5.0 A REVIEW OF RELEVANT PREVIOUS PREDICTIVE MODELS

Archaeological predictive models were developed in association with the U.S. 13 Corridor Project (1984), the Route 896 corridor study (1987), and the previous U.S. 301 corridor study (1993). All three of these models considered some or all of the area that falls within the current project's study area. A review of these models, preceded by a brief general discussion of the use of archaeological models, follows below.

5.1 General Discussion on the Use of Predictive Models

Predictive modeling in archaeology evolved from the discussion of settlement patterning or "settlement archaeology" that emerged in the New Archaeology in the 1960s. Essentially, the approach assumes that if the pattern of site locations can be established for a particular cultural area, that patterning can be used to predict where undocumented sites might be found within similar cultural settings. Effective, comprehensive models can quickly and efficiently predict areas of archaeological interest and save time and energy by focusing surveys on only those areas with the greatest potential to contain sites. In the context of the current study, this modeling can also be used as planning tool, to aid in the identification of archaeologically sensitive areas during the course of the development of designs for projects that could have deleterious impacts on such deposits. The advent of the use of GIS in archaeology has also fostered this approach, in that it permits the control and manipulation of vast amounts of data (Levy et al. 2001), and its use in predictive modeling can be equated with the application of much more sophisticated and precise methods (Westcott and Brandon 2000). GIS has permitted investigators to consider simultaneously a variety of environmental variables in predicting likely archaeological site locations.

There are several important theoretical and interpretive elements to consider in any discussion of site predictive modeling. In general, there are two distinctive approaches to the problem of predicting the location of archaeological sites. These have been classified as either *empirical/inductive* or *deductive* (Hay 1993:40-41) in the literature. The contrast between these is similar to that seen between two concepts intrinsic to the discussion of settlement archaeology itself. In that discussion, the generally perceived goal is to discover what factors are responsible for settlement patterning and why. To achieve that goal, a methodological distinction is made between the terms *settlement pattern* and *settlement system*. A *settlement pattern* is defined as a non-random distribution of archaeological sites on a landscape. This pattern is solely an archaeological manifestation, consisting of various types of sites, their physical characteristics, and their location in an explicitly defined geographic region. The patterning is thus determined inductively. A *settlement system* is defined as the cultural adaptation that produced the observed distribution pattern of archaeological sites. As such, a settlement system is a theoretical construct, a model based on the archaeologists' understanding of which criteria were considered significant by the people in a culture as they exploited their physical and social environment. In settlement archaeology, characterizing settlement pattern is a proximate goal; the ultimate goal is to characterize and then test the settlement system, of which the pattern is the physical manifestation. The deductive approach to site prediction begins by modeling the abstract settlement system, from which a settlement pattern may be predicted. This analysis produces two linked models, that of a settlement system and settlement pattern.

The Empirical Approach

The empirical/inductive approach has received much greater attention than the deductive methodology from archaeologists working in the United States. At its simplest, an inductive/empirical approach is based on low-order generalizations based on empirical observations (e.g., "sites are found within 300 feet of water" or "sites are found on level, well-drained soils"). These generalizations are based on observed correlations between the location of archaeological sites and attributes (as defined by the researchers) of the natural or cultural landscape.

Warren and Asch (2000:6) characterize predictive "...models are tools for projecting known patterns or relationships into unknown times or places". According to these authors, predictive models in prehistoric archaeology are based on three basic assumptions: 1) prehistoric sites are distributed non-randomly; 2) natural environmental features are, in part, responsible for that non-random distribution; and 3) those natural environmental features are visible in contemporary maps and databases.

Methodologically, the empirical approach begins by defining what features in the environment are correlated in a statistically significant way with the location of documented archaeological sites (e.g., Rose et al. 1995; Duncan and Schilling 1999). Once the environmental feature(s) or factor(s) have been isolated, predicting the location of undocumented sites becomes a matter of mapping all locations within a given study area where the determining environmental feature(s) are found. Verification of the predictive model is achieved through archaeological survey of those areas that, according to the model, have a high probability of containing undocumented sites. If the model has accurately correlated the locations of known archaeological sites with the determining environmental variables, sites will be found where, and only where, they are predicted.

Perhaps the most fundamental criticism of the empirical/inductive approach derives from its being based (of necessity) on the set of previously reported sites. It is generally understood in the archaeological community that the site location data available to researchers is often of variable reliability and reflects the biases inherent in the methods by which sites are identified and recorded (Hay 1992, 1993; Duncan and Schilling 1999; Miller et al. 2000). These biases will necessarily be replicated in an empirically-derived predictive model. The empirical/inductive approach is also limited in its ability to characterize the type of site that might be encountered in a particular setting, as the method does not include as a criterion any consideration of the underlying behaviors expressed at the site.

