear that this step is a 'self-fulfilling prophecy", one must be reminded that a variety of data were used to develop the initial predictive model - not solely the data derived from the primary stage survey. Based upon the hypotheses generated earlier, variables can be introduced or removed from the process, or the weighting of the variables can be adjusted until the model is able to predict the highest percentage of sites possible.

A second field survey program in an area near the first is necessary to collect more baseline data and/or test the model. It is recognized however, that this may not always be feasible because of external limitations such as time and money. Once again, the strength of correlation between known site locations and the identified independent variables should be measured. This information should be incorporated into a new, "second generation" predictive model. This model would then be applied to both the primary and secondary stage survey areas. The variables would be modified in such a way as to produce a model predicting the highest percentages of known archaeological sites. Once this has been achieved, tertiary stage modelling may begin.

1.7.3 Tertiary Stage Modelling: Application and Refinement

A tertiary stage predictive model begins when the secondary stage predictive model predicts the location of the highest number of sites possible in the two previous survey areas. It is at this point that the model may be considered applicable in a real sense. At this point, testing procedures have been carried out, and have demonstrated the validity and integrity of the model. The researcher must be vigilant at this point to ensure that the model is not applied blindly. Any continuing application, (or expansion of applicaion) of the model must undergo thorough testing, much like that conducted in the secondary stage to ensure the ongoing validity of the model. New data must be incorporated into the model year after year in an effort to produce the most robust model possible. In addition, as repeated applications of the model are effected, sites that were once 'anomalous' may now become patterned. Such observations must be formalized, carefully interpreted, and then integrated into the model to maintain and update its integrity.

Few, if any, models have achieved tertiary stage development because of the nature of the agencies employing them. Most cultural resource management agencies are limited by their role as resource managers and have neither the time nor resources for ongoing research and development. They are interested in identifying the location of resources in order to facilitate informed planning and management. Given the urgency of these goals, it is often an irresistible temptation to implement partially tested models as a part of routine resource management and planning. As a result, predictive models developed for such agencies often resemble a procedural 'cook book'. The variables identified in the prototype model become tranformed from untested indicators of site distribution into routine "red flags" of site location. As the predictive model is being routinely used, it gains unwarranted credence and uncritical acceptance. Users may reason that if certain steps are followed, a scientifically valid result will follow.

The three stage modelling process outlined here reduces the likelihood that such a 'cook book' approach will result. An additional point may be raised concerning the prospect that a predictive modelling approach will supplant conventional archaeological field work. There should never be a point where predictive models take the place of field work. In the context of an academic modelling exercise, the negative implications of a poor model are relatively minor. However, in the context of cultural resource management, the implications of applying a poorly tested model can be severe, and perhaps even disastrous. At the same time, it is not realistic to expect resource management agencies to do nothing until all possible sites have been field inspected. Clearly a reasonable compromise is possible and best serves all agencies involved.

> I have no objection to the use of multivariate locational models for research and planning purposes, but they simply cannot provide sufficient evidence to warrant the granting of archaeological clearance without the benefit of field survey. Any such reliance on predictive models to 'write off' areas of low projected site density constitutes both an abuse of statistical methods and an abrogation of … management responsibilities (Berry 1984:845-6).

Once again, the development of predictive models is a dynamic process where models are rigorously tested over many years and in many different areas. The results of each year's testing must also be incorporated into the existing model. Information gained in future years of application are also incorporated into the model development process. Ideally, this process should never stop.

1.8 SUMMARY

Predictive modelling is a research methodology used by archaeologists to identify prehistoric activity locations. It has its basis in settlement pattern analysis, and the results of predictive modelling continue to further the aims of settlement pattern studies. Predictive models have been identified as operating within inductive and deductive theoretical frameworks. They are also developed using distinct methodological approaches. Since the mid 1970s, predictive models have become less associated with the settlement studies from which they emerged. For the most part, predictive models have developed within the sphere of cultural resource management, and their application has been primarily in this management context. However, the breadth of application is increasing as archaeologists recognize the potential of predictive models as an academic research tool.

GEOGRAPHIC INFORMATION SYSTEMS AND ARCHAEOLOGICAL RESEARCH

2.1 DEFINITION OF A GEOGRAPHIC INFORMATION SYSTEM (GIS)

This discussion of geographic information systems focusses upon the type of system utlized in this research although general information about other types of GIS systems is also presented.

A geographic information system (GIS) has been defined as "a powerful 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 1986:6). A GIS is a computerized information system that describes objects from the real world in terms of:

> 1)their position with respect to a known coordinate system
> 2)their attributes that are unrelated to position (such as pH, temporal order)
> 3)their spatial interrelations with each other (topological relations) (Burrough 1986:7).

A geographic information system consists primarily of computer software that functions in conjunction with computer hardware to transform and manipulate spatial data. A GIS should be considered as more than a simple data manipulation tool.

> Because these data can be accessed, transformed, and manipulated interactively in a geographical information system, they can serve as a test bed for studying environmental processes or for analysing the results of trends, or for anticipating the possible results of planning decisions. By using the GIS in a similar way that a trainee pilot uses a flight simulator, it is, in principle, *possible for planners and decision-makers to explore a range of possible scenarios and to obtain an idea of the consequences of a course of action before the mistakes have been irrevocably made in the landscape itself* (Burrough 1986:7, emphasis added).

Perhaps the best way to represent the operation of a GIS is to visualize a standard topographic map. A topographic map contains a tremendous amount of information. One may view the various types of information as a range of data categories that encompass roads, railways, water bodies, buildings, vegetation, legal boundaries and elevation among numerous others. While each of these categories is a separate class of information, they are all schematicaly represented and printed on one piece of paper, and therefore are understood as permanently combined in the mind of the map viewer.

A geographic information system may use all of these categories of information, but stores them as separate map layers. The map layer concerned with roads has only 'road' information on it. It may document the location of various roads and also their quality (i.e., highways, secondary, dirt roads). The map layer concerned with water may contain information about the type of water body (i.e., river or lake), its name, volume or stream order. While each of these separate map layers contains only one category of data, they may be manipulated in such a way as to produce a desired result. For example, a map layer of 'roads' and a map layer of 'water' may be combined to determine where all intersections of roads and water exist in a given area. In these manipulations lies the power of a geographic information system. Data may be manipulated in countless ways to model any number of scenarios. For example, it may be necessary to identify all areas within one kilometre of the junction of two streams. To search for and calculate these data manually using paper maps would take a considerable amount of time. Using a GIS, this information could be identified within minutes.

The geographic information system software used in this research is MAP II (Pazner *et. al.* 1989). This software allows for the easy entry, storage, manipulation and analysis of spatial data using Macintosh personal computers. Map II is also notable for its powerful data transformation capabilities. MAP II is a raster-based GIS in that it organizes spatial information in a 'cell' format. Information is stored in cells which are organized in rows and columns. Raster data representation assumes that spatial information is treated as if it were a flat Cartesian surface. Thus, each cell in the electronic database is associated with a square parcel of land in reality. The value assigned to each cell reflects the 'attribute state' of a specific variable which is deemed characteristic of that corresponding square of land.

Data may be entered in a raster-based GIS in a number of ways. Perhaps the most practical method of data entry involves the use of a flatbed scanner. A scanner is a device very similar in its workings to a photocopy machine. A desired image (e.g. a topographic map) is placed on the glass surface of the scanner, which then makes a 'photocopy' of the image. Instead of immediately printing the image on paper, the scanner translates the image into a digital representation: the analog image is automatically transformed into a series of rows and columns with each cell of the image being assigned a different brightness value. When viewed at a large scale, the image appears as a number of large coloured squares. However, when viewed at a small scale, the image appears as a very close representation of the original. The various brightness values can be correlated with different entities in the image. For example, water will be represented by similar brightness values while a grassy field will be represented by a group of different brightness values. Understanding and correlating which brightness values correlate with landscape features allows users to classify the composition of large areas.

Data may also be entered by means of automated data gathering devices such as LANDSAT or SPOT satellites. Both satellites collect data in raster format, rendering them immediately usable by raster-based GIS software. Additionally, data may be entered manually through the use of screen tools available within the MAP II software. Information may also be encoded into the database via a mouse/pen input device resulting in a database that can be directly updated, revised or even created by the user right on the computer screen.

The resolution of raster data describes the relationship between the grid cell size in the database and the actual size of the cell on the ground. For example, a resolution of 30m indicates that every grid cell in the database represents 30 metres by 30 metres on the ground.

