Causality and Cross-Purposes in Archaeological Predictive Modeling

Thomas G. Whitley
Brockington and Associates, Inc., USA

See the CD for the extended version

Abstract

In recent years numerous archaeological approaches to predictive modeling have been presented in the literature. Most of these have taken the "inductive" perspective of applying known site locations to an analysis that estimates probable site location based on a mathematical equation and presents predictive surfaces in a GIS. Conversely, "deductive" models have also been used in which "expert systems" or site selection variables have been quantified as probability surfaces. There has been little discussion, though, of the differences between CRM and academic-based predictive modeling and how it has influenced the state of the "science" today. Generating more refined inductive predictive models either through the use of higher quality site location data or through more complex statistical techniques, runs counter to the implicit goals of CRM-based predictive modeling. A simple deductive GIS approach which assumes a causal explanatory relationship creates comparable or better results (especially in homogenous areas) with no negative effects on these limited goals. Ultimately, the dichotomy between inductive and deductive approaches is not in theoretical orientation, rather it is embodied in our understanding (or failure to understand) that predictive modeling is really a tool useful only for land management, not interpretive archaeology.

The primary use of predictive models to date is almost invariably for large scale Cultural Resource Management (CRM) applications, and they typically occur in North America. The second primary purpose for which predictive models have been developed is that of understanding past land use or what can be called site selection processes (Whitley 2000). It was recognized early on that predictive models provided a quantitative aspect to what was well understood as the qualitative realm of settlement pattern analysis (e.g. Allen et al. 1990). Clearly, predictive modeling is seen as a means by which we might address some of the complex issues of human/landscape interaction, typically in a GIS framework.

Though Bayesian statistics is not invoked in all predictive models, the underlying structure of Bayesian probability is assumed for all of them. The first fully articulated assumption of all archaeological predictive modeling is a form of the Bayesian rule of "total probability" (Pearl 2000:3). This assumption can be stated as: The probability of any land unit being a site (or a portion of a site) is the sum of the probabilities of all exhaustive and mutually exclusive variables that cause a land unit to be chosen as a site or to unintentionally be made one. Put simply, this means that the presence of a site in any area is a factor of all possible influencing variables (intentional or unintentional). This invokes a direct causal relationship between archaeological sites and a host of possible variables; the underlying principle of all predictive models. This is further elaborated by the "conditional rule of Bayesian probability" (Pearl 2000:3-4). This can be stated in archaeological terms as: The probability of any land unit being a site is the sum of the probabilities of each causal variable multiplied by the conditional probability of that variable. Put simply again, this means that not every factor is as influential as every other factor; that they each respond to other conditions, or in effect provide an individual weight to the final probability of any land unit being part of an archaeological site. The diversity in influence of different variables allows predictive modelers to statistically assess, or make expert decisions about, what they find to be the most influential factors and build their models accordingly.

So, what are the theoretical implications of the foregoing characterizations? The most obvious is that the site must be a valid and useful concept to explain human activity. Also, that sites are definitive and absolute in their ability to be identified, and that the characteristics used to identify sites are strictly correlative with human behavior, and not other phenomena. That correlation is assumed to mean that the most common or abundant behavior(s) are the most significant with respect to the purposes of predictive modeling, and that the success of predictive models is a factor of their ability to identify the most common site-selection behaviors.

Are these implications necessarily true, though, and what is not implied by these assumptions, but may be true? First of all, sites may not be easily distinguished by empirical evidence, and they may not be considered strictly representative of human activity (Dunnell 1992). Different activities result in different kinds of sites, and different kinds of sites may have many different kinds of manifestations. Strictly speaking, significance of archaeological sites cannot be equated with abundance, and models geared toward identifying the most common elements of site selection may, in fact, overlook very significant sites in supposedly low potential zones. This becomes manifested in the distinction between so-called inductive and deductive models.

Interestingly, archaeologists employ the terms inductive and deductive in ways that are quite limited with respect to their meaning in the realm of scientific explanation. In philosophy of science, the terms inductive and deductive refer to the difference between logical arguments based on universal laws and statistical tendencies. Deductive explanations are based on mathematically consistent and provable laws, while inductive ones are based on observing statistical trends (Hempel 1965, Salmon 1998). In archaeological predictive modeling, however, inductive and deductive do not refer to methods of explanation, rather to practical means by which probability values are calculated. In the end, all predictive models are trend-based, or inductively explanatory; even so-called deductive ones. Instead, the distinction should be made between models as being either correlative or cognitive in nature.

