An Explanatory Framework for Predictive Modeling Using an Example from Marion, Horry, Dillon, and Marlboro Counties, South Carolina

Thomas G. Whitley and Gwendolyn Burns

Brockington and Associates, Inc.
Atlanta, GA, and Charleston, SC, USA
tomwhitley@brockington.org

Abstract

It has been argued at CAA and other conferences that archaeological predictive models that explain the relationships between the environment and human activity, rather than merely identifying presumed correlations, have the greatest potential to inform land management decision-making. An explanatory approach to archaeological predictive modeling within a GIS framework was designed and used for a large-scale highway development project in South Carolina. Covering more than 6,500 sq. km in the Coastal Plain, this model was an ideal test for some of our notions about the nature of human settlement, procurement, and interactive behaviors. The results suggest that an explanatory approach is more enlightening, flexible, efficient, effective, and ultimately more useful than any other approach for this largely homogenous region. They also indicate that the approach could be employed anywhere, can be used to establish regional and/or local baselines, and is adaptive to the needs of a particular project or question.

1 Introduction

Last year, Brockington and Associates, Inc., developed an archaeological predictive model as part of the alternatives analysis for the proposed I-73 highway corridor under a contract with the South Carolina Department of Transportation (SCDOT) and the Federal Highway Administration (FHWA). The model was designed to assess the relative areas of lowest to highest archaeological sensitivity within Dillon, Horry, Marion, and Marlboro Counties, South Carolina, as a means to evaluate survey costs and choose the most cost-effective alternative (Figure 1). Through the GIS/statistical analysis of a series of measured and derived environmental and cultural attributes, we developed a set of sensitivity surfaces modeling archaeological probability for different settlement/subsistence patterns and behavioral categories. These surfaces were then combined into an overall land management model expressed as a continuous sensitivity value for every 30-m land unit in the study area.

Our goal in this paper is not just to provide a brief overview of the methods by which this model was accomplished, but also to illustrate some of the key concepts that were employed relating to both archaeological predictive modeling specifically, and to cognitive landscape studies in general. In fact, we argue here that predictive modeling *per se* is probably the least interesting application of this kind of modeling for archaeologists. Predictive models have immediate uses within the large scale of alternative analyses, but there is a distinct price to pay in reducing the rich tapestry of prehistoric and historic cognitive spatial evaluation to probabilities useful for modern-day land management purposes (see especially Church et al. 2000; Wheatley and Gillings 2002; van Leusen et al. 2002; Whitley 2003). As archaeologists, our interests tend to be in the experiences of modeling and what that tells us (or fails to tell us) about past people, rather than in the land management issues themselves.

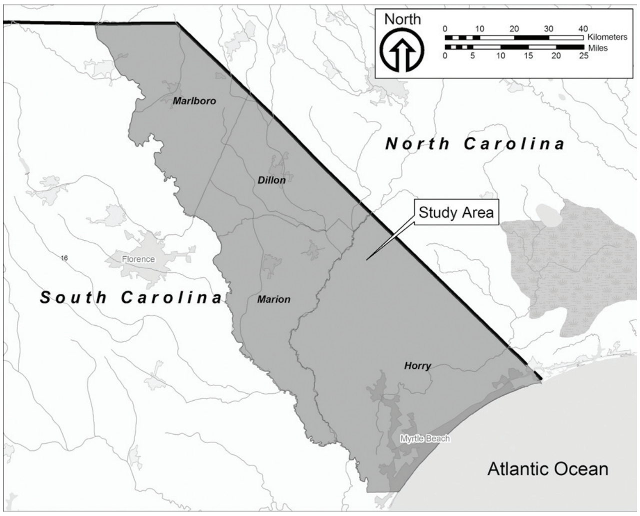

Figure 1. Project location.

Note, that we are not referencing the vast literature on predictive modeling as we are attempting to narrow the discussion rather than provide yet another overview of the practice (for additional discussion and references on predictive modeling please refer to any of the sources cited herein and follow their bibliographic trail). Similarly, there are a number of sources which address the prehistory or history of South Carolina in general, and/or this region in particular which were used in the development of this model. Because of limited space, we are excluding them, and the general overview of regional prehistory that they provide, from this discussion.

The causal-explanatory framework from which we developed this model is, in fact, useful for much more detailed analyses of economic and spatial motivation, agency and social interaction, than the mere identification of site locations. The real benefit of the cognitive GIS evaluation of archaeological landscapes is in the potential for a real understanding of how people in the past exploited, lived in, and experienced all aspects of their environment. Far more than just a tool for predicting survey and mitigation costs, an explanatory framework within which GIS models can operate provides a visual canvas upon which we can illustrate some of our more complex notions about human behavior.

2 The Study Area

The Coastal Plain of South Carolina is a broad expanse of largely homogenous terrain that extends from the low rolling hills in the north to wide marshy flatlands and barrier islands in the south. Unlike many semi-arid and highly dissected areas of the country, permanent water sources are virtually ubiquitous and ground slope tends to be relatively flat. Thus, neither "slope" nor "distance to water" are significant factors in limiting human settlement.