There are a number of advantages to using a raster-based GIS. Perhaps the most important one is the simplicity of the data structure. The 'row' and 'column' structure of the data translates into very simple mathematics for computation by the computer. This results in relatively fast analysis and manipulation of data when compared with systems using other data structures (i.e. vector-based systems).

A second important advantage is the compatibility of rastor systems with existing remote sensing technology. Satellite and airborne sensors store the data in raster format. Very little modification is necessary to render that information usable by a raster-based geographic information system.

Because of the nature of the data structure, various kinds of spatial analysis can be easily carried out. Generally speaking, in a raster-based system, every cell is the same size and the maps generated must also be the same size. Therefore, the computer does not need to calculate and account for landscape characteristics that are arranged in irregular or unorthodox shapes. As a result, the mathematics involved in simultaneously manipulating several map layers is relatively simple. Operations such as the overlay of innumerable map layers is facilitated, greatly expanding the usefulness of a raster-based GIS.

A further advantage of a raster-based GIS pertains to inter-application compatibility. Ultimately, all data can be reduced to a text-based, or ASCII, format. This data format is readily accepted by a wide variety of software packages, including analytical statistical packages. This allows the researcher to transfer data from a GIS to a different software package for further, more refined, analyses.

Another advantage of a raster-based GIS relates to the price of the technology. These systems are relatively inexpensive with software and hardware being energetically developed and marketed by several different firms. Concurrent with the development of GIS software, the revolutionary improvement of the power of micro-computers makes them comparable with minicomputers. The linkage of new raster-based GIS with 'high-end' personal computers offers a powerful and versatile set of analytical tools.

Another consideration which makes MAP II an attractive GIS option is that it uses a map manipulation language called cartographic modelling (Tomlin 1990). This 'natural language' approach to GIS greatly simplifies training and use since data transformation functions utilize English-like commands to construct manipulation command lines. Cartographic modelling also forces one to explicitly define what is actually sought through the modelling exercise. The clear delineation of goals greatly simplifies the task of identifying data needed for the modelling process.

Whatever the type of GIS used, such clear-sighted definition of goals and data needs is an important precondition of predictive modelling. Flowcharts can be drawn outlining the various map manipulations to be conducted, and numerous scenarios can be offered and compared: 'what if' analyses may be undertaken, and numerous options to potential difficulties can be identified and addressed. Thus, even before the computer is turned on, the needed data can be identified, the data manipulations that are expected to be performed are indicated, options and alternatives are outlined, and an expectation of the goal to be achieved can be calculated.

There are also a number of disadvantages to a raster-based geographic information system. The first relates to the 'cell' and 'column' structure of the data. This data structure leads to a tremendous amount of redundancy of data and, consequently, data files can consume large amounts of physical disk space. The problem of large file sizes has been alleviated to some degree by clever programming which can reduce the size of raster data files by 50% or more.

A second disadvantage to raster-based systems, again, relates to the cell format of the data. The cell resolution of a map is set by the user. Once cell resolution has been identified, the smallest spatial phenomenon must be represented by that cell resolution. For example, imagine a situation whereby each cell in the digital database has a cell resolution (i.e. is equivalent to) of 100m x 100m on the ground. If we are mapping an area of 50 square kilometres, the resultant data files will consist of a very manageable 5,000 cells of information. However, the smallest spatial phenomenon must be represented as being at least 100m wide or 100m long. Thus, a small stream or creek will be represented as being at least 100 metres wide. Such gross generalization can quickly invalidate the modelling exercise.

If the cell resolution is set very fine (e.g. 5m x 5m), then a very large map is needed to represent a 50 square kilometre map area. Such a map would result in about 2,000,000 cells of information being needed to represent the map area. While topological features will be represented in detail, the map size will be large, computational time will be great, and data storage needs will be huge. A further discussion of this subject is presented below.

Finally, whereas in a vector-based system, phenomenologically recognizable structures can be accurately depicted, in a raster-based system those same structures can only be identified by an aggregation of cells. This ultimately results in a jagged figure where over or under-representation of areas may occur due to the restricted manner in which they can be displayed. This factor is again dependent upon cell resolution.

For more detailed discussion of the use of GIS in archaeological research, a number of studies have recently appeared outlining the advantages and disadvantages of this approach (Altschul 1990; Bailey et. al. 1985; Briur 1988; Brown and Rubin 1982; Calamia 1986; Carmichael 1990; Ferguson 1985; Forney et. al. 1985; Hasenstab 1983; Kvamme 1986, 1989, 1990; Kvamme and Kohler 1988; Limp 1987; Parker 1986; Zubrow 1987; Zulick 1986).

GIS AND ARCHAEOLOGY: PREDICTIVE MODELLING RESEARCH DESIGN CONSIDERATIONS

3.1 INTRODUCTION

The ultimate focus of a heritage resource predictive modelling project should be at the level of the archaeological site. An archaeological site is defined as:

> a specific locality containing the remains of past human activities; such remains may range from the traces of a campfire to the ruins of a settlement occupied for centuries or to the site of an historic shipwreck. A site can also be viewed as a set of information derived from object, features, structures, and human physical remains found in specific geographic, environmental and geological contexts (Department of Communications 1988:29-30).

In order to properly and accurately develop a predictive model of archaeological site location, there are a number of important considerations that must be addressed. These include discussions of data acquisition and integrity, data entry, data manipulation and data resolution.

3.2 DATA ACQUISITION AND INTEGRITY

Anyone doing predictive modelling using GIS will gather data from a number of varied sources. Some potential sources of data are listed in Table 3.1. Data gathered from various sources must be evaluated as to its overall quality and usefulness in predictive modelling. Any of the data sources described below are subject to innumerable biases that affect data quality to varying degrees and, therefore, have the potential of introducing error into the GIS database. These potential sources of error may include (but are not limited to) map scale, age of data, format, accessibility, areal coverage, and density of observations.

3.2.1 Map Scale

The scale at which original maps are drawn directly reflects the quality of data and the amount of detail which may then be obtained from the map. It is extremely important that the scale of source maps match the requirements of the study. For example, a Canada Land Inventory (CLI) map depicting soil series for a given area is published at the relatively small scale of 1:1,000,000. At this scale, the soil differences between eskers and poorly drained muskeg swamps may be generalized into one broad category. At such small scales, distinct lithic outcrops may be generalized into broader categories reflecting their common origin. Furthermore, geological maps tend to focus upon the distribution of economically valued minerals and ore bodies or major bedrock deposits. However, highly silicious bedrock deposits that might be of economic value to hunters utilizing a stone-based technology might be deemed too small or unimportant to be mapped. Thus, a distinction which may prove to be important in a study of heritage resource potential is not available on maps of suitable scale. Ironically, data represented on very large scale maps may contain too much information. That is, the large map scale may result in mapped distinctions which encompass too small an area to be represented at the scale of resolution used in the GIS database. In such a situation, the analyst must selectively ‘generalize’ the data to match the operational map scale.

For the purposes of predictive modelling, it is often necessary to obtain data wherever it might be available, even if at a range of map

Table 3.1. Potential Sources Of Data For Inclusion In A GIS Database

GENERAL DATA SOURCE	EXAMPLES
Public (government) Paper Map Libraries	•NTS Topographic Map Series • Canada Land Inventory Map Series • Ontario Base Map Series
Private (Corporate) Paper Map Libraries	• Forest Industry Map • Ontario Hydro Map Library
Public Electronic Cartographic Databases	• Dept. of Energy, Mines and Resources Digital Map Library • Northwestern Ontario Forest Ecosystem Classification Program
Private Electronic Cartographic Databases	• Proprietary Corporate Digital Cartographic and GIS databases
Electronic Heritage Databases	• Canadian Heritage Inventory Network
Remotely Sensed Data from Air - or Spaceborne devices	• LANDSAT TM and LANDSAT MSS • SPOT • Fluorescence Line Imager (FLI) • Synthetic Aperture Radar (SAR) • Aerial Photography
Published Written Sources	• Manuscripts - e.g. Mercury Series • Journals - Canadian Journal of Archaeology
Unpublished Written Sources	• Consultants' reports to private industry • Consultants' reports to government • Internal government reports • Academic Conference Papers • Workshop Proceedings • Ongoing Regional Research Reports
Field Reconnaissance/Ground Truthing	• Archaeological Survey/Excavation • Geological/Geographical Survey • Wildlife/Vegetation Surveys

scales. In Canada, particularly useful maps are the NTS topographic map series available from the Government of Canada, Department of Energy, Mines and Resources. These maps are routinely published at both 1:50,000 and 1:250,000 scales. Each map contains at least two primary sources of data that can be utilized using a GIS database.