Because the empirical approach to predictive modeling is entirely inductive, it is an atheoretical approach to archaeological inquiry. Site predictive models depend on empirical discovery of statistically significant correlations and produce a series of generalizations that are valid only for the area from which the data are derived. Each study is independent, and the findings are not cumulative, in the sense that the results of one study cannot be used to refine the predictive model for future studies in other areas. This problem occurs because the approach provides no theoretical link between predictor variables or between those variables and human behavior. These generalizations are based on observed (i.e., etic) correlations between the location of archaeological sites and their settings' attributes. There is little or no theoretical backing or rationale as to how those variables relate to each other, much less how they relate to cultural adaptations. (For example, one study considered variables of "minimum depth to seasonal high water table," "soil erodibility," and "soil parent material," without providing any rationale for how those variables actually affected prehistoric decision making regarding site location [Warren and Asch 2000:15]). Consequently, our understanding of how or why the variables predict site locations is not developed beyond the obvious, thereby limiting the usefulness of the models.

The Deductive Approach

A deductive approach to the problem of site predictive modeling shares a common assumption with the empirical approach and with archaeological settlement pattern analysis – archaeological sites are distributed non-randomly across the landscape and environmental and cultural factors are, in part, responsible for their non-random distribution. The deductive approach differs in that the analytical point of departure is not with the distribution of known sites, but with the recognition that human beings select the location of their settlements on the basis of conscious decisions grounded on a set of physical and social needs. By understanding these needs, a researcher can predict not only where settlements are located, but also why they are located where they are.

As the name implies, the deductive approach begins with a model (or models) of cultural adaptation to a specified area. The pattern of settlement locations is derived (deduced) from the distribution of resources that are believed to have been of value to the members of a culture. Not only can the physical location of archaeological sites be predicted on the basis of settlement pattern, but the types of sites (i.e., what types of resources were exploited at each site location) can also be predicted. This approach aspires to bringing the archaeologist closer to an understanding of the worldview of people being studied.

This method has the additional advantage of providing a closer link between predictive modeling and subsequent field testing. By providing predictive statements not only on the probable location of archaeological sites but on the *types* of sites likely to be found, Phase I archaeological testing can become a first step in a deductive research program that uses the site-identification phase of the cultural resource management process to begin testing not only the settlement pattern model, but also the settlement system model that underlies it. A predictive model that provides both site location and site type can be utilized to formulate the most appropriate Phase I archaeological field and analytical methods to test the model and its assumptions.

There are, of course, shortcomings to this approach as well. Perhaps most fundamental is the limitation imposed by the inaccuracies of the theory on which the model is based. Particularly in the case of models involving prehistoric peoples, assumptions about valuations of aspects of the environment are likely to be in error, particularly when they involve considerations beyond subsistence activities. These biases will inevitably prejudice the model's predictions. This is also a significant factor when attempting to model the earliest historical uses of the landscape. Except for those areas for which accurate mapping of early transportation routes and settlements exists, models will of necessity be based in current theory, which may prove to be grossly inaccurate (DeCunzo, personal communication, 2005). For example, Fithian (personal communication, 2005) has suggested that several of the orthodoxies regarding the earliest (seventeenth and early eighteenth century) European settlement of Delaware's interior are not supported by the results of his recent studies in Kent and Sussex Counties.

In the case of prehistoric studies, another important factor is the role paleoenvironmental reconstructions play in developing models. Of course, the dramatic impacts urbanization and modern agriculture have had on landscapes obscure what would have been important ecological settings and topographic features in prehistoric times. Paleoenvironmental reconstruction is further limited by researchers' ability to accurately depict the climate and associated biota of the Early Holocene, or to delimit climatic cycles within the Middle and Late Holocene. A general characterization of the paleoenvironment of the Delmarva Peninsula in the period postdating the emergence of the deciduous forest and the relative stabilization of sea levels after the glacial retreat can perhaps be reconstructed with some confidence. However this extends back only to the Archaic period (Jacobson et al. 1987:282), and finer resolution beyond broad depictions is problematic.

Another problem associated with a deductive methodology, limited largely to the prehistoric period, is the problem of defining archaeological site types. Much of the contemporary discussion of prehistoric site function, settlement patterns, and settlement systems, particularly in relation to hunter/gatherer adaptations, is based on a series of studies by Wagner (1960), Steward (1968a), Cleland (1976), Gamble (1978) and Binford (1980, 1982). These investigators linked subsistence strategies with settlement patterns and with the characteristics of the natural environment that determined the pattern of subsistence/settlement strategies employed.

