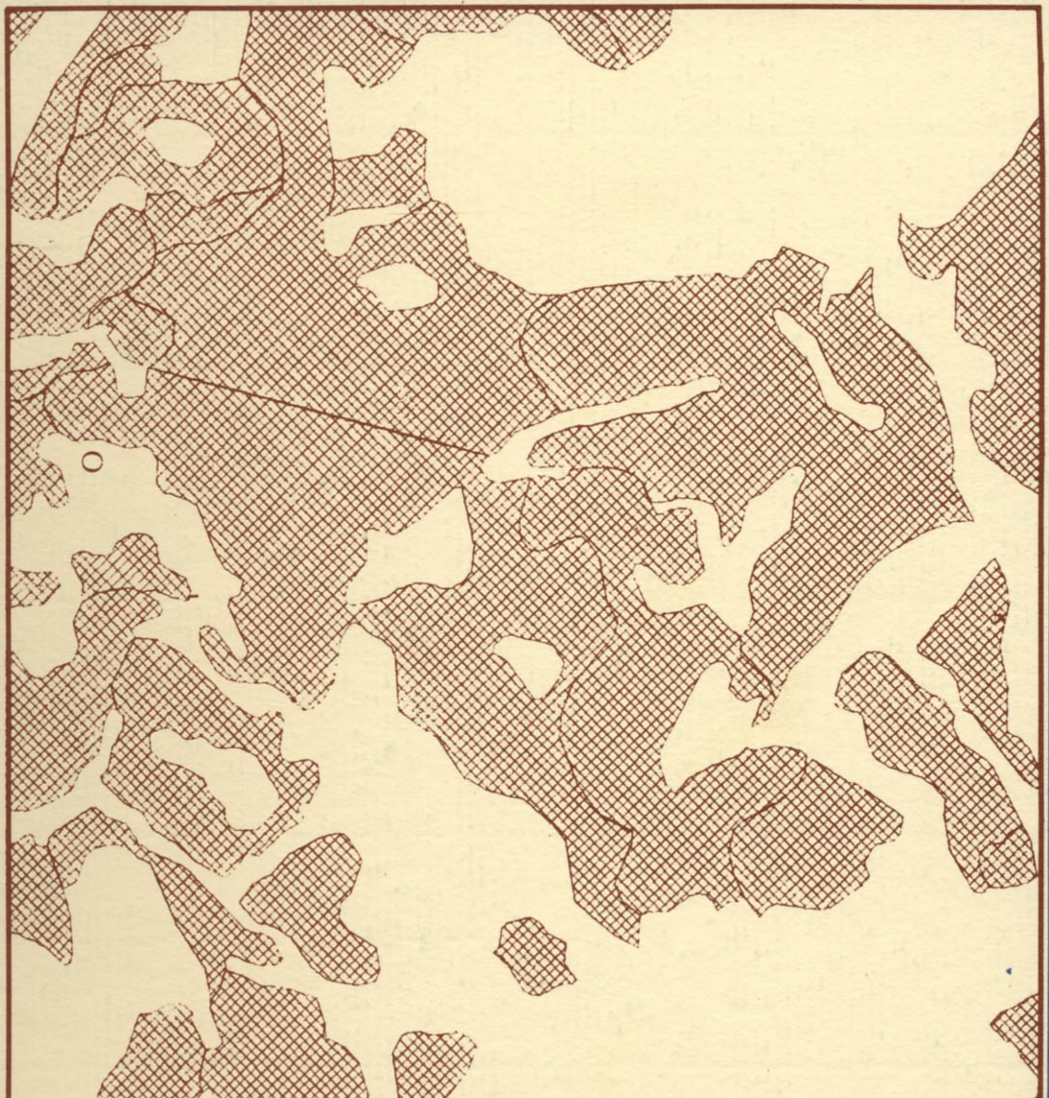


Archaeological Predictive Models

A New Hanover County Test Case

by Conran A. Hay
Catherine E. Bollinger Alan N. Snavely
Thomas E. Scheitlin Thomas O. Maher

North Carolina Archaeological Council
Publication Number 18

ARCHAEOLOGICAL PREDICTIVE MODELS: A NEW HANOVER COUNTY TEST CASE

by

Conran A. Hay
Alan N. Snavely
Thomas E. Scheitlin
Catherine E. Bollinger
Thomas O. Maher