Human occupation first occurred in the region possibly as early as 12,000 to 14,000 years ago. At that time the environment was colder than today, and boreal forest and grasslands covered the study area. In essence, conditions were probably much like parts of Eastern Canada, with temperate swamps in the south and a mixed deciduous/coniferous forest covering much of the rest. The coastal zone of complex river and tidal channels includes salt and brackish marsh, and the Pleistocene coastline was farther out to sea than today. As a result, many early sites are probably located underwater on the Continental Shelf.

Through the Holocene, the climate warmed and became more moist. Oak-hickory forest began to dominate in the interior, while longleaf pine and live oak forest became abundant in the coastal zone. Cypress and sweet gum became the climax marshland species. Isolated grasslands were now less common, and the coastline retreated nearly 100 km due to rising sea levels. Today, the region has warm summers and mild winters, with a fairly high annual rainfall. Torrential thunderstorms are common, and hurricanes are particularly prevalent between June and September, though strong hurricanes striking the Southern Atlantic Coast of the US (outside Florida) occur, on average, only once per decade.

3 The Cultural Framework

Our first step in the modeling process was to determine which time periods and which behaviors we were interested in. Once again, the purpose of the model was not to establish any different survey or treatment strategies for high, moderate, or low potential zones, but to assess the relative costs of survey, testing, and excavation of significant sites within all planned highway alternatives; but maintaining all of the current South Carolina archaeological standards. This means that regardless of how our ultimate model evolved, we would still survey all areas of the chosen alternative equally and according to the standards established and already used in the state. The model was meant merely as a tool to determine how expensive it would be to survey any given alternative, and how likely it would be that significant sites which needed to be avoided or excavated could fall within it.

Therefore, we decided that we would need to be most concerned about the time periods and behaviors likely to produce significant sites (i.e., all of them). This is a point of departure that should explicitly be noted because it goes to the fundamental flexibility of the explanatory framework. Correlative models assume that all sites used in the analysis represent the same behaviors, or at least that spatial decision making is uniform enough to support the notion that a common pattern of variable selection will result in a useful model; despite the many different cultural groups or time periods these sites represent. An explanatory model, on the other hand, assumes from the outset that archaeological sites are the byproduct of hundreds of kinds of behaviors, by many different people, each with their own cognitive landscapes functioning within distinct cultural traditions (Whitley 2005). Thus, the focus should be put on developing individual models for each behavior and using the sites as a testing dataset, rather than combining all of the sites and finding the lowest common denominator of all behaviors and all time periods.

It is possible as an archaeologist to have an interest in distinct, perhaps extremely unique or complex behaviors, and to build explanatory models for them in exactly the same way in which we approach this model. There is, in fact, no limitation to model even behaviors which leave no archaeological trace *because the sites themselves are not used as a means to build the model*. Any behavioral model can be tested with even a small dataset of sites, or possibly through logical abstraction from the absence of observed sites. Ultimately, an explanatory model of human settlement and spatial patterning in a region could include hundreds, or even thousands, of distinct spatial models all reflecting individual behaviors, of individual cultural traditions, or unique cognitive landscapes.

Because of our mandate to create a model useful for land management purposes, our goal was to find some middle ground between a generic lowest common denominator type model (where all sites and behaviors are lumped together) and the extreme opposite perspective (where every specific behavior is modeled on its own). The compromise was reached by classifying the human occupation in the region into several larger categories representing what we believed were the behaviors or practices that produced sites most likely to be considered significant by the South Carolina Department of Archives and History (SCDAH). In this way, our model evolved largely in the context of the land management practices for which it was required (i.e., the purposes of the SCDOT and the FHWA), but constrained by the past and current standards and practices of the state review agency (i.e., the SCDAH).

To meet this end, we defined five categories of settlement/subsistence, based principally on the mobility and resource extraction methods defined by archaeologists working in the region. There is a great deal of temporal overlap in the transitions between these patterns, and it should not be assumed that they are mutually exclusive. But the following larger categories were defined:

- *Prey-based Nomadism (PBN)* - This category represents the earliest period of occupation in the region, where nomadic hunters followed migrating groups of large prey, typically in grassland areas, but also in woodlands. It is described as "prey-based" because mobility patterns were largely dictated by the movements of the prey and not tied to spatial boundaries or territories. This period coincides predominantly with the Early Paleoindian period (ca. 14,000 to 9,000 years ago).
- *Wide-area Ecosystemic Nomadism (WAEN)* - This category is defined by the transition from large migratory prey to smaller locally abundant resources. Though still nomadic, hunter-gatherers from this time frame exploited large-scale ecosystems for a much greater diversity of species than the earlier time periods. Population pressure was still low, however, and there may have been a great deal of regional migration between very different habitats during different times of the year; especially with regard to accessing lithic resources. Temporally, this settlement/subsistence pattern equates with the Late Paleoindian and Early Archaic periods (10,000 to 6,000 years ago).
- *Constrained Ecosystemic Nomadism (CEN)* - This category correlates to a time when population pressure began to create a change in the dynamics of settlement and subsistence in the region. Though still nomadic and still hunting and gathering locally available resources, territories were evolving and becoming constrained. There was likely much greater trade of utilitarian goods, and overwater travel probably became more dominant than in previous periods. In all, this pattern represents the period in which trade and social networks were evolving. The cultural designations associated with this pattern were the Middle and Late Archaic periods (7,000 to 4,000 years ago).
- *Seasonal Sedentism (SS)* - This category represents the period in which previously nomadic people began to re-occupy the same localities year after year in a seasonal round and for longer periods, thus becoming somewhat sedentary. They probably transitioned between larger ecosystems such as the Piedmont and the Coastal Plain, in a fairly regular and predictable pattern. Traveling primarily along waterways (both overland and overwater), they would have maintained proximity to established trade routes and exploited very predictable local resources in regular and familiar ways. Horticulture became a primary food production method during this time, further tying people to the larger more fertile river valleys. The temporal periods associated with this pattern are the Early and Middle Woodland periods (4,000 to 2,000 years ago).
- *Permanent Sedentism (PS)* - This category represents the transition to full scale agriculture and the establishment of permanent villages. Tied very closely to established trade routes, especially within the largest river valleys, political relationships evolved through access to exotic prestige goods. Very complex relationships, territories, and resource exploitation patterns evolved into the Mississippian societies that came to dominate the Southeast. The most significant sites of this time are the mound centers, located often at the fall line of major river valleys. This settlement/subsistence pattern also includes the later Euro-American occupations through today. Temporally, the Late Woodland, the Early and Late Mississippian, and Historic periods (3,000 years ago to today) can all be subsumed by this pattern.

With this framework of settlement/subsistence patterns as a guide, we defined three primary categories of behaviors. Much like the patterns defined above, these should not be assumed to be entirely mutually exclusive, but could easily be shown to have a lot of overlap:

- Resource Acquisition (RA) - This includes the hunting or gathering of food species as well as horticultural or farming behaviors, the exploitation of lithic or other resources, and accessing exotic or non-utilitarian items (typically for trade).
- Domestic/Production (D/P) - These behaviors included the establishment of settlements, building dwellings or storage structures, production activities such as lithic tool manufacture, making and firing ceramics, weaving textiles and processing utilitarian resources, etc. They also include cooking food, disposing of refuse, and other domestic activities, as well as manufacturing exotic trade items.
- Social Interaction (SI) - This includes many kinds of social behaviors, such as interaction with neighbors for trade or political purposes, building of mounds or ceremonial centers, maintaining territorial boundaries, warfare and other ritual activities.

By intersecting these behavioral categories with the defined settlement/subsistence patterns, we correlated them with a series of archaeological occurrences defined and recorded by previous researchers in the region. Furthermore, we were able to extract from these sites the types of sites which tend to be considered significant most frequently (by the SCDAH) and limit our investigations to the patterns and behaviors that represent them, instead of everything else. Once those kinds of sites were defined, we developed the following key explanatory understandings of the relationships between behaviors and the landscape:

Resource Acquisition is keyed to source locations, habitat, predictability, and local access routes.

- Domestic/Production is keyed to comfort, resource exploitation patterns, pre-existing sites, both regional and local travel corridors, and a compromise between resource and social needs.
- Social Interaction is keyed to regional travel and trade routes, territorial boundaries and markers, military engagements and incidents, and regional social relationships.

Additionally, there is a transition from low to high intensity of exploitation of local resources for the earliest periods through the Late Archaic, then a decline again to somewhat lower intensity during the periods when agriculture is most relied upon. This includes both upland and wetland resources. There is as well a transition from predominantly overland to overwater travel; reaching its peak during the Mississippian.

A key to understanding our approach was to consider that we could further define three primary levels of spatial knowledge with respect to all behaviors. These were based on the mobility patterns defined earlier. But in short, we would expect that all residents in a region would have spatial knowledge of local resources in direct proportion to the amount of time spent in proximity to them. Thus, highly nomadic groups that travel through a region but do not significantly exploit local resources would have only limited knowledge of those resources, and they would be expected to stay close to regional travel corridors. More constrained nomadic groups who intensively exploit local resources on a seasonal basis would have more substantial resource familiarity, and thus greater spatial knowledge, and would be expected to travel further from regional travel corridors, but perhaps only in limited circumstances. Permanent residents of a region (whether they are nomadic or sedentary) would be expected to have the most familiarity with regional resources and should be expected to be least tied to regional travel corridors.

However, that relationship to regional travel corridors could also reflect the dependence on trade relationships. Thus, permanent residents who were socially and politically tied to extra-regional trade would still be tightly bound to regional travel corridors despite their extensive knowledge of regional resources. Similarly, the methods of travel and the direction of travel would largely determine the specific travel corridors taken. Given these cultural parameters, we were able to develop hypotheses for probabilistic formulas once we gathered the available spatial data.