One data source that may be obtained from a 1:50,000 topographic map is elevation. When evaluating the suitability of elevation information in the form of contour lines, one must consider the interval between adjacent contour lines. Clearly, a narrow interval, such as <15m (~50 ft.) provides enough elevation detail to properly represent the general nature of the terrain. Contour intervals greater than 15m result in increasingly severe generalization of the terrain, resulting in the loss of important features such as beach ridges and small moraines. Contour lines in themselves are generalizations from accurate point data. Thus, a contour line does not identify a particular area as having a specific elevation. Rather, contour lines are usually derived from detailed interpretation of stereoscopic pairs of air photos, and identify areas which share average elevations.

A second primary data source that may be obtained from a 1:50,000 topographic map is hydrography. All topographic maps contain information about surface water resources, and include seasonal streams, perennial rivers, seasonal swamps, organic bogs and muskegs, deep water lakes, natural watersheds and artificial reservoirs. These data may be utilized by a GIS to provide map layers representing hierarchies of water bodies, drainage patterns, headwaters/outflow lakes, and so on These data can be organized as polygons or like-groups of grid cells.

One difficulty associated with obtaining hydrographic information from topographic maps relates to seasonal water bodies. It has been the experience of the author that water bodies, designated as seasonal on NTS topographic maps, may range from running streams to dry creek beds. The nature of the watercourse is dependent upon seasonal fluctuations in precipitation as well as yearly fluctuations in precipitation. Therefore, while a topographic map may indicate that a seasonal water body occurs in a given area, it is not guaranteed that there is water at that location. This information could be supplemented by field survey and/or aerial reconnaissance or interpretation of aerial photos.

3.2.2 Age of Data

An important adage in map intepretation is that "…with the exception of geological data, the reliability of data decreases with age" (Mead 1982). Old data may be unsuitable if it was collected under a system of standards that is no longer acceptable. In any assessment exercise, the evaluation of data integrity is a necessary procedure. It becomes imperative when the data used were collected in the past by other agencies for other purposes. In addition, old data may be important when evaluating the historic nature of a landscape. This may include considerations of historic features vs. modern features, pre- vs. post-dam construction impacts, and vegetation changes among a host of others.

3.2.2.1 Geological Data

Some data, such as geological features remain fairly constant over time, although there will be a refinement of data quality and precision as new research is published. A possible variable affecting the quality of geological data is human activity in the form of mineral extraction, transportation/utility corridors, and hydroelectric development with consequent reservoir development. These activities, among others, may cause significant change upon surface or subsurface geological formations relevant to proper heritage resource assessment. Recognition of these factors and their proper identification is an important measure that must be undertaken.

3.2.2.2 Elevation Data

The quality of elevation information tends to remain fairly constant over time, although future refinement and publication of larger scale versions can greatly augment data quality. Factors that may affect the quality of elevation data are major mineral extraction activities and/or natural erosional activities such as earth loss (e.g. river erosion) or movement (e.g. sink holes, mud slides). By reviewing the date of 'culture-check' on the topographic map with any subsequent activities, concerns over the elevation data can be alleviated.

3.2.2.3 Hydrographic Data

The quality of surface water data can remain fairly constant over time. However, long term landscape transformation can come about through deglaciation, isostatic rebound, and short term changes regularly occur as a function of river bank erosion and natural lake level fluctuations. Other factors that may affect surface water distribution are major earth-moving/mineral extraction activities that result in altered watercourses, the creation of new water bodies or the destruction of existing water bodies. Hydroelectric development may also result in significant alteration of water resources. Upstream from a dam, new reservoir lakes may be created, flooding extensive tracts of land. Downstream from a dam, decreased water flows may result in altered water courses associated with slight to significant changes in the topography of the valley bottom. In addition, massive surges of water from a dam can accelerate bank erosion, create large estuarine outflow features at downstream mouths. By cross-checking the date of 'culture-check' on the topographic map with any subsequent activities, concerns over the surface water data should be alleviated.

Subsurface water data tend to change significantly over relatively short periods of time. Factors that affect the level of subsurface water tables are major groundwater extraction activities such as community drinking wells. Also of concern are activities that alter groundwater replenishment areas such as swamps, marshes, and muskeg. Activities which result in drainage of these standing-water localities will ultimately lead to reduced subsurface flows. Such changes may have significant affects upon vegetation communities and surface stream flows. Data which reflect these contemporary altered flows may not represent flows that were extant in prehistoric or early historic times.

3.2.2.4 Vegetation Data

The quality of vegetation data relates to two factors. The first is the nature of data collection. While the forest industry may provide a complete and detailed catalogue of tree species and quality, their catalogue of forest floor plants may be rudimentary, or at least focused upon the distribution of economically valued plants. Such existing data may provide excellent coverage of specific tree species and sometimes plant communities. However, such forest inventories seldom address the structure and diversity of plant communities, nor do they discuss the kinds of plants which might have been important to prehistoric hunters and gatherers. Also, vegetation surveys carried out for various purposes in the past, may not reflect the current structure of the plant community.

As has been discussed in Volumes 1 and 2, the structure of any vegetation community is extremely dynamic, and subject to constant change as a function of a number of causal factors. Small changes in atmospheric and/or environmental factors may result in dramatic changes in plant communities. For example, increased industrialization has an impact upon the natural environment. Increased sulphur emissions has been linked to acidification of precipitation which in turn results in a degradation of forest vegetation and the decline of particularly vulnerable species. Clearly a very serious issue for predictive modelling is whether a given contemporary vegetation community is representative of the vegetation in the past. Additionally, the composition of a given forest tract can only be accurately assessed by 'ground truthing'. That is, determining whether interpretations derived from air photo and satellite data are valid by physically examining representative sample tracts.

3.2.2.5 Archaeological Data

Archaeological site inventory data accumulates over the long term through the efforts of a number of researchers. Such data collection is seldom, if ever, complete, and it can be safely assumed that any site inventory is subject to the collective biases of all contributing researchers. In areas where archaeological reconnaissance is very incomplete, the current distribution of sites is more useful for measuring where archaeologists have worked rather than reflecting the collective inventory of heritage resources. Thus, the spatial distribution of sites is often deceptive, and must be used only after a careful considertaion of the systematic biases at work. Archaeological site inventories must also be reviewed in search of clerical and typographic errors in defintion of map coordiates and physical placement upon maps. Furthermore, a known site may no longer exist if it has been subjected to activities such as earth-moving and erosion.

Archaeological site inventories are constantly being updated. It is important that the latest site location catalogue is used when evaluating the existing archaeological site database. Interpretations about site function also change as a result of scholarly debate and rebutal. Careful evaluations of site functions must be undertaken to properly place a site within its category (e.g. campsite, lithic procurement site, kill site).

3.2.2.6 Transportation Corridors/ Cutting Zones

Because of the ongoing nature of industrial activity northern Canada, the most recent data available must be utilized when identifying transportation/utility corridors. While the position of established transportation and utility infrastructures tend to remain fairly constant, new service roads, transmission lines and gas pipelines are installed every year. For example, established logging roads will provide information about access to relatively isolated areas, but they may not be recorded upon currently published maps. Ironically, logging roads currently recorded upon published maps may be presently out of service and impassible. A proposed transmission line right-of-way may allow identification of areas under consideration for development. The resulting impact to the heritage resource base may then be evaluated accordingly.

3.2.3 Data Format

Three categories of data format may be addressed at this time: maps and air photos available on paper; cartographic data available in a digital format (i.e. existing GIS data) and; non-cartographic data published as prose (i.e. reports, journals, field notes). Each of these three data categories offer strengths and difficulties that need be addressed.

3.2.3.1 Cartographic Data Available on Paper

The use of paper maps for predictive modelling can sometimes be problematical as is apparent from the discussion in the previous section. Some of these problems related to the age of data, issues of map scale, and the high cost of physically updating and republishing new versions of old map sheets. Given the high cost of updating data on paper maps, many maps currently in circulation utilize data that is upwards of 20 to 30 years old. While paper maps may include tremendous amounts of data relevant for use in a GIS-based study, several factors must be considered in order to ensure proper use. Data on paper maps must be transformed into digital format for use by a GIS. Digital transformation of ‘paper’ data may result in the introduction of error through distortion as part of the digitization process, or at the stage of digital data classification.