NORTH CAROLINA ARCHAEOLOGICAL COUNCIL
PUBLICATION NUMBER 18

Published Jointly by the
North Carolina Archaeological Council
and the
Archaeology Branch
Division of Archives and History
North Carolina Department of Cultural Resources

Raleigh, 1982

TABLE OF CONTENTS

PREFACE

Predictive models of archaeological site location were identified as an important tool for planning and environmental review early in the development of the statewide plan for archaeological survey in North Carolina. As such, predictive models have been considered important elements of the State's automated system for managing cultural resources. However, before predictive modeling procedures could be adopted into the State's management system, important questions about the costs, reliability and utility of archaeological predictive models generated from site file information had to be addressed.

In late 1977 two developments caused the Archaeology Branch to initiate a test of predictive models. First, a computerized geographic information system was acquired by the State's Land Resources Information Service (LRIS). This system allowed predictive models to be geographically isolated. Second, a year long archaeological survey of New Hanover County was initiated under the Comprehensive Employment Training Act (CETA). Cooperative agreements were reached between the Archaeology Branch, LRIS, and the director of the CETA survey for conducting a predictive model test case. The CETA survey was to record site information on the State's computerized site forms, store the needed environmental data at LRIS and test a predictive model of site location developed by the staff of the Archaeology Branch. Time delays precluded the development of these models during the CETA survey. Late in 1978 the Archaeology Branch contracted with Dr. Conran A. Hay to develop, test and refine predictive models of site location using the New Hanover data. The implicit goal of the project was to evaluate the general utility of predictive models developed from standard archaeological site files for cultural resource management. This report summarizes that project.

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The New Hanover County Predictive Models project was initiated late in 1978 by the Archaeology Branch of the Division of Archives and History, North Carolina Department of Cultural Resources. The project was conducted in three phases. First, a descriptive predictive model based on previously collected site information was developed. It utilized four environmental variables: soil type, type of nearest water, elevation, and distance to water. The second, or field phase of the project, was designed to evaluate the preliminary model. A field survey based on probabilistic sampling procedures was conducted to establish the true distribution of archaeological resources in the county. After evaluating the preliminary model, alternative models were generated based on the controlled sample data from the field survey.

The final model was developed by means of regression analysis. Soil fertility and drainage character were the independent variables selected, and the extent of archaeological material to be expected within any given unit of land surface was chosen as the dependent variable. The precision of the model was found to be limited, to some extent, by the small size of the survey upon which it was based and by the accuracy of currently available data concerning the distribution of soil units within New Hanover County.

As a result of this project, certain criteria which predictive models should meet were developed:

1) Predictive models should be derived from probabilistic, representative samples of units of land surface, rather than from nonprobabilistic and potentially nonrepresentative samples of archaeological sites.

2) Predictive models derived using mathematical analytical procedures capable of handling the complexities of multivariate prediction may be required for many areas.

Models meeting these criteria, though initially more expensive, will be more efficient for the purposes of cultural resource management and should therefore facilitate the compliance process.

ACKNOWLEDGMENTS

The predictive model study reported in this monograph represents the coordinated work of many individuals and agencies. The coauthors of this report synthesized and reported this work. Thus, it is important to acknowledge and thank those that have had an impact on this project.

The Archaeology and Historic Preservation Section of the North Carolina Department of Cultural Resources was created in mid-1977. The Archaeology Branch was formed as part of the Section at this time. Jacqueline R. Fehon, Chief Archaeologist for the Archaeology Branch, and Brent D. Glass, Administrator for the Section, realized the need for the development of predictive models of archaeological site location and supported this study through appropriate staff assignments and the allocation of necessary funds.