4 The Spatial Data

For this model, we gathered a series of available spatial data. Most prominent was the 30 m digital elevation model (DEM) extracted from the National Elevation Dataset (NED; http://seamless.usgs.gov/). From this data we were able to extract, standardize, and create a great deal of our derived data (including slope, modeled hydrology, flow accumulation, flow direction, and surface aspect). Additional datasets which form the basis of the model or were used to test portions of it include (but are not limited to): the digitized counties (from the supplied ArcGIS data); the detailed digitized soils for each county (http://www.ncgc.nrcs.usda.gov/products/datasets/ssurgo/); modern roads, highways, wetlands, and waterways (all available from the ArcGIS dataset or directly from several federal agencies such as the Environmental Protection Agency (EPA) http://www.epa.gov); historic roadways (hand digitized from hard copy historical maps); and previously recorded archaeological sites (from the State Sitefiles Database—digitized by the South Carolina Institute of Anthropology and Archaeology). Since the study area was large (covering nearly 2500 square mi; 6500 square km) we did not incorporate aerial photographs as a primary dataset.

The primary dataset was converted to a series of surfaces useful for the model. All final data was transformed into 30 m raster surfaces for analysis. This was due to the elevation data being originally in that format, but was also fortuitous since South Carolina uses a standard archaeological survey technique of 30 m shovel testing. This means that our final probability surface could be directly translated into predicted numbers of positive shovel tests without any intermediate mathematical transformations.

Degree of slope was the initial surface derived using the standard ArcGIS slope extraction routine. The resulting surface was then standardized so that the minimum slope (0 degrees - or level terrain) was equivalent to a value of 0 and the maximum (about 32 degrees in the study area) to a value of 1. This allowed the later transformation of slope into different friction surfaces by multiplying the standardized value by weighted travel costs.

The DEM was also used to extract a hydrology surface. Because we were interested in understanding the travel costs across the surface with water as both a travel medium (i.e., via watercraft) and a travel barrier (opposing foot travel where streams and rivers are too deep), we needed to generate a hydrology surface that could reflect some general understanding of flow rates (not part of the existing digitized stream data). The flow rates would help us extract locations of possible stream crossings. Thus, we used the built-in ArcGIS hydrologic analysis as a means to extract flow accumulation. Granted, there are drawbacks with this method (e.g., the assumption of uniform rainfall patterns, and the inability to incorporate stream widths), but in a more detailed study we would expect this to be ironed out. For our purposes it was sufficient. The final flow accumulation surface was then transformed into another standardized variable ranging from 0 (all flow rates lass than 200—meaning fewer than 200 30-m land units drain into them) to 1 (the highest flow rate and where the Pee Dee River leaves Horry County), which could be weighted and added later as considered appropriate.

Aside from the transformation of all other digital data into standardized decimal values ranging from 0 to 1, the remaining primary derived surfaces were extracted from the soils data. The published soils data for each county include a table generalizing the capacity classes; in other words, the soil's potential for several different kinds of habitat. Rather than key site locations to a named soil type, we chose to assign numerical values to each soil type based on their categorical descriptions in these tables. For example, every soil type had a value of *Excellent, Very Good, Good, Neutral, Poor, Very Poor,* or *None* for 10 different categories, and the potential for Wetland Wildlife, Woodland Wildlife, Openland Wildlife, Shallow Water, Wetlands, Coniferous Forest, Hardwood Forest, Wild Herbaceous Plants, Grasslands, and Grain Crops. These were translated into decimal categories ranging from -1 to 1, with neutral as zero.

From this categorical transformation, ten individual raster surfaces were produced reflecting the soil potentials of each category for the entire study area; the assumption being that prehistoric people would be seeking out the benefits of soils by such categories individually, not by a soil type *per se* (which is an averaging of all of these categories). These were taken further to create habitat potentials, such as for megafauna (combining the potentials for grasslands, hardwoods, and openland wildlife), upland wildlife (combining woodlands, openlands, hardwoods, and wild herbaceous plant potentials), general wildlife (averaging all wildlife potentials), agriculture (combining grain crop potential and potential for grasslands), and horticulture (equal parts grain crops, grasslands, and wild herbaceous plants).

5 Regional Travel Corridor Extraction

Once all of the primary and derived data sources were standardized and rasterized for analysis, it was possible to extract another series of surfaces. In a previous paper (Whitley 2004), it was argued that all human settlement is time-series dependent and should not be seen as comparable to, or tested by its relationship to, random point placement. This is a fundamental flaw of most predictive models, and several assumptions were made here to provide a different approach. First, it was assumed that travel corridors would form the baseline from which probability could be assessed. These travel corridors were assumed to represent two types of travel: local (primarily for resource acquisition) and regional (principally for trade or migration). Furthermore, we assumed two methods of travel: overland travel (either by foot or horseback) and overwater travel (by watercraft). Assuming travel as a baseline allowed us to incorporate a time-series dependency that more accurately represents human behavior than equally assessing all areas of the spatial manifold.