Binford (1980) created a number of well-defined archaeological correlates to each type of hunter/gatherer adaptation. Site types corresponding to the forager adaptation included: (1) the *residential base*, the "hub of subsistence activities"; and (2) the *location*, where extractive tasks are performed. Site types associated with the collector adaptation were more varied and included: (1) the *residential base*, the "hub of subsistence activities;" (2) the *location*, where extractive tasks are performed; (3) the *field camp*, a site created by specific task groups; (4) the *station*, temporary lookout point to find game; and (5) the *cache*, created when resources gathered are too bulky to return all goods back to camp; they are stored on-site.

In 1982, Binford expanded his discussion of hunter/gatherer adaptations and the associated archaeological site types to model the use of space and patterns of movement by hunter/gatherers. He defined the following spatial contexts for individual archaeological sites: (1) the *camp area* included the immediate perimeter around a site; (2) the *foraging radius* included the area that could be exploited within a day's foraging trip (foraging groups produce "location" types of sites within the foraging radius); (3) the *logistical radius* included the area that task groups operated in and which produced "field camp" types of sites during their collecting trips of several days duration; and (4) the *visiting zone* included the area within which people would travel for other purposes.

5.2 1984 Reconnaissance Study for the U.S. 13 Relief Route Corridor

Custer et al. (1984) proposed a predictive model in support of planning studies for the U.S. 13 Relief Route Corridor Project. The predictive prehistoric component of this model was formulated based on a set of environmental variables defined by the authors as significant to site location choices. The model used LANDSAT imaging with a resolution of 500 feet to delineate environmental settings. Settlement models were proposed for each specific prehistoric period

(Paleoindian, Archaic, Woodland I, and Woodland II), and topographic settings likely to contain these sites were identified. For the Paleoindian Period, the model proposed a correlation between base camps and "well-drained ridge in areas of maximum habitat overlap" (pg. 59). It also identified swamps and bay/basin features as being associated with sites. Archaic sites in the Midpeninsular Drainage Divide were likely to be found in association with stream terraces, swamps, and confluences. Woodland I and II sites would fall on "well-drained knolls at springs and stream confluences" or on "well-drained knolls at swamp and springs" (pp. 68, 75).

The approach defined probabilities using an inductive methodology based on the population of known sites and a logistical regression analysis that defined relative probabilities between 0 and 1. Three probability classes were defined by the model: high probability had a p value greater than 0.75, moderate one between 0.50 and 0.75, and low one less than 0.50. (Custer [personal communication, 2005] noted that in this scheme low probability areas could have as much as a 50% chance of containing a prehistoric site.) To test the model, reported sites and previously studied areas that had been removed from the sample before the model was constructed were evaluated by the criteria of the model. High and moderate probability zones were considered, and in both instances, the model proved effective.

The historic archaeological model was straightforward. High potential was assigned to areas near standing historic structures or near the location of ones no longer standing that appeared on historic maps. The intent of the approach was to base the model on the known historic-period settlement pattern.

In subsequent reports on the results of Phase I archaeological surveys undertaken in association with the project, Custer and Bachman (1986) and Bachman et al. (1988) concluded that the model had accurately predicted the contexts within which sites were ultimately found. In addition, Custer and Bachman (1986) discussed at length the important role bay/basin features had played in prehistory. Studies undertaken by Egghart et al. (2003) on the Frederick Lodge Site Complex (7NC-J-97-99), a site complex located on a rise above two bay/basin features, tended to support this assertion.

5.3 1987 Phase I & II Archaeological Investigations of the Route 896 Corridor

Lothrop et al. (1987) reported on the results of a Phase I/II survey undertaken in a corridor along Route 896. Nearly all of the APE associated with that survey falls within the APE for the U.S. 301 Project Improvements. In their discussion of the initial evaluation of the prehistoric archaeological sensitivity of the APE for that project, the authors report that they reviewed the CRS files at the Bureau of Archaeology and Historic Preservation. That review identified 34 sites within or near the project's APE, all of which fell within the Midpeninsular Drainage Divide. Of those 34 sites, 31 lay within 200 meters of an existing or remnant drainage, and 13 lay within 100 of such water features.

The authors suggested that because of its relative topographic uniformity, the Route 896 APE would have represented an area with "a low level of environmental diversity in the prehistoric period" (pg. 28). They used distance to water as the sole criterion for determining an area's relative prehistoric archaeological potential, defining two classes of potential: high, for those areas that fell within 200 meters of a water source, and low, for those that fell further than 200 feet from water. Historic archaeological potential was believed to be high near Glasgow, a transportation node, and along the historical alignment of Route 896.