Paper maps may possess errors due to misinterpretation of the air photos which are the primary information source. Sometimes these errors are not identified as part of the editorial process, and the long time interval between map updates sometimes results in the perpetuation of these errors for decades. Paper is not a static medium, and is subject to stretching and shrinking, resulting in the inappropriate placement of datapoints in the digital database. This error source will be exaggerated if paper reproductions of paper maps are used in map digitization.

Error may be introduced during the scanning stage when a paper map is transformed into raw digital data. Information from a paper map may not be recognized by the computer scanner and may also be missed by the technician.

Error may also be introduced during the classification stage when raw digital data is processed to extract desired information, such as water resources, elevation information or transportation corridors. During classification, the locations of roads, rivers or key elevation marks may be displaced from their actual location by misinterpretation of the data, faulty inputting or misclassification. Further, errors may also occur when the data required are found upon two or more map sheets, requiring each sheet to be scanned seperately and then electronically refitted. Such edge mapping is sometimes difficult to achieve, and requires a great deal of skill and practise.

3.2.3.2 Cartographic Data Available in Digital Format

As discussed above, digital GIS data conforms to two broad formats: vector and raster-based formats. The numerous GIS software packages that are commercially available save data in proprietary formats. In the vector GIS market, a few companies have established a dominant position among GIS users, and as a consequence, the format by which they store their data has become something of an informal standard. Perhaps the most popular of these formats is ARC-INFO.

In the raster GIS market, many different formats exist due to the large number of software packages available. Some data format standards do exist, and include the IGES standard, and the Graphics Kernal System standard. Due to the nature of the raster data structure, information may also be saved in a number of generic formats. These formats include ASCII (American Standard Code for Information Interchange), EBCDIC and binary. Of these, ASCII is capable of being translated by all microcomputers.

The GIS system in use by the Centre for Archaeological Resource Prediction at Lakehead University is MAP II. This GIS software saves its information in a proprietary format. However, it has the capability to import data stored in a number of different formats including SPANS, SYLK, TIFF, PICT, MAP II Interchange Format (ASCII, binary, decimal, hexadecimal [LANDSAT/SPOT]) and USGS DEM Format. MAP II data may also be exported from the program in a variety of formats that include SYLK, Interchange Format, PICT, TIFF and SPANS). This combination of data exporting and importing capabilities makes MAP II a powerful GIS. The data created in MAP II is not limited in utility only with the MAP II program. It can be exported for use with most programs available on the Apple Macintosh as well as numerous programs available for MS-DOS-based computers. Among other potential uses, this allows MAP II generated data to be subjected to analysis by statistical software and included in written reports with a minimum of tedious (and error prone) manual re-entry of data.

A further type of cartographic data available for GIS is available from satellite-based remote sensing devices such as the LANDSAT satellite and the SPOT satellite. Both LANDSAT and SPOT data are commercially available and may be directly imported into MAP II, and once interpreted can provide current information concerning forest cover, water resources, transportation routes and geological data, among others.

3.2.3.3 Non-Cartographic Data Published as Literature

The information concerning the existing heritage resource base is primarily in the form of text-based reports. Very little of that information exists in a cartographic format aside from the spatial placement of sites across the landscape. No cartographic heritage information exists in a digital format which is directly usable by a GIS. Other information that may be necessary to update wildlife habitats, new geological finds or vegetation data also exist in printed literature and therefore may also be of use to a project.

In the case of printed literature as potential data, an understanding of the nature of the data must be effected. Also, the printed literature will, at some point, have to be translated into cartographic form which will result in some generalization of the data being considered.

3.2.4 Data Accessibility

There are two major factors that may affect data accessibility. The first of these relates to proprietary information. In many instances, government and private industry are reluctant to release data for which they have proprietary rights - especially if it is deemed to be sensitive in any way. Obviously, such a lack of availability can compromise the modelling effort. One might be forced to go through the time and expense of re-acquiring the data, or attempting the research without access to important variables.

A second major factor affecting data accessibility is cost. The purchase of proprietary data and/or the purchase of commercial data (i.e. LANDSAT, SPOT) can be costly. For example, LANDSAT TM full scenes are approximately $3900 (US) and SPOT full scenes are approximately $1800 (US). While the cost of purchasing data may initally appear prohibitive, the cost of generating it 'within house' or undertaking modelling without key information makes it clear that such purchase costs may be quite 'affordable'. On the other hand, a considerable amount of data can be obtained at little cost. For example, elevation and most hydrographic data can be obtained from NTS topographic maps at a cost of four to five dollars per map sheet (plus the costs of hiring data entry technicians). If data is to be generated 'in house', then careful selection of a GIS with low training costs is very important.

3.2.5 Areal Coverage of Data

Many large scale land development projects take place in regions removed from heavily populated areas of the country. There is often comparatively little spatial information about these remote and sparsely populated areas, and generally what data is available is quite out of date. Much of the extant data were gathered on a 'project-oriented' basis. That is, data were gathered to provide information about a specific project (i.e. mineral exploration, forest harvesting) in a specific comparatively constrained area such as a river watershed or a highway right of way. With the prominent exception of the federal and provincial government topographic mapping programs, very few data have been gathered on a regional basis. Therefore, if information is to be gathered about large areas (i.e. thousands, or hundreds of thousands of square kilometres), innumerable data sources must be brought together.

Since data coverage is not complete or at a consistent scale for many parts of northern Canada, decisions must be made concerning how the necessary data uniformity is to be achieved. There are two options available. One option is to collect more data. Clearly, this option involves a considerable expenditure of time and money. While this may be the desired option, it may not always be feasible. A second option is to generalize detailed data to the point that it matches the less detailed data. This involves making decisions about which details should be removed from a map in order to match it to a less detailed map. In many cases, this operation can be done by computer which may, for example, "sample" out every fifth cell of information to achieve less detail. While this option is less expensive in terms of time and money, there may be some loss of information and data quality during the various stages of generalization.

Also, since projects will undoubtedly span several years and perhaps longer, decisions must be made towards how to achieve data uniformity into the future. This is vitally important when it is considered that the data might be used by several researchers for multiple purposes.

3.2.6 Density of Observations

One stage in the development of a predictive model involves the study and evaluation of the existing archaeological database. This archaeological information is then utilized to provide information about the spatial distribution of sites, the range of site types reported as well as their temporal distribution.

As is the case with much of Canada, archaeological coverage of northern Ontario is 'spotty'. Much of the research and many of the surveys conducted have been the result of development activities in the form of transportation/utility development (i.e. roads and hydroelectric dams). While some areas of northern Ontario have limited coverage, large portions of the region have received no archaeological coverage.

In the past, the shorelines of rivers and lakes have been the primary focus of archaeological investigation. As noted earlier, this littorial bias is a function of both preconceptions about prehistoric human land use, and the contemporary reality of accessibility and site visibility in northern Ontario (Hamilton 1990:3). Little or no systematic archaeological reconnaisance has ever been conducted in the upland regions removed from these shorelines. This inadequate knowledge base is also a function of the limited number of archaeologists working in the region. This also serves to perpetuate the 'spotty' archaeological coverage, and also the underlying assumptions about prehistoric land use patterns.

Thus the development of predictive models of site location in many parts of northern Canada requires extensive archaeological surveys of test areas as part of the development of preliminary predictive models. Failure to generate this substantive and representative knowledge base may result in the predictive model merely replicating contemporary baises.

3.3 DATA ENTRY

As the current research project exclusively uses the MAP II geographic information system, the following discussion uses this software for illustrative purposes. However, it can also be treated as a generic discussion of how GIS data can be collected and integrated for predictive modelling purposes.

There are three means by which data may be entered into the MAP II GIS: 1) data entry by scanning; 2) entry of existing digital information and; 3) manual entry of data.

3.3.1 Data Entry by Scanning

Data that currently exist in a non-digital format must be transformed into a digital format for use in the MAP II GIS. An effective means of carrying out this transformation is through the use of a flatbed scanner.