To derive travel corridors we used two surface analyses. First, we were interested in regional travel corridors through the region; this would allow us to better understand the mechanisms by which trade routes arose and by which people migrated in and out of the study area. To do this, we first had to create a friction surface. Friction is generally assumed to be a relative cost to travel across one land unit in the analysis. In this case our land units were 30 m across, and we set the baseline travel cost across a completely frictionless 30-m land unit as a cost value of 1. This baseline cost was then altered by addition of whatever we believed altered the friction. In this case, slope is one friction variable. We added the standardized slope value transformed into a relative cost. In this case, we believed that slope is an exponential cost, and that as slope increases, the added travel friction increases exponentially; thus a slope of 0 adds a cost of 0 to the friction surface, while if a slope of 10 degrees adds a cost of 1, a slope of 40 degrees would add a cost of 16 (not 4 as would be the case in a straight linear transformation).

Water is a friction modifier in two ways. First, in overland travel, streams and large rivers create barriers to movement. Thus, the flow accumulation surface was used as a friction additive where the more water that flowed through the land unit, the greater the friction cost. This was used in a straight linear sense and was not transformed exponentially. Second, in overwater travel, streams and large rivers provide a medium to movement. Thus, when calculating overwater travel routes it was assumed that some sort of watercraft was used and friction was reduced for all waterways over a certain size (i.e., they had to be navigable).

From these primary surfaces (slope, flow accumulation, and digitized waterways) we created two final friction surfaces, overland travel friction and overwater travel friction. We would like to have included paleovegetation patterns as part of the friction surfaces since they could also have been significant barriers to movement, however those data are much sketchier for the region and were deemed not really pertinent for the general predictive model we were creating.

To make use of the friction surfaces, we created a series of points spaced evenly at a 1 km distance around the entire perimeter of the study area (in this case all four counties). We used the built-in ArcGIS routine for calculating the least-cost pathways between these points to simulate travel from any direction outside of the study area. We ran this analysis for both overland and overwater travel and produced a map indicating the least-cost pathways through the region (Figure 2). Of course, we can't assume that overland or overwater travel only used the single least-cost pathways, but our use of closely spaced points all along the perimeter of the study area maximized our potential for extracting a large number of travel corridors. Additionally, the digitized historic roadways were added as known regional travel corridors during the historic period.

6 Local Travel Corridor Extraction

Once we had extracted what we believe to be the regional travel corridors, we were also interested in how people exploited local resources. To do this we took a slightly different tack. We were not interested in the single least-cost pathway into the local resource areas (i.e., the uplands or wetlands depending on the resource), rather we wished to extract all pathways into the local resource areas. So, we first took the regional travel corridors as a starting point and created two pseudo-elevation models by calculating cost distance from the regional travel corridors to all points in the spatial manifold using each of the friction surfaces.

This essentially gave us the ability to view a false terrain as an expression of how much relative effort would be required to travel in any direction starting at the nearest point along a regional travel corridor. This false topography could then be used, exactly as a digital elevation model would be, in a hydrological analysis. The result would display the most efficient pathways away from the regional travel corridors to exploit local resource areas. These we called local travel corridors, and formed the basis for formulas of those periods showing intensive exploitation of local resources.

One additional point has to be expressed with respect to the local travel corridors. We believe that depending upon the settlement/subsistence pattern employed, there are mitigating factors in how the friction surface could be used. So far, we employed two different friction surfaces (overland and overwater friction) representing two different ways of using water as either a travel medium or travel barrier. However, friction could be further moderated by subsistence choices and expressed by the soil capacity classes.

For example, Early Paleoindian people would have been intently following herds of large herbivores. Those herbivores would be more likely found in areas of grassland rather than wetlands or conifer forest, and consequently people pursuing them are likely also to be found in such areas. Therefore, we could further moderate the friction surfaces to reduce friction in land units that contain good megafauna habitat, and increase it in land units with poor megafauna habitat for prey-based nomads. This provides a way to decrease cost distances from regional or local pathways if megafauna habitat is high. The same potential exists for each of the other five settlement/subsistence patterns. This approach was used to create several different overland local travel corridor analyses (overwater travel was not affected because soils values were uniform in all waterways).

This sounds complicated, but the essence is that any cost distance evaluation is moderated by the friction surface, and friction can incorporate many different kinds of costs and benefits, including social variables. We did not include territorial boundaries in the analysis because not much is known or presumed about our study area regarding such social boundaries. But we see no reason why they could not be included in other studies as moderators of friction, or even more complicated social variables such as knowledge or predictability. There is really no limit to what could be imagined, and friction/cost distance evaluation holds the key as a proxy for all sorts of cognitive hypotheses.

7 The Formulas

With all of our standardized primary and derived datasets in hand, we were ready to begin the development of probabilistic formulas. From the outset, we understood that in an ideal situation we would have an unlimited amount of time and resources to dedicate to building hundreds of formulas that expressed our hypotheses for the settlement/subsistence patterns defined along with the behavioral categories we were interested in; or, conversely, to focus on very specific detailed behaviors. Bearing in mind, however, our goal of a generalized model useful for a comparative alternatives analysis, we decided to limit our building of formulas to three to six examples for each pattern/behavioral combination (Figure 3). All formulas were weighted-additives of each standardized variable such that the output always ranged as decimal values between -1 and 1 (with -1 representing very low potential and 1 representing very high potential).

