ANALYSIS OF SITE SURVEY RESULTS

This section of the report will describe analyses of the results of the site survey.

The report on planning in the northern portion of the project area (Custer and Bachman 1986) included analyses of both prehistoric and historic site locations. Unfortunately, there are insufficient data on historic site locations in the southern area to warrant analysis, although these historic site locations will be included in a future analysis of all 2000 historic site locations from New Castle and Kent Counties in the initial Route 13 Corridor study. Therefore, the prehistoric site location analysis presented here will consist of an analysis of site locations with respect to the previously developed predictive model (Custer et al. 1984) and an analysis of environmental variables affecting site locations. The comparison of the predictive model's expected results and the observed results of the field survey provide an evaluation of the prehistoric predictive model. Analysis of site location characteristics provides a broader cultural perspective on prehistoric man-land relationships. All of these analyses include comparisons to other survey results from the Middle Atlantic region. Finally, a general discussion of the survey results for each of the major time periods is included.

EVALUATION OF PREHISTORIC PREDICTIVE MODEL

The following provides a discussion of the relationship between the present survey's results and the prediction of prehistoric site locations provided in the original Route 13 Project Planning report (Custer et al. 1984:76-90, Attachment V). Although the predictive model derived from logistical regression analysis of LANDSAT data has been successfully tested, in a preliminary manner, in southern and central Kent County (Eveleigh et al. 1983; Custer et al. 1986; Eveleigh 1984), and although a similar logistical regression model was successfully tested in south central New Castle County (Wells et al. 1981; Wells 1981), continued testing, and subsequent refinement, of predictive models is always desirable. Therefore, the survey results of this project were used to test the local predictions generated from the logistical regression analysis. Special consideration of testing of the predictive model's results is also presented here because several recent articles (Berry 1984; Tainter 1984) have questioned the use of predictive models derived from multivariate statistical analyses of site locations, such as logistical regression and discriminant analysis (Kvamme 1981; Burgess et. al. 1980), in cultural resource management situations. Because many of the criticisms focus on the models' accuracy, it was hoped that an analysis of the accuracy of the results of the LANDSAT model used for the Route 13 Corridor would support its use as a cultural management tool.

The best field test of the models'predictions would be a survey based on a stratified random sample. Stratification could be based on two factors: 1) predicted probability zone, and 2) environmental factors used in the development of the model, such as surface water setting, presence/absence of modern and ancient marshes, and local soils. Initial tests of the logistical model's predictions incorporated some of these features, but relied mainly on previously recorded sites rather than new field data. The present Route 13 fieldwork incorporates some of these aspects as well; however, for the most part it represents a non-random and uncontrolled "grab sample" with many inherent biases.

Probably the most important bias in the sample is introduced by the fact that the study areas were chosen for field survey because they included high concentrations of cultural resources. Consequently, the high and medium probability zones are over-represented in the samples studied. This over-representation means that tests of model predictions are more likely to detect errors of inclusion rather than errors of ommission. In other words, we will be able to detect errors where the model says a site is present and a site is not present. However, we will not be able to study as effectively errors where the model says a site is not present and a site truly is present. The nonproportional representation of some kinds of probability and environmental zones also introduces numerous sources of sampling error, especially in the absence of a random component in the sampling design. Because of the non-random nature of the sample, non-normality of some of the data and the various uncontrolled sources of sampling error and biases, complex statistical tests of goodness-of-fit between model predictions and field results were not used.

As a final introductory note it should be pointed out that all fieldwork within the nine study areas was carried out in a "blind test" setting. Although the field crews, including the project manager, knew that they were surveying within areas that were likely to find sites, ie. that they were in "problem areas", none knew exactly where the high, medium, and low probability zones were located. Therefore, all areas were field inspected with the same intensity and coverage. "Likely areas" within the study areas did not receive special treatment.

The first tests of model predictions focused on measures of site occurrence. These tests considered the goodness-of-fit between expected site counts and densities within the varied probability zones, based on predictive models, and the observed field results. Table 29 lists the data used for all areas combined. High and medium probability zones are over-represented in the test data and figures for low probability zones are somewhat suspect.

The first comparison of observed and expected results utilized a simulated 400m grid cell data base. This grid cell data base was utilized for a number of reasons. First, the 400m grid matched the grid system used for the development of the

TABLE 29

DENSITY MEASURE DATA

Probability Zone	#400m Grid Cells	#Cells w/ Sites	%Cells w/ Sites
H	36	28	82
M	110	74	60
L	96	23	30

initial predictive model (Custer et al. 1984:76-90). Also, a 400m grid was used in the initial testing of the predictive model (Eveleigh 1984; Eveleigh et al. 1983; Custer et al. 1986). Most importantly, application of the 400m grid allowed the generation of comparable measurements of site occurrence for analysis. The predictive model for site locations generated a series of probability values for the occurrence of sites within specified units of area. Field observations generate counts of sites within specified units of area. In order to directly compare results of field testing with the model predictions, some kind of data transformations must be undertaken. Statistical distributions which transform site densities into probability values, such as the Poisson, are of limited utility (see Custer 1979; 1980) because archaeological data do not clearly meet the necessary requirements of independence-of-events (Thomas 1975). In fact the logistical regression of distribution was used to avoid the problems encountered in applying the Poisson distribution. Consequently, it is easier to transform the probability values into expected frequencies of sites and then directly compare the expected site frequencies within different probability zones with the observed frequencies generated from sample survey data.