A flatbed scanner converts non-digital data into digital data in much the same way a photocopier copies paper documents. It does so by measuring the varying degrees of darkness and lightness on the original map. Many commercially available flatbed scanners have a glass surface approximately 8.5 x 14 inches in size, and convert non-digital images into an 8-bit computer image. An 8-bit image is capable of depicting an image using up to 256 shades of gray.

An example of data entry using a flatbed scanner is as follows. A map is placed on the glass surface of the scanner. The scanner is instructed, via computer software, to take a 'photocopy' of the map surface and the resulting image is sent to a computer as opposed to being printed on a piece of paper. That image is saved in a format compatible with MAP II importing capabilities (TIFF) and is subsequently imported into MAP II.

At this point, the scanned image is merely patterned represntation of the original map using a variety of shades of gray. In order to render the image usable by MAP II, manual classification must be undertaken. For example, if one seeks to isolate and define elevation contour lines, then each contour line must be 'traced' and assigned a legend value. In this case, the legend value will likely corresond to the elevation value, but values can be arbitrarily assigned. In effect, the process of classification involves assigning an unique 'attribute state' name to map cells that are to be treated as the same. In this way, the computer may then treat that map category, in this case a given elevation, as a separate variable, consisting of several attributed states, and isolate it for further manipulation.

Past experience using flatbed scanners to input data into MAP II has proven to be quite effective. A tremendous amount of information can be quickly and accurately entered into the computer. Since one is essentially 'tracing' information, positional accuracy of data is of very high quality. In addition, since a scanned image of a topographic map will also include geopositioning data such as UTM gridlines, placement of additional data is greatly facilitated.

As most scanners have the capability to reduce or enlarge an image, one can arrive at a desired map scale with some ease. Furthermore, if working with several data sources rendered at different scales, map scales can be electronically adjusted to match a data standard. Should the reduction or enlargement capabilities of a scanner not be sufficient, a standard photocopy machine can be used to make a reduction/enlargement of a given area. That photocopy may then be scanned by the scanner and entered into the GIS. Of course, with every photo-reproduction of the original paper map, the degree of distortion of the mapped data increases.

3.3.2 Entry of Existing Data in Digital Format

As discussed above, data that are already stored in digital form exist in a variety of different formats. Some of these formats are immediately compatible with MAP II while others will require translation in order to be imported into MAP II. There are two categories into which digital data may be classed: 1) data that require no intermediate translation and; 2) data that require intermediate translation.

3.3.2.1 Digital data that require no intermediate translation

MAP II has the capability to directly import a number of different data formats without the need for any intermediate data translation. These conform to a variety of standard microcomputer data formats.

One such format is the Tagged Image File Format (TIFF) available on both Macintosh and DOS computers. The value of a TIFF image is that it stores information concerning the various levels of gray apparent in the scanned image. Another standard format is the Interchange Format which includes files saved in ASCII format. This generic file format exists on virtually all computer platforms including DOS and Macintosh. Any type of information can be saved in ASCII format including text and numerics. A third format is the Symbolic Link (SYLK) format used primarily to store the relatively complex information derived from computer spreadsheets. This format also exists on both the DOS and Macintosh platforms. The PICT format exists only on the Macintosh platform and is used primarily to store picture information, both black/white and colour. The importance of the PICT format is that it translates data, drawn as polygons in various software packages, into raster images that are acceptible into MAP II. This is extremely useful if some maps have been drawn using computer aided design (CAD) software. Finally, a translation module allowing the importation of SPANS data created on the SPANS GIS platform has been added to the MAP II package, adding to its flexibility. Thus, data that are created on virtually any microcomputer may potentially be imported directly into MAP II.

In addition, the MAP II Interchange Format allows for direct importing of commercially-available remote sensing data including LANDSAT MSS, LANDSAT TM, and SPOT. These data are stored upon tapes and, therefore, must be downloaded to a microcomputer. This process can often be achieved via a mainframe. An alternate solution is the acquisition of a reel-to-reel tape drive with a microcomputer interface.

The resolution of remote sensing data ranges from low resolution MSS data (80m) to high resolution SPOT data (20m and 10m). These data can greatly aid in the identification of current land use practises, and the spatial arrangement and extent of subtle geographical features in poorly mapped areas. The U.S. Geological Survey (USGS) Digital Elevation format can be imported directly into MAP II. While USGS data may not be necessary for some projects, the potential for its use allows MAP II to be used for projects in the United States as well as other parts of the world.

3.3.2.2 Digital data that require intermediate translation

While MAP II has the capability to directly import a number of different data formats, it cannot currently import formats that are not in raster format. Thus, GIS data created using vector-based systems, such as an ARC-INFO system are not immediately usable by MAP II.

However, Dr. Micha Pazner at the Department of Geography (U. Western Ontario), the head of the Map II development team, has developed translation modules that enable MAP II to translate vector-based GIS data,(such as SPANS), into a format readable by MAP II.

3.3.3 Manual Data Entry

In any project involving the use of a GIS, creating and editing data requires considerable expenditures in time and money. If the processes of data generation and editing can be expedited, the project can realize substantial savings. One of the powerful features of the MAP II geographic information system is the ability to create or edit data directly on the screen using the ‘screen tools’ available within the package. Using these screen tools, spatial information ranging from individual archaeological sites, to road networks, to forest quality classifications can be entered directly on the screen without the need for error prone digitizing tables or data translation procedures.

While manual entry is possible for any of the types of data, it is unrealistic for the large volumes of data associated with geology, forest quality, hydrology, elevation, among other variables to be entered in this manner. In fact, the power of manual data entry comes in adding, deleting, or revising data that were entered via scanners or downloaded from tapes. A newly discovered archaeological site can be added to the database simply by opening the map layer associated with archaeological sites, selecting the appropriate screen tool, moving the tool to the location of the site on the map and then clicking the mouse. The map cell associated with this site is identified and saved into the database. Thus, the necessary tasks of data revision and creation are simplified through the use of screen tools available in MAP II.

3.4 DATA MANIPULATION

The major difference between geographic information systems and other computer-based cartographic systems

> …is the provision of capabilities for transforming the original spatial data in order to be able to answer particular queries…Geographic information systems, however, provide a very much larger range of analysis capabilities that will be able to operate on the topology or spatial aspects of the geographical data, on the non-spatial attributes of these data, or on nonspatial and spatial attributes combined. These analysis capabilities will, in most geographic information systems, be provided in such a way that a user can work interactively in order to perform the analyses and syntheses required (Burrough 1986:81).

MAP II uses of a method of data manipulation called cartographic modelling (Tomlin 1990). A cartographic model can be envisioned as a bound collection of thematic maps, much like an atlas. However, it is a collection of maps that are organized such that each of these maps contain information pertaining to a common study area.

> It is important to note that a cartographic model conveys information about its study area in both implicit and explicit form. ...each of the layers of data within a cartographic model will explicitly describe the nature of each location in its study area in terms of a stated characteristic. ...a great deal of information that is not explicitly recorded (at least not at first) is nonetheless implicit in the spatial and logical relationships among those data that are recorded and in the meanings attributed to them. This implicit information must also be considered part of a cartographic model (Tomlin 1990:4).

This method of manipulating map layers involves what Tomlin defines as ‘map algebra’. Since the average user has little training in computer programming, an approach to spatial analysis that does not require a technical computer background facilitates research by allowing non-GIS technicians to apply GIS to their own research fields. The average computer user will, however, have a good general knowledge of mathematics as well as more specific knowledge on the subject which his/her project is focused. Map algebra takes advantage of this knowledge base by not assuming the user has a background in computer programming. Map algebra enables spatial analysis and cartographic modelling by utilizing natural language commands. The utility of this approach is apparent in the discussion in Volume 4 of this report series.

In MAP II, each map layer has a distinct name. For example, the map layer describing topographic relief might be called ‘Elevation’. A second, different map layer might be called ‘Water’. The various spatial analysis operations in MAP II also have ‘natural’ names. For example, the operation which enables one map layer to be combined with a second is called COMBINE. If one wishes to combine the ‘elevation’ map with the ‘water’ map, the necessary command sequence would be: COMBINE <Elevation> with <Water>.

3.4.1 Advantages to Cartographic Modelling

Cartographic modelling using the ‘natural language’ approach has a number of very important advantages. First, because the user is communicating with the computer using terms similar to everyday language, learning advances at a very quick pace. Novice users can be taught sophisticated cartographic modelling procedures in a matter of weeks. For example, at the Department of Landscape Architecture (University of Manitoba), the author directed a studio project aimed at modelling Manitoba Provincial land use policies. In conjunction with provincial and rural township officials, a graduate studio used MAP II to model thirteen rural land use policies in an effort to evaluate land available for rural residential development. The students in the class were 1st and 2nd year graduate students with no experience using MAP II and little experience using Macintosh computers.