Correlative models are those which use existing site data and currently measurable environmental variables to build statistical relationships which can then be generalized from previously surveyed areas to those which have not been surveyed (such as through regression analysis). As such, correlative models are strictly empirical by nature and make the assumptions of determinism. Deterministic explanation comes in several forms (Hempel 1965 Salmon 1998) but in general the Bayesian rule of conditional probability, in a deterministic framework, would embrace the idea that the inaccuracies of any predictive model are the result of not having enough information. All probabilities could be identified, in principle, if we only had the ability to do so.

Cognitive models are those which are not limited by existing archaeological data. Hypotheses of site placement are built on understanding more complex issues involved in the cognitive selection of suitable areas. Then, presumed important variables are measured and classified in a way hypothetically similar to how prehistoric populations may have done so, and probability surfaces are projected across the entire landscape. A single cognitive model may, in fact, produce many different permutations which can be tested. Once complete, the known dataset of archaeological sites is then compared to the projected probability zones and accuracy and precision estimates are made. Cognitive models are not necessarily deterministic, though they can be. But they do have the potential to embrace indeterminism, since they are not built from correlative evaluations.

A good way to understand the difference between deterministic and indeterministic explanation can be illustrated by this example (adapted from Mackie 1974:40-41 and Salmon 1998:145-147). Imagine three candy machines. The first is purely deterministic; if you put in a Euro, a candy bar is always ejected. No other coin or object will cause the candy bar to be released, and no candy bar will ever be inadvertently released without the insertion of a Euro. We know that the function of inserting a Euro is both necessary and sufficient to cause a candy bar to be ejected. We can therefore deductively explain the presence of a candy bar with the deterministic rule that a Euro must have been inserted.

A second machine, though, is somewhat different. With it the insertion of a Euro will always produce a candy bar, but sometimes insertion of other items will have the same effect. Thus, a Euro is a sufficient cause of the presence of a candy bar, but it is not the only possible (or necessary) cause. The presence of a candy bar, therefore cannot be deductively explained by a rule stating a Euro must have been inserted, rather it becomes an issue of inductive probability. There is a certain likelihood that a Euro may have caused the appearance of the candy bar (which can be calculated by assessing the state of the system over a period of time or a set number of observations). The machine is still deterministic, though, because we assume that if complete information were available, we could always explain the presence of a candy bar.

A third machine differs in that inserting a Euro or other items triggers an instantaneous analysis of the spin of a quantum particle trapped in the machine. If the spin is in one direction it will result in the release of a candy bar, but in the other direction it will not. Similarly, sometimes with no insertion of any coins, the particle's spin will be measured and a candy bar will be ejected. Here, the insertion of a Euro is neither sufficient nor necessary to explain the presence of a candy bar. We do not have any explicable deterministic way of addressing the presence of a candy bar from this machine. We could assume that more information is needed, but clearly even when we have a suspected causal factor, it does not always result in the release of a candy bar (hence prediction is impossible). In order to fully understand the nature of the machine we need to address all of its constituent parts and the mechanistic effects of all possible causal factors, even if some of those may be reliant on fundamentally indeterministic variables (such as quantum mechanics or human free will).

When dealing with archaeological sites, we are faced with a similar circumstance. Correlative models assume that all variables are either present and have been measured, or that they could be if we only had the information. Ironically, this is termed deductive chauvinism, (Salmon 1998:142-163) and it implies that given all variables and all parameters, all probabilities could be determined and prediction would be 100 % accurate and 100 % precise. Correlative modelers, though, do not deny the influence of human behavior and cognition in making site placement decisions. Yet, the acceptance of this perspective assumes that all human cognition is a deterministic system and implies an absence of free will (Salmon 1998:28). Thus, correlative models are only capable of identifying necessary factors to produce archaeological sites, and only those which are both frequently necessary and commonly observable. Correlative models are not capable of providing insight into the sufficiency of site placement factors to explain the presence of a site, nor the mechanisms of how site selection processes are determined. This is in direct opposition to the second purpose of predictive modeling.

Cognitive models assume, at least, the limitations of the second machine, but allow the possibility of the third (where some aspects of site placement decisions may be, in principle, inexplicable). This assumes that some aspects of human systems are dynamical. Dynamical systems do not imply the absence of predictability for all aspects of the system, merely that they range on a scale between entirely predictable to entirely unpredictable. A correlative analysis, though, is by its very nature limited to only those aspects which are highly predictable, and cognitive explanation must focus on causal-mechanistic issues instead (Salmon 1998). This requires the adoption of explanatory frameworks which deal with indeterministic phenomena.