The field survey undertaken in association with the project identified 15 archaeological sites, 11 of which had prehistoric components and six of which had historic components. All of the sites with prehistoric components fell with 200 meters of a water source. The authors reported that "the criteria used for assigning portions of the Project Area to high or low probability of prehistoric site location were valid" (pg. 189). The authors go on to state that "in areas of low environmental diversity, archaeological survey designs which rely on one or only a few criteria for site location prediction may be appropriate" (pp. 190-191).

The six sites with historic components were found to be associated with transportation resources, most notably with Glasgow and then with the road itself. They were related to agriculture, as had been expected. Because no sites associated with the residences of landowners were encountered, these were presumed to have been built at a distance from the road outside the limits of the project's APE. The one eligible historic residential site found during the survey, the Cazier Site, was a tenant house that had lain closer to the road and, as a consequence, fallen within the APE.

5.4 1993 Reconnaissance Study for the Route 301 Corridor

Kellogg (1993) proposed an archaeological predictive model in a report association with a previous evaluation of transportation needs within the U.S. 301 corridor. Using LANDSAT characterizations of the locations of known sites, the report derived a formula for discriminating between likely site locations and unlikely ones. The results of this analysis were then used to predict the prehistoric archaeological potential within the study area. In addition, for the part of the study area that extended north of the Chesapeake and Delaware Canal, an Automated Environmental Resources Information System, or AERIS, a precursor to GIS, was used to evaluate prehistoric archaeological potential. Distance to drainages and distance to wet/dry soil ecotones were determined to be significant variables in predicting prehistoric archaeological potential in the AERIS-based analysis.

Pre-1770 historic archaeological potential was determined using vicinity to historic roads and navigable waterways as a criterion for determining areas of high, moderate and low probability. Within the current project's study area, an area of moderate probability was defined along (Old) Choptank Road. Post-1770 historic archaeological potential was defined in relation to an area's vicinity to standing historic structures or to the locations within which such structures had once stood, as determined by reference to historic maps.

The method presented in Kellogg (1993) represented an innovative approach to archaeological predictive modeling in Delaware, in that it used a computer program to manipulate remote sensing data derived from satellites (Landsat) and data from an early version of GIS (AERIS) to define areas of relative sensitivity. At its heart the approach was deductive. For prehistoric archaeology, it assumed that there was a correlation between the site types thought to be associated with the various prehistoric periods and specific environmental niches. For historic archaeology, it assumed an association between archaeological deposits and the locations of early transportation routes and historic structures.

The model's principal shortcoming lay in the inaccuracy of the data available to it. Landsat characterizations of areas were based on surface reflectivity, a non-specific characteristic that is relatively ambiguous in its implications. Variations between areas could be attributed to a number of factors or combination of factors, and there was (at the time) no definitive way to select the actual determining attribute or attributes. AERIS's limit of resolution was 500 feet, and only a relatively small set of environmental data were available in this format in 1993.

Given these limitations, along with the increase in the power of computers and improvements in the capabilities of the software available to manipulate data, it was decided to develop a new predictive model. The model presented in this report is akin to Kellogg's in the sense that it is at heart deductive and assumes characteristics of site selection on the part of Native American groups of the prehistoric period based on environmental criteria. The model was formulated to take advantage of the wide variety of environmental data now available in a GIS format and the much finer spatial resolution (limiting cell size = 10 meters) at which this data is reported.

Of the three models proposed for areas included within or in the near vicinity of the U.S. 301 Project Development, the model developed for the Route 896 Corridor study is the one most relevant to the current project. Like the current project, the Route 896 study fell for the most part within the Midpeninsular Drainage Divide. The Route 896 model defined vicinity to water as the most significant criterion (indeed, as the only criterion) for determining prehistoric archaeological potential. The field work subsequently undertaken along the corridor served to test the model, and within the parameters of the field methodology, the results appeared to support the model's predictions: all of the prehistoric sites found during the course of the survey fell within areas defined as being high probability. In those areas evaluated by a pedestrian survey (i.e., plowed fields with good surface visibility), the data were comparable between high and low probability areas and therefore provided a better evaluation of the model. In those areas where subsurface excavations (shovel test pits) were used instead, the data were not comparable, as a shorter interval between tests was used for the high probability areas (20 meters) as opposed to the low probability areas (30 meters). Still, the results of the survey seemed to confirm the model. Interestingly, there was no discussion of the three outlier sites (distance to water > 300 meters) identified from the site files.