Within one month, all primary data had been entered by the students and data manipulations were undertaken. The Provincial land use policies included development restrictions due to prime agricultural land reserve, wildlife impact, water resource impacts, recreation impacts, industrial impacts, transportation impacts among others. Within two months of the commencement of the project, a presentation was made to provincial and township government officials. In this presentation, graphic and statistical data were used to illustrate areas coming under development constraint as a result of land use policies. In addition, a map was generated illustrating areas where rural residential development could take place and, equally important, areas where development could not take place due to policy restrictions.

In this exercise, students with little or no computer experience and no MAP II software experience effected sophisticated analyses in a very short time period. They manipulated millions of pieces of data addressing sophisticated issues. Clearly, the cartographic modelling approach utilized by the MAP II package greatly facilitated their task. The utility of such ‘user friendly’ data querrying and manipulation by personnel trained in heritage resource management rather than GIS data management is obvious.

A second advantage of the cartographic modelling approach is that the analyst is forced to clearly define the modelling issue to be addressed. For example, a cultural resource manager requires the identification of areas of heritage value. In addition, the identification of areas that have little or no heritage value is also required. The clear delineation of this goal greatly simplifies the task of identifying those areas. Once the goal is known, primary data can be acquired. Flowcharts outlining the various map manipulations required to manipulate the data can be constructed indicating the various procedures necessary to identify high-value heritage areas. Various scenarios can be offered and compared; ‘what if’ analyses may be constructed, and numerous solutions to potential difficulties can be identified and addressed.

Thus, even before the computer is turned on, the necessary data that need to be acquired are identified; the data manipulations that are expected to be performed are indicated; options and alternatives are outlined; an approximation of time expenditures may be plotted and an expectation of goal achievement can be calculated. Such focused planning offers significant savings in time and money.

3.4.2 Disadvantages to Cartographic Modelling

There are two cautions regarding cartographic modelling. One is based on the assumption that all the information contained in the various data layers is error free. To assure that this assumption is valid, all data should be carefully reviewed and corrected. If errors persist in a data layer, they will be perpetuated throughout model construction. Of course, such error checking is necessary no matter what GIS software is used. As each stage of modelling and data transformation progresses, such errors and their cumulative affect become more difficult to identify and purge from the model.

A second difficulty with cartographic modelling is that the person doing the modelling must have an intimate knowledge of the database. That person must also understand how specific data manipulation operations modify the database. Failure to understand the nature of these data manipulations can easily result in the production of derived data layers that may be meaningless. It is incumbent upon the cartographic 'modeller' to not only be cautious with the operations being performed, but also to ascribe the appropriate meaning to the results of those transformations.

Thirdly, it is important that the user understand the subject matter and understand the logic behind the correclations. This is necessary to ensure that that causative interpretations are made of the results, not spurious correlations.

3.5 DATA RESOLUTION

In a project involving the use of geographic information systems, an extremely important consideration is data resolution. As discussed earlier, raster data representation assumes that spatial information is treated as if it were a flat Cartesian surface. Each cell in the database is then associated with a square parcel of land in reality. When entering data into a digital database, one must consider what scale to assign each cell. The size each cell represents in reality is called cell resolution.

Cell resolution is an issue that deserves considerable attention. It not only involves discussion of how representative the database is of reality but also brings into consideration pragmatic issues of file size and computation time. Cell resolution, file size and computation time are all discussed in more detail.

3.5.1 Cell Resolution

Before any data is entered into a database, a decision must be made concerning cell resolution. For example, when it is stated that a map's cell resolution is 25 metres, each cell on the map represents 25m x 25m (625 sq. m.) on the ground.

In a raster database, maps may be assigned a cell resolution of any value. As noted earlier, a cell may be assigned a resolution of 1 km resulting in each cell on the map representing 1 square kilometre. Thus, the smallest feature on the map must be represented as being at least 1km x 1km in size. Small streams will be represented as being 1km wide. Roads and utility corridors will be represented as being at least 1km wide. A single archaeological site is represented as not being smaller than 1 square kilometre.

With a cell resolution of 1km, a map 100 x 100 cells can depict an area of 10,000 square kilometres. Furthermore, an area of that size can be stored on a hard disk using very little space. Because of the relatively small map size, computations involving maps derived from this database will take very little time, and spatial operations in MAP II proceed very rapidly.

Depending upon the needs of some projects, this coarse cell resolution may be quite adequate. However, many topological features will be grossly overrepresented. In the case of a project involving archaeological sites, a cell resolution of 1km is far too coarse - it does not represent 'reality' precisely enough.

Conversely, a map with a cell resolution of 1 metre permits the representation of cultural and natural landscape features at the precision of 1 metre. Ideally, such precision would serve archaeological predictive modelling admirably. Isolated find spots, such as the location of a single projectile point could be represented with great accuracy. Topological features could also be represented with great accuracy, rendering a computer representation of the landscape very credible. Such a cell resolution would be appropriate for research projects that involve the analysis of individual sites and their immediate environs.

However, a map that was 100 x 100 cells in size would only represent an area of 10,000 square metres, or 1 hectare, albeit with great accuracy. However, such a fine data resolution is currently unrealistic for regional predictive modelling encompassing thousands of square kilometres. Using such a cell resolution to map even a few square kilometres would become cumbersome to perform spatial analysis and transformation. A project would become figuratively mired in a computational swamp. Even the simplest operations would take hours; the more complicated might take days.

In sum, cell resolution is not machine dependent, but rather, is project dependent. The needs and goals of a project will guide the selection of an appropriate cell resolution. Most cultural resource managers focus their concerns at the level of the archaeological sit, and in northern Canada an archaeological site can range from one stone chip encompassing only a few centimetres to as large as several hectares, encompassing multiple occupations. While there is no such thing as average archaeological site size, sites discovered to date tend to average smaller than 50m x 50m in size.

Therefore, a cell resolution of 50m could appropriately represent many archaeological sites in northern Canada, although some sites will be overrepresented. However, 50 metres might represent only the width of a substantial river, and while some highways may approach a width of 50 metres, most roads in northern Ontario are considerably smaller.

A cell resolution of 30 metres offers a reasonable compromise between over-generalization and unwieldy file size and data redundancy. Unavoidably, some data such as small creeks will be generalized to a small degree.

A 30 metre cell resolution is an attractive option for archaeological projects. Using a resolution of 30m x 30m, a map 300 x 300 cells in size could represent an area of 81 square kilometres. This is a substantial area to be represented electronically given contemporary computer capabilities. Furthermore, 30 metre cell resolution is the same resolution in which LANDSAT TM data is obtained. While some geometric rectification is necessary to correct the TM data to map sheet datum references, once corrected the satellite data is immediately compatible with project maps. This possibility provides exciting possibilities. For example, the LANDSAT TM data could be overlaid with GIS produced data to compare logged areas with areas of predicted high heritage value.

3.5.2 Map Size and File Size

The issue of file size should not be confused with map size. While they are related, they are in fact two different subjects. Map size refers to the number of rows and columns making up the actual map and, consequently, the area covered by the map. File size refers to the amount of storage space the map takes up on the computer hard disk.

In MAP II, the absolute maximum number of rows in any map is 64,000 and the absolute maximum number of columns in any map is also 64,000 (Pazner et. al. 1989:4). If one were to construct a map 64,000 x 64,000 cells in size with a cell resolution of 30m, the total area represented by the map would be 1920 km x 1920 km or 3,686,400 square kilometres.

However, while the possibility exists to represent a huge area using one map, such a map is presently unreasonably large to manipulate. One solution is to divide the large map into several smaller maps that together represent the large region. Again, the decision concerning map size is essentially project dependent, and must be evaluated on a case by case basis.

File size relates to the physical storage space a map file requires on a computer disk. A map that is 600 x 400 cells in size with one legend entry takes up 285k on a disk. When it is considered that an 80 Mb hard disk can store up to 80,000k of data, 280 files at 285k each can be stored on one hard drive. It is unrealistic to assume that all files will be 285k in size. An average file size of 400k is more realistic for comparison purposes. Thus, a 80 Mb hard drive could store up to 200 map files. The size of files already generated by the Centre for Archaeological Resource Prediction range between 1100k and 284k. The majority of maps are less than 400k in size.