The decimal range of -1 to 1 is an expression of the underlying framework of the model as a cost-benefit analysis. Rather than a typical probability range of 0 to 1 (or 0 to 100%), using -1 to 0 indicates an accumulated cost for the pattern/behavioral combination (with 0 being neutral or no cost, and -1 being the highest observed cost) while the range of 0 to 1 denotes an accumulated benefit (with 0 once again being neutral or no benefit, and 1 being the highest benefit observed). This allows a much better explanatory understanding of the nature of cognitive landscapes and how they were employed to make cultural decisions.

Ultimately, we created 46 formulas (Figure 4) that were applied to the spatial manifold and compared with both the previously recorded archaeological data and the spatial distribution of probability values. Our goal was to select the best formula for each type of site and each settlement/subsistence pattern based on a comparison with the known data.

To do this, we first converted all known sites that we believed represented the settlement/subsistence pattern and behavioral category of each formula into a raster surface of 30-m pixels valued either 0 (no site present) or 1 (site present). Then the resulting raster surfaces were multiplied by the formula surface, producing zeros everywhere no site had been previously recorded, and the formula value everywhere an applicable site was known. A histogram was then created from that data illustrating the distribution of applicable sites with respect to the formula values. Any formulas where applicable sites occurred in low or very low predicted values were then thrown out. The formula which

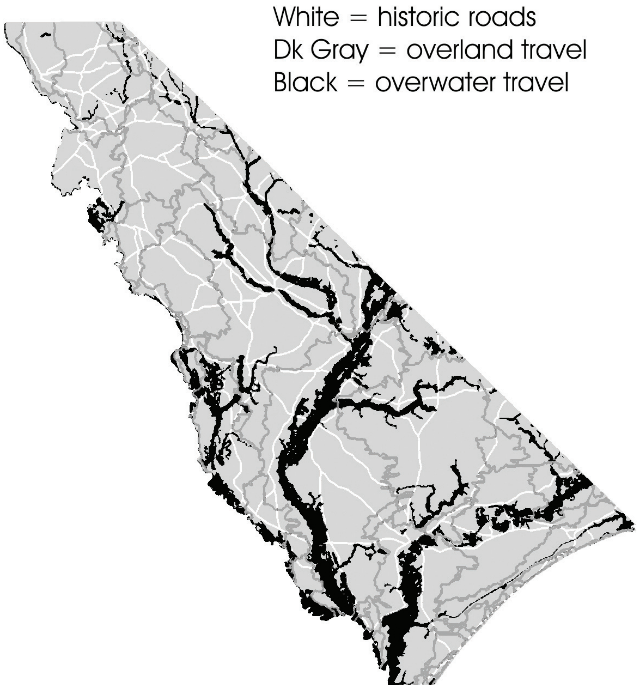

Figure 2. Regional travel corridors (pathways).

Formula Allocations

Settlement/ Subsistence Pattern:	Behavioral Categories: RA	D/P	SI
PBN	F1, F2, F3	F4, F5, F6, F7	none
WAEN	F8, F9, F10	F11, F12, F13, F14	none
CEN	F15, F16, F17	F18, F19, F20, F21	none
SS	F22, F23, F24	F25, F26, F27, F28	F29, F30, F31, F32
PS	F33, F34, F35	F36, F37, F38, F39, F40, F41	F42, F43, F44, F45, F46

Figure 3. Formulas allocated by behavioral categories and settlement/subsistence patterns.

Formulas

F1: 0.5MH+0.5RL
F2: 0.25MH+0.75RL
F3: 0.75MH+0.25RL
F4: 0.5WP+0.25RL+0.125S+0.125MH
F5: 0.25WP+0.5RL+0.125S+0.125MH
F6: 0.25WP+0.25RL+0.25S+0.25MH
F7: 0.5WP+0.5RL
F8: (UWP+LL+RL)/3
F9: 0.5UWP+0.25LL+0.25RL
F10: 0.25UWP+0.5LL+0.25RL
F11: 0.2UWP+0.2LL+0.2RL+0.2WP+0.2S
F12: 0.1UWP+0.1LL+0.3RL+0.4WP+0.1S
F13: (UWP+LL+WP)/3
F14: 0.5LL+0.5WP
F15: 0.33GWP+0.33LW+0.33RL
F16: 0.5GWP+0.25LW+0.25RL
F17: 0.25GWP+0.5LW+0.25RL
F18: (GWP+LW+RL+WP+S)/5
F19: 0.1GWP+0.1LW+0.3RL+0.4WP+0.1S
F20: 0.33GWP+0.33LW+0.33WP
F21: 0.5LW+0.5WP
F22: (H+LW+RW+GWP)/4
F23: 0.1GWP+0.1LW+0.3RL+0.4WP+0.1S
F24: 0.5H+0.125GWP+0.25LW+0.125RW
F25: (H+GWP+LW+RW+WP+S)/6
F26: 0.3H+0.1GWP+0.1LW0.2RW+0.2WP+0.1S
F27: (H+LW+WP)/3
F28: (H+GWP+LW+WP)/4
F29: (H+RW+WP+S)/4
F30: 0.4H+0.2RW+0.3WP+0.1S
F31: 0.2H+0.4RW+0.2WP+0.2S
F32: 0.2H+0.5RW+0.3WP
F33: (A+GWP+LW+RW+HR)/5
F34: 0.1A+0.3GWP+0.1LW+0.1RW+0.4HR
F35: 0.4A+0.1GWP+0.1LW+0.1RW+0.3HR
F36: (A+GWP+LW+RW+WP+S)/6
F37: 0.3A+0.1GWP+0.1LW+0.1RW+0.4WP
F38: (A+RW+WP)/3
F39: (A+GWP+HR)/3
F40: 0.4A+0.2GWP+0.4HR
F41: 0.2A+0.2GWP+0.6HR
F42: (A+RW+WP+S)/4
F43: 0.4A+0.2GWP+0.3WP+0.1S
F44: 0.2A+0.4RW+0.2WP+0.2S
F45: 0.2A+0.5RW+0.3WP
F46: 0.5RW+0.5WP