So, this brings us to the point at which we might ask; why do so many applications of predictive modeling tend to be strictly correlative? The first reason is probably one of convenience. It is relatively easy to take an existing dataset of archaeological sites and environmental values, perform some standard statistical analyses (such as multiple nonlinear regression), and produce a handy formula. The formula can then be turned around and applied in a GIS and a fairly accurate and precise model sometimes results. Second, correlative models are often presumed to have some level of objectivity. With respect to predictive models, I think it is often assumed that more complex statistics means greater objectivity. Such methods have their uses, but the old maxim of "bad data in, bad data out" applies regardless of the methods of mathematical manipulation in between. This was illustrated quite succinctly by Cowgill's warnings against placing too much emphasis on mathematical models at the expense of theoretical profundity (Cowgill 1986:387).

But, ultimately, why is it that correlative models seem to work, at least on occasion? The main reason a correlative model does achieve success is based on what might be called the lowest common denominators for site selection. These are several variables that correlate highly with archaeological sites simply because they are limiting factors on all human behavior; primarily slope and distance to water. Invariably, every successful correlative predictive model uses slope and distance to water, in some form, as key factors in developing correlative formulas, and they almost all occur in semi-arid or highly dissected areas. Causality, though, has not been addressed in such correlative models. It is assumed that because of the correlation between known archaeological sites and particular variables, that they were actively used as suitability indicators for sites. In reality, though, I argue that most environmentally correlative variables (especially slope and distance to water) act primarily as auto-correlations and were probably rarely cognized as variables of choice. In that sense, they may be considered necessary factors for site selection, but their importance is auto-correlative with all human activity not causally conditioned by it. Thus, they cannot be seen as sufficient cause for site placement.

So the upshot of this discussion is; how do correlative models fail to meet the goals of predictive modeling? Several things would be required for successful application of a correlative model. First, the project would need to be located in a region that is sufficiently arid and dissected enough that slope and distance to water are meaningful limiting factors, or suitable alternative limiting factors can be identified. Second, it would require a large dataset of accurate and well described archaeological sites from all temporal periods of similar function, and which do not cross multiple cultural boundaries. Third, it would require an additional dataset which could be used to test the accuracy and precision of the model, or alternately a large enough initial dataset of sites that a "jack-knife" sample can be held back and used for testing. Fourth, it would require a well developed environmental dataset from which measurable variables can be extracted.

The problem is that there are few locations within which all of these criteria are met. Almost invariably the project areas do not have such ideal conditions or large and accurate datasets, and those conditions can only be met by vast amounts of data gathering at a very high cost. It therefore becomes too costly or even impossible to do a correlative predictive model in many cases, and ultimately the resulting model does not provide better insight into site placement processes than intuition. As the cost of doing the predictive model increases, the reasons it was initiated become increasingly irrelevant from a land management perspective. Since slope and distance to water are the primary important limiting factors in most correlative models, they contribute the greatest percentage to a model's success. Any variables beyond the primary ones add proportionally less to the gain statistic and therefore are of smaller and smaller consequence. Conversely, though, the means of extracting the influence of those additional variables adds increasingly to the cost of the model.

Ultimately, this begs the question of; how should we create and apply quantitative archaeological predictive models? What purposes are appropriate, and who should pursue them? I argue that we need to recognize that correlative predictive models (regardless of their methods and area of application) have severe theoretical and explanatory limitations. They can be used in some situations to give insight into land management activities; specifically alternatives analysis. But they should not be mistaken as tools of interpretative archaeology. Likewise, it would be inappropriate to consider correlative models as a means to protect significant archaeological sites or high potential areas, as they would be limited to only abundant kinds of sites and a few, probably auto-correlative, environmental factors. Without causality, predictive modeling may be a useful land management tool in some severely limited settings and with nearly unlimited funding, but it provides no explanatory power and forces a deterministic and facile understanding of human cognition.

References

ALLEN, K., GREEN, S. and ZUBROW, E., 1990. Interpreting Space: GIS and Archaeology. Taylor and Francis, New York.

COWGILL, G., 1986. Archaeological Applications of Mathematical and Formal Methods. American Archaeology, Past and Future; A Celebration of the Society for American Archaeology 1935-1985. Meltzer, D., Fowler, D. and Sabloff, J. (eds.), Smithsonian Institution Press, Washington DC.

DUNNELL, R., 1992. The Notion Site. Space, Time, and Archaeological Landscapes. Rossignol, J. and Wandsnider, L. (eds.), Plenum, New York.

HEMPEL, C., 1965. Aspects of Scientific Explanation. Free Press, New York.

MACKIE, J., 1974. The Cement of the Universe. Clarendon Press, Oxford.

PEARL, J., 2000. Causality: Models, Reasoning and Inference. Cambridge University Press, Cambridge, UK.

SALMON, W., 1998. Causality and Explanation. Oxford University Press, Oxford.

WHITLEY, T., 2000. Dynamical Systems Modeling in Archaeology: A GIS Approach to Site Selection Processes in the Greater Yellowstone Region. PhD Dissertation, Anthropology, University of Pittsburgh, Pittsburgh, PA.