The number of legend entries depends upon the variety of data being displayed on one map. MAP II maps can contain a maximum of 32,000 different legend entries (Pazner et. al. 1989:4). Indeed, a legend of this size will result in a file that is several megabytes in size. Perhaps the map layer that is the most demanding in terms of legend entries is the elevation map, which may contain several thousand different elevation categories. In general, most maps will not exceed 25-30 legend entries while some may reach 100.

3.6 SUMMARY

The integration of geographic information systems into archaeological research has considerable benefits to archaeologists. First, a GIS allows a researcher to consider areas much larger than could be done using more traditional, manual cartographic and database methods. Secondly, and specifically with respect to predictive modelling, analysis of regions can be done uniformly over the entire area.

The introduction of computerized techniques results in the consideration of many factors that are not traditionally addressed by archaeologists. These issues include age and type of data, differing digital formats, means of entering, and exporting and presenting data.

Overall, the theoretical basis upon which predictive modelling rests and the practical considerations integral in the development and application of a predictive model are important in understanding, applying and interpreting the prototype predictive model presented in Volume 4 of this report series.

REFERENCES CITED

Acheson, Steve and D. French
1992 *Plotting Archaeologists or Plotting Sites: An Evaluation of Survey Objectives and Techniques in Selected Areas of B.C.*. Unpublished paper presented at the 25th annual Canadian Archaeological Association meetings, London, Ontario.

Adams, Robert McC. and Hans J. Nissen
1972 *The Uruk Countryside: The Natural Setting of Urban Societies*. University of Chicago Press, Chicago.

Allen, Kathleen M.S.
1990 Manipulating Space: a commentary on GIS. In *Interpreting Space: GIS and archaeology*, edited by K.M. Allen, S.W. Green and E.B.W. Zubrow, pp. 197-200. Taylor & Francis Ltd., London.

Allen, K.M., S.W. Green and E.B.W. Zubrow
1990 *Interpreting Space: GIS and archaeology*. Taylor & Francis Ltd., London.

Altschul, Jeffrey H.
1989 *Man, Models and Management: An Overview of the Archaeology of the Arizona Strip and the Management of its Cultural Resources*. Report prepared for USDA Forest Service, USDI Bureau of Land Management, Denver.

1990 Red flag models: the use of modelling in management contexts. In *Interpreting Space: GIS and archaeology*, edited by K.M. Allen, S.W. Green and E.B.W. Zubrow, pp. 226-238. Taylor & Francis Ltd., London.

Bailey, R., S. Hackenberger, D. Howes, and D. Wherry
1985 *Geographic information processing in land use modeling and testing in the Columbia River basin*. Paper presented at the 50th annual Society for American Archaeology meeting, Denver.

Berry, Michael S.
1984 Sampling and Predictive Modeling on Federal Lands in the West. *American Antiquity* 49(4):842-853.

Briuer, F.L.
1988 *Geographic information systems and automated archaeological information: Investments and dividends for archaeological resource management*. Paper presented at the 53rd annual Society for American Archaeology meeting, Phoenix.

Brown, Margaret Kimball (ed.)
1981 *Predictive Models in Illinois Archaeology Report Summaries*. Illinois Department of Conservation, Division of Historic Sites.

Brown P.E. and B.H. Rubin
1982 Patterns of desert resource use: An integrated approach to settlement analysis. In *Granite Reef: A study in desert archaeology*, edited by P.E. Brown and C.L. Stone, pp.267-305. Anthropological Research Papers No. 28, Arizona State University.

Brown, P.E. and C.L. Stone (eds.)
1982 *Granite Reef: A study in desert archaeology*. Anthropological Research Papers No. 28, Arizona State University.

Burrough, Peter A.
1986 *Principles of Geographical Information Systems for Land Resources Assessment.* Clarendon Press, Oxford.

Calamia, M.A.
1986 *Geographic information system applications for cultural resource management*. Bureau of Land Management, Denver.

Carmichael, D.L.
1990 GIS predictive modelling of prehistoric site distributions in central Montana. In *Interpreting Space: GIS and archaeology*, edited by K.M. Allen, S.W. Green, and E.B.W. Zubrow, pp. 216-225. Taylor & Francis Ltd., London.

Carr, Christopher
1985 Introductory Remarks on Regional Analysis. In *For Concordance in Archaeological Analysis*, edited by C. Carr, pp. 114-127. Westport Publishers, St. Louis, Mo..

Chang, K.C. (ed.)
1968 *Settlement Archaeology*. National Press Books, Palo Alto.

Dalla Bona, L. and Linda Larcombe
1992 *Evaluating Archaeological Potential.* Report submitted to the Quebec Ministry of Cultural Affairs, Quebec City.

1993 *Modelling Prehistoric Land Use in Northern Ontario*. Paper presented at the Centre for Archaeological Investigation's 10th annual Visiting Scholar's Conference, Southern Illinois University, Carbondale.

Dean, J.S.
1983 Environmental aspects of modeling. In Theory and modeling: refining survey strategies for locating prehistoric heritage resources. *Cultural Resources Document 3*, edited by L.S. Cordell and D.F. Green, pp. 11-27. Southwestern Region: Forest Service, USDI.

Dept. of Communications
1988 *Federal Archaeological Heritage Protection and Management: A Discussion Paper*. Government of Canada, Ottawa.

Ebert, James I. and Timothy Kohler
1988 The Theoretical and Methodological Basis of Archaeological Predictive Modeling. In *Quantifying the present and predicting the past: Theory, method, and application of archaeological predictive modeling*, edited by W.J. Judge and L. Sebastian, U.S. Government Printing Office, Washington, D.C..

Ferguson, T.A.
1985 *Use of GIS to recognize patterns of prehistoric cultural adaptation*. Paper presented at the 50th annual Society of American Archaeology meeting, Denver.

Flannery, Kent V.
1968 Archaeological Systems Theory and Early Mesoamerica. In *Anthropological Archaeology in the Americas*, edited by B.J. Meggars, pp. 67-87. Washington, D.C..

Flannery, Kent V. (ed.)
1976 *The Early Mesoamerican Village*. Academic Press, New York.

Forney, S.J., K.A. Schneider, R.A. Mead, and C.M. Robinson
1988 *Applications of a geographic information system for cultural resources on the Ocala National Forest*. Paper presented at the 53rd annual Society for American Archaeology meeting, Phoenix.

Green, Ernestene L.
1973 Location analysis of prehistoric Maya sites in Northern British Honduras. *American Antiquity* 38(3):279-293.

Haggett, Peter et al.
1965 *Locational Analysis in Human Geography*. St. Martin's Press, London.

Hamilton, Scott
1990 *Heritage Resource Predictive Modelling in Northern Ontario*. Unpublished paper presented at the Canadian Archaeological meetings, Whitehorse, Yukon.

Hasenstab, R.J.
1983 *The application of geographic information systems to the analysis of archaeological site distribution*. Paper presented at the 48th annual Society for American Archaeology meeting, Pittsburgh.

Haury, Emil W.
1956 Speculations on Prehistoric Land Settlement Patterns in the Southwest. In *Prehistoric Settlement Patterns in the New World*, edited by G. Willey, pp.3-10. Viking Fund Publications in Anthropology No. 23, New York.

Hay, C., et. al.
1982 Archaeological predictive models: a New Hanover test case. *North Carolina Archaeological Council Publication 18*.

Heizer, Robert F. and Martin A Baumhoff
1956 California Settlement Patterns. In *Prehistoric Settlement Patterns in the New World*, edited by G. Willey, pp. 32-44. Viking Fund Publications in Anthropology No. 23, New York.

Jochim, Michael A.
1976 *Hunter-Gatherer Subsistence and Settlement: A Predictive Model*. Academic Press, New York.

Judge, W. J. and L. Sebastien (eds.)
1988 *Quantifying the present and predicting the past: Theory, method, and application of archaeological predictive modeling*. U.S. Government Printing Office, Washington, D.C..

Kohler, Timothy A.
1988 Predictive Locational Modeling: History and Current Practice. In *Quantifying the present and predicting the past: Theory, method, and application of archaeological predictive modeling*, edited by W.J. Judge and L. Sebastian, U.S. Government Printing Office, Washinton, D.C..