MH = Potential for Megafauna Habitat
RL = Cost Distance to Regional Overland Pathways
WP = Cost Distance to Permanent Water Sources
S = Relative Slope
UWP = Upland Wildlife Potential
LL = Cost Distance to Local Overland Pathways
GWP = General Wildlife Potential
LW = Cost Distance to Local Overwater Pathways
RW = Cost Distance to Regional Overwater Pathways
H = Potential for Horticulture
A = Potential for Agriculture
HR = Cost Distance to Historic Roadways

Figure 4. Formulas used in the analysis.

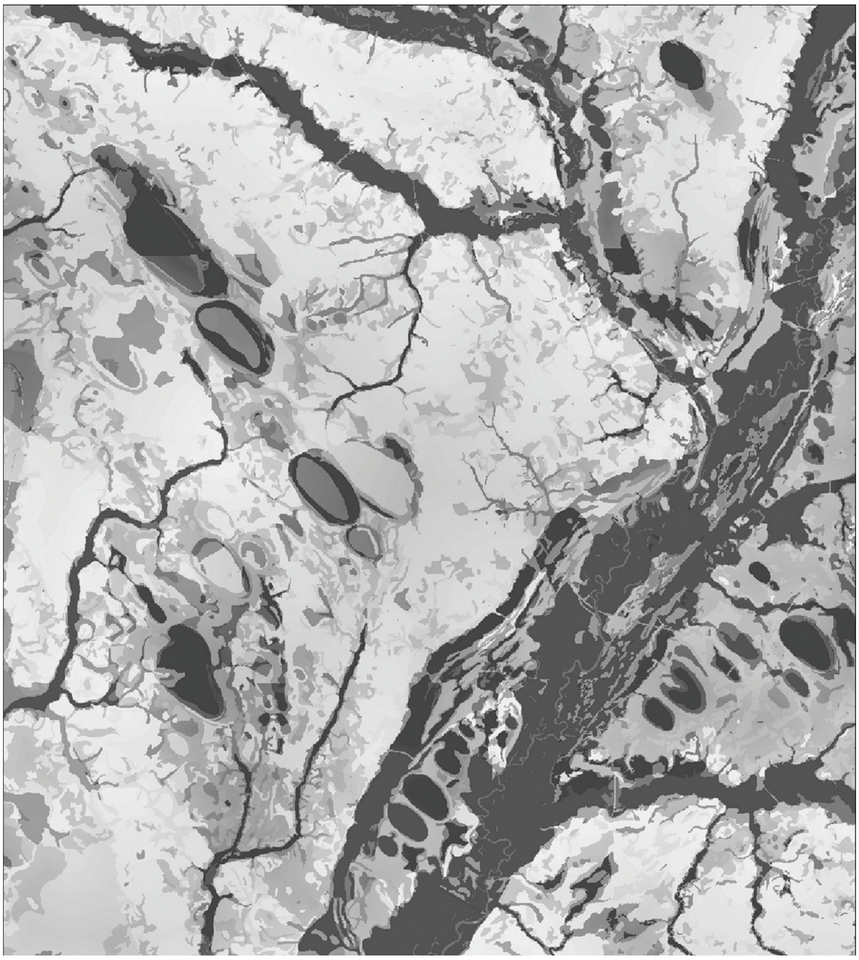

Figure 5. One of the 46 probability surfaces (detail): lighter = greater potential for archaeological resources.

showed the greatest percentage of applicable sites with high predicted values was selected. This is the same as an assessment of the formula's accuracy.

Additionally, however, a histogram of the total number of land units by value for each formula was also produced. Any formula that tended to produce only very high predicted values was thrown out as being not discriminative enough. This is the same as an assessment of the formulas precision (or specificity).