The transformation of the model's probability values was accomplished by first overlaying a 400m grid template on maps of the project area. The number of 400m cells falling within each of the probability zones was then recorded. The number of cells which contained archaeological sites was also recorded. In order to calculate the expected number of 400m cells containing sites, the midpoint value of each probability zone of each probability class was multiplied by the number of cells within the probability zone. The observed number of cells with archaeological sites was simply a count of the cells which contained archaeological sites.

Table 30 lists the observed and expected values which were then compared using the chi-square distributions. The test value, degrees-of-freedom, and probability value are all included in Table 30. The test statistics all indicate no statistically significant differences between the observed and expected frequencies of grid units containing archaeological sites. As an additional test of the model's predictions using the 400m grid data, the grid counts noted in Table 30 were converted to percentages and the expected and observed percentages were compared using the chi-square test (Table 31). No significant difference was noted. To summarize, the observed site frequency data fit well with the expected results.

TABLE 30

OBSERVED AND EXPECTED SITE COUNTS FOR 400m GRID CELLS

Probability Zone	Expected	Observed
H	32	28
M	69	74
L	24	23

Chi Square = .90 D.O.F. = 1 $.25<p<.50$

TABLE 31

OBSERVED AND EXPECTED SITE PERCENTAGE FOR 400m GRID CELLS

Probability Zone	Expected	Observed
H	88	82
M	62	60
L	25	30

Chi Square = 1.30 D.O.F. = 1 $.10<p<.25$

SETTLEMENT PATTERN AND SITE LOCATION ANALYSIS

The data collected on prehistoric archaeological sites during this survey of the southern portion of the Route 13 Corridor can be analyzed to provide numerous insights into the lifeways of the prehistoric cultures of the Low Coastal Plain area of the Delmarva Peninsula. The data from the present surveys are especially useful for site location analysis because they cover important areas of major drainages, such as the Smyrna and Leipsic Rivers, which had not been intensively studied previously. A series of cross-tabulations of site location variables and site characteristics and simple tabulations of frequencies of sites by environmental variables will be presented. Variables chosen correspond to those seen as important in other regional site location analyses (Custer 1980, 1984; Cavallo and Mounier 1980; Gardner 1978, 1982; Stewart 1980; Grossman and Cavallo 1982; Galasso 1983).

One variable recorded for the sites found in the Route 13 Corridor is distance to nearest water source in meters. Descriptive statistics (minimum value = 10, maximum value = 480, mean = 49.83, median = 24.00, standard deviation = 54.18, coefficient of skewness = 3.43, coefficient of kurtosis = 20.40) and the water distance Q-Q statistic (.8084, $.25>p>.10$) indicate a non-normal distribution. Figure 57 shows a box plot of the distance to nearest water source and indicates that 75% of the sites are within 100m of a water source. The skewed, non-normal distribution is caused by a few outlying large values ranging up to 400 meters. Because the study areas were focused on the Mid-Drainage Zone of major drainages, there was little or no variability in the types of surface water. It can be noted, however, that 163 of the sites (38%) were associated with a confluence of streams.

Table 32 shows the distribution of sites among the varied geomorphological settings. As might be expected, the stream-related settings predominate with stream terraces more common than floodplain settings. In Delaware's Low Coastal Plain floodplain settings are more commonly poorly drained than terrace or bluff settings. Therefore, the better drained terraces are preferred site locations. Nonetheless, some of the better-drained floodplain settings were still utilized. The importance of soil drainage is also underscored by looking at the distribution of sites by soil settings (Table 33). Only 2% of the sites are located on poorly drained soils.

TABLE 32

GEOMORPHOLOGICAL SETTINGS OF PREHISTORIC SITES

Geomorphological Setting	Count
Terrace	282
Floodplain	125
Bluff	12
Interior Knoll	5

FIGURE 57

Box Plot: Distance to Water

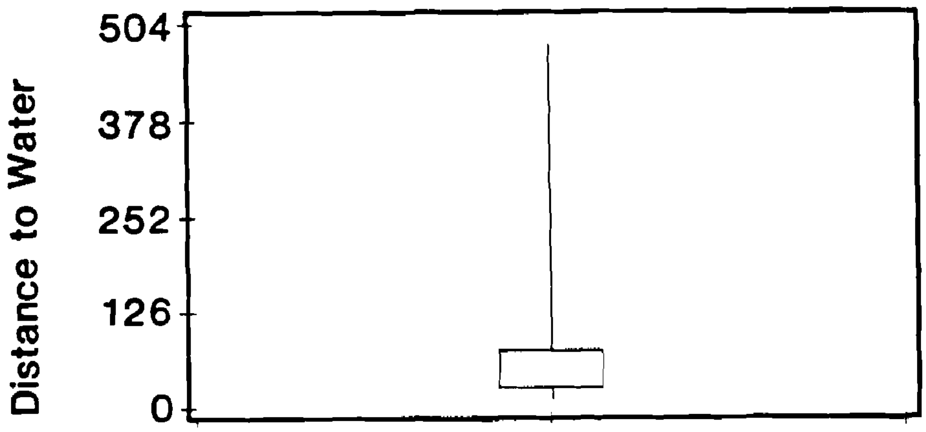

TABLE 33

SOIL SETTING OF PREHISTORIC SITES

Soil Setting	Count
Sassafras	343
Matapeake	40
Fallsington	10
Woodstown	10
Mattapex	9
Other	12

Note: All Soils are well-drained except for Fallsington

Table 34 shows the distribution of sites with respect to terrain aspect. It can be seen that there is no preferred site aspect, probably because the low relief of the Low Coastal Plain makes aspect and exposure unimportant components of site selection criteria. In sum, because the study areas covered in this survey were all similar floodplain settings there is little site location variability to analyze.