Kohler, Timothy A. and Sandra C. Parker
1986 Predictive Models for Archaeological Resource Location. In *Advances in Archaeological Method and Theory Vol. 9*, pp. 397-452. Academic Press, New York.

Kvamme, Kenneth L.
1985 Determining Empirical Relationships Between the Natural Environment and Prehistoric Site Locations: A Hunter-Gatherer Example. In *For Concordance in Archaeological Analysis*, edited by C. Carr, pp. 208-238. Westport Publishers, St. Louis, Mo..

1986 The use of geographic information systems for modeling archaeological site distributions. In *Geographic information systems in government Vol. 1*, edited by B.K. Opitz, A. Deepak Publishing, Hampton, Virginia.

1988a Using Existing Data for Model Building. In *Quantifying the present and predicting the past: Theory, method, and application of archaeological predictive modeling*, edited by W.J. Judge and L. Sebastian, U.S. Government Printing Office, Washington, D.C..

1988b Development and Testing of Quantitative Models. In *Quantifying the present and predicting the past: Theory, method, and application of archaeological predictive modeling*, U.S. Government Printing Office, Washingtion, D.C..

1989 Geographic Information Systems in Regional Archaeological Research and Data Management. In *Method and Theory in Archaeology Vol.1*, edited by M. Schiffer, University of Arizona Press, Tucson.

1990 One-Sample Tests in Regional Archaeological Analysis: New Possibilities Through Computer Technology. *American Antiquity* 55(2):367-381.

1992 A Predictive Site Location Model on the High Plains: An Example with an Independent Test. *Plains Anthropologist* 37(138):19-40.

Kvamme, K.L., and T.A. Kohler
1988 Geographic information systems: Technical aids for data collection, analysis, and display. In *Quantifying the present and predicting the past: Theory, method, and application of archaeological predictive modeling*, edited by W.J. Judge and L. Sebastian, U.S. Government Printing Office, Washington, D.C..

Lafferty, R.H.
1981 Distribution of Archaeological Materials. In *Settlement Prediction in Sparta*, edited by Lafferty and R.H., pp. 163-206. Arkansas Archaeological Research Series No. 14, University of Arkansas, Fayetteville.

Lewis, R. Barry and Susan A. Murphy
1981 Central Illinois Unit (IV). In *Predictive Models in Illinois Archaeology Report Summaries*, edited by M.K. Brown, pp. 33-40. Illinois Dept. of Conservation, Division of Historic Sites.

Limp, W.F.
1987 *The identification of archaeological site patterning through integration of remote sensing, geographic information systems, and explanatory data analysis*. Ms. on file at the Arkansas Archaeological Survey, Fayetteville.

Limp, W. Frederick and Christopher Carr
1985 The Analysis of Decision Making: Alternative Applications in Archaeology. In *For Concordance in Archaeological Analysis*, edited by C. Carr, pp. 128-172. Westport Publishers, St. Louis, Mo..

Mead, D.A.
1982 Assessing data quality in geographic information systems. In *Remote sensing for resource managers*, edited by by C.J. Johannsen and J.L. Sanders, pp. 51-62. Soil Conservation Society of America, Chapter 5, Ankeny, IA..

Parker, Sandra C.
1985 Predictive Modeling of Site Settlement Systems using Multivariate Logistics. In *For Concordance in Archaeological Analysis*, edited by C. Carr, pp. 173-207. Westport Publishers, St. Louis, Mo..

1986 The role of geographic information systems in cultural resource management. In *Geographic information systems in government Vol. 1*, edited by B.K. Opitz, pp. 133-140. A. Deepak Publishing, Hampton, Virginia.

Pazner, M, K.C. Kirby, N. Thies
1989 *Map II: Map Processor*. John Wiley & Sons, Ltd., New York.

Peregrin, Peter
1988 Geographic Information Systems in Archaeological Research: Prospects and Problems. *GIS/LIS '88 Proceedings Vol. 2*, pp. 873-879.

Plog, Fred and James N. Hill
1971 Explaining variability in the distribution of sites. In *The Distribution of Prehistoric Population Aggregates*. Anthropological Reports No. 1, pp. 7-36, Prescott College Press, Prescott, Az..

Reynolds, Robert, G.D.
1976 Linear Settlement Systems on the Upper Grijalva River: the application of a Markovian Model. In *The Early Mesoamerican Village*, edited by K. Flannery, pp. 180-194. Academic Press, New York.

Roper, Donna C.
1981 Review Comments on Five Predictive Models Reports Units IV, V, VIII, IX, and X. In *Predictive Models in Illinois Archaeology Report Summaries*, edited by M.K. Brown, pp. 147-151. Illinois Department of Conservation, Division of Historic Sites.

Schermer and Tiffany
1985 Environmental Variables as Factors in Site Location: An Example from the Upper Midwest. *Midcontinental Journal of Archaeology* 10(2):215-240.

Steward, Julian H.
1938 *Basin-plateau aboriginal sociopolitical groups*. Bureau of American Ethnology Bulletin 150. Smithsonian Institution, Washington.

Sullivan, Alan P., and Michael B. Schiffer
1978 A Critical Examination of SARG. In *Investigations of the Southwestern Anthropological Research Group: The Proceeding of the 1976 Conference*, edited by R. Euler and G. Gumerman, pp. 168-175. Museum of Northern Arizona, Flagstaff.

Tainter, Joseph A.
1983 Foragers in the Southwest. In *Theory and Model Building: Refining Survey Strategies for Locating Prehistoric Heritage Resources: Trial Formulations for Southwestern Forests*, edited by L.S. Cordell and D.F. Green, pp. 28-38. Cultural Resources Document No. 3, USDA Forest Service.

Thomas, D.H.
1986 *Refiguring Anthropology: The First Principle of Probability and Statistics*. Waveland Press, Prospect Heights, Illinois.

Tomlin, C. Dana
1990 *Geographic Information Systems and Cartographic Modeling*. Prentice Hall, Englewood Cliffs, N.J..

Trigger, Bruce D.
1968 The Determinants of Settlement Patterns. In *Settlement Archaeology*, pp. 53-78. National Press Books, Palo Alto.

Vita-Finzi, C., and E.S. Higgs
1970 Prehistoric Economy in the Mount Carmel Area of Palestine: Site Catchment Analysis. In *Proceedings of the Prehistoric Society*, edited by J. Coles. Vol. XXXVI, December 1970:1-37.

Warren, Robert E.
1990 Predictive modelling of archaeological site location: a case study in the Midwest. In *Interpreting Space: GIS and archaeology*, edited by Allen, Green and Zubrow, pp. 201-215. Taylor & Francis Ltd., London.

Williams, Leonard, D. H. Thomas, and Robert Bettinger
1973 Notions to Numbers: Great Basin Settlements as Polythetic Sets. In *Research and Theory in Current Archaeology*, edited C.L. Redman, pp. 215-237. John Wiley & Sons, New York.

Williams, Stephen
1956 Settlement Patterns in the Lower Mississippi Valley. In *Prehistoric Settlement Patterns in the New World*, edited by G. Willey, pp. 52-62. Viking Fund Publications in Anthropology No. 23, New York.

Willey, Gordon R.
1953 *Prehistoric Settlement Patterns in the Virù Valley, Peru*. Bureau of American Ethnology Bulletin 155. Washington, D.C..

Willey, Gordon R. (ed.)
1956 *Prehistoric Settlement Patterns in the New World*. Viking Fund Publications in Anthropology No. 23, New York.

Willey, Gordon R. et al.
1965 *Prehistoric Maya Settlement in the Belize Valley*. Papers of the Peabody Museum, Vol. 54, Cambridge, Mass..

Wood, John J.
1978 Optimal Location in Settlement Space: A Model for Describing Location Strategies. *American Antiquity* 43(2):258-270.

Wright, G.A. and J.D. Dirks
1983 Myth as environmental message. *Ethnos* 48(3-4):160-176.

Zubrow, E.B.W.
1987 The application of computer-aided GIS to archaeological problems. In *Proceedings of the first Latin American conference on computers in geography, San Jose, Costa Rica*, pp. 647-676. Editorial Universidad Estatal a Distancia, Costa Rica.

Zulick, C.A.
1986 Application of a geographic information system to the Bureau of Land Management's resource management planning process. In *Geographic information systems in government Vol. 1*, edited by B.K. Opitz, pp. 309-328. A. Deepak Publishing, Hampton, Virginia.