We were saddled with a problem, though, in that the existing archaeological site data (1,051 sites previously recorded in these four counties) do not include descriptions of site types or probable associated behaviors; merely temporal periods and eligibility status. Our selection of the most appropriate formulas to use in the final analysis relied on some assumptions about the nature of eligible sites versus non-eligible ones, and for some settlement/subsistence patterns the testing dataset was rather small. In the end, this model was for alternatives analysis, and therefore, once a preferred alternative is selected and field survey has begun, the results may further validate or repudiate some of the selected formulas.

The formulas created and selected (Figure 5 is a detail of one of the selected formulas) are not abstract compilations of variable attributes put together by finding the lowest common denominator among a set of already biased archaeological sites. Rather, they represent real approximations of the environmental and social variables that theoretically were most important to past people, based on the research done by the archaeologists working in the region. Thus, they are explanatory and provide causal relationships (cf., Salmon 1998) between past human actions (behaviors) and their effects (archaeological sites). Though these formulas are general and somewhat reductionist on the large scale, they are far more enlightening than a correlative model that provides no causal reference and no explanation.

8 Adapting the Formula for Land Management

One further step was entailed in the process of producing the predictive model, however adapting it for use by non-archaeologists. In its application, the model was to be incorporated into an overall cost-benefit analysis of each alternative. This analysis incorporated other parameters such as wetlands, endangered species, socioeconomic issues, historical structures, land purchase agreements, and engineering cut and fill constraints, to name a few. Archaeological potential was merely one small aspect of the overall analysis. Therefore, the results of the model had to be reduced to a single surface that meant something tangible to a non-cultural resource specialist.

To do this, all of the final formulas were combined into a composite surface and transformed to range in decimal value between 0 and 10, where the value of 0 indicates an area known not to contain significant archaeological resources (i.e., it is completely disturbed, is unsurveyable, or is part of an archaeological site which has been identified and found to be not significant), and 10 indicates the location of a known significant site. All other values represent the relative likelihood of encountering archaeological resources, with the caveat that the relative value reflects the average density of sites in the region, and density of positive shovel tests within the average site in the region.

This being the case, a comparison was possible by averaging the probability values for each alternative to get a general idea of which one is likely to be most expensive to survey or mitigate any adverse effects to significant sites. Conversely, it was also possible to define categorical limits to "high," "moderate," and "low" potential zones and compare alternatives by summing the acreages of each category. Ultimately, we produced both raster continuous data, and vector categorical data, so that the choice of which to use was up to the design engineers and the alternatives analysts.

To date, we have found no general formulas which contradict the findings of previous regional research (almost all of which has been published only as gray literature) regarding the possible settlement of different time periods or site types. However, the lack of data is still vast, and as more research is conducted and the models are tested further we may find some interesting results. This may particularly be the case if very specific behavioral patterns are targeted – none of which we have done at this point. The results of the I-73 survey itself are also not yet complete, so evaluation of our general models in the context of this study is also still incomplete.

For our purposes as archaeologists, the individual surfaces and the manner in which we formulated explanatory understandings of how past people cognized, experienced, and utilized their natural and cultural environments was of the most interest and applicability. The potential to produce much more detail and test many more ideas is what holds the greatest promise.

References Cited

Church, Tim, Brandon, R. Joe, Burgett, Galen R. 2000. GIS applications in archaeology: method in search of theory. In, *Practical Applications of GIS for Archaeologists: A Predictive Modeling Kit.* K. Wescott and R. Brandon, eds., pp 135-156. London: Taylor and Francis.

Salmon, Wesley C. 1998. *Causality and Explanation*. Oxford: Oxford University Press.

van Leusen, Martijn, Deeben, Jos, Hallewas, Daan, Zoetbrood, Paul, Kamermans, Hans, Verhagen, Philip. 2002. *Predictive Modelling for Archaeological Heritage Management in the Netherlands*. Baseline Report for the NWO (Humanities Section) of the BBO (Stimuleringsprogramma Bodemarchief in Behoud en Ontwikkeling), Amersfoort.

Wheatley, D. and Gillings, M. 2002. *Spatial Technology and Archaeology*. London: Taylor and Francis.

Whitley, Thomas G. 2003. Causality and cross-purposes in predictive modeling. In, The E-way into the Four Dimensions of Cultural Heritage. CAA2003.Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 31st Conference, Vienna, Austria, April 2003. Magistrat der Stadt Wien – Referat Kulturelles Erbe – Stadtarchäologie Wien, eds., pp 236-239. BAR International Series 1227. Oxford: Archaeopress.

Whitley, Thomas G. 2004. *Re-thinking Accuracy and Precision in Predictive Modeling*. Paper Prepared for the Computer and Quantitative Applications in Archaeology 2004 Conference, Prato, Italy, April 18-22.

Whitley, Thomas G. 2005. A brief outline of causality-based cognitive archaeological probabilistic modeling. In, *Predictive Modelling for Archaeological Heritage Management: A Research Agenda*. M. van Leusen and H. Kamermans (eds), Nederlandse Archeologische Rapporten 29, pp 125-139. Amersfoort: ROB.