The modeling study was implemented partially because of the Archaeology Branch's access to the Land Resources Information Service (LRIS). LRIS' computer system was used to isolate models geographically and to store environmental information. A cooperative agreement was reached between former LRIS directors Steven Pratt and Carol Simmermacher and the Section to test the utility of LRIS in isolating predictive models. Terry Ellis, of LRIS, trained Comprehensive Employment Training Act (CETA) and Branch staff on the use of the system. Karen Siderelis, also of LRIS, wrote several programs which facilitated the georeferencing process. Alan K. DeWitt, Dinah Wilde-Ramsing and Rick Ballenger, CETA employees, stored much of the geographic and archaeological data in the LRIS system. Additional storage and editing of the data was performed by Thomas O. Maher of the Archaeology Branch staff. Thomas E. Scheitlin, also of the Branch, edited and georeferenced the models.

Another major factor in the initiation of this predictive model project was the assistance provided by the personnel involved in the New Hanover County CETA survey. Mark Wilde-Ramsing, director of the CETA survey, collected the archaeological data used for the modeling study. His staff also stored most of the environmental information used by the study at LRIS. Their assistance was invaluable.

The Archaeology Branch contracted with Dr. Conran A. Hay to direct the project. He has donated many hours since the fulfillment of his contract toward the completion of this report. The statistical models for the project were developed by Hay and Alan N. Snavely. Snavely was responsible for the computer work associated with model formulation including the regression model. The people who collected the field data endured the most physically strenuous part of the project. Hay directed the fieldwork. Snavely served as the field chief. The crew consisted of Thomas O. Maher, Ellen "Corky" Piwoz and Sarah D. Tichnor. Members of the Archaeology Branch staff who assisted in the field work included Jacqueline R. Fehon, Mark A. Mathis, Elaine S. Nelson, Sandra O. Perry, Thomas E. Scheitlin and Carol S. Spears. Laboratory processing and analysis were conducted by Hay, Maher and Snavely.

We would also like to thank Dr. Gordon R. Murdock, former director of the Marine Resources Center at Fort Fisher, for permitting the Archaeology Branch

to erect a museum display in the center and for providing lab space during the field phase of the project. This gave the Branch the opportunity to generate public interest in the project by developing a display entitled "Computers and Coastal Archaeology". Carol S. Spears and Linda B. Luster designed, built and erected the display.

Technical support for the project was provided by a number of individuals. The final copy of the report was typed by Ida Landis. Numerous drafts of the report were typed by Lucille L. Walker and Sandra O. Perry. The initial graphics for the report were created by Linda B. Luster, and final graphics were produced by Margaret B. Pierce. Many people have reviewed and edited the report, including Catherine E. Bollinger, Conran A. Hay, Sandy Hay, Mark A. Mathis, Joseph Mountjoy, Thomas E. Scheitlin and Alan N. Snavely. All members of the Archaeology Branch staff have reviewed and commented on the report at one time or another.

On behalf of the Archaeology Branch I would like to thank all of the above and those whom I have neglected to mention. They all made valuable contributions to the project.

Thomas E. Scheitlin

CHAPTER ONE

INTRODUCTION

Current federal legislation requires the identification and evaluation of all significant historic and prehistoric properties that will be adversely affected by federally funded or licensed projects. During the last decade, predictive models have become increasingly important in the compliance processes mandated by this legislation. Using the locations of recorded sites as a guide, predictive models seek to identify the parts of a region that are most likely to contain other, undiscovered archaeological sites. Two important benefits are thus provided. First, agencies and corporations can assess, in a preliminary fashion and at a low cost, the magnitude and nature of project impacts to historic and prehistoric resources. Second, planners can design cost-effective strategies for the more complete resource inventories required by law prior to project implementation.

The State of North Carolina, through the Archaeology Branch of the Division of Archives and History, Department of Cultural Resources, is in the process of developing and testing predictive models and evaluating the role they play within the general context of cultural resource management. The New Hanover County Predictive Models Project was conceived and implemented as an initial step in this evaluation process. The primary aims of the project were twofold: 1) to develop a predictive model for the prehistoric archaeological resources of New Hanover County, North Carolina, and by so doing, 2) to address a number of more general issues of considerable importance in predictive modeling: i.e., (a) What kinds of data are required for the formulation of a reliable predictive model? (b) What analytical procedures should be used? and (c) What roles should predictive models play in the management and conservation of archaeological resources?

The New Hanover County Project was conducted in three phases: 1) initial predictive model development, 2) model testing, and 3) model revision. During the first of these phases a preliminary predictive model for the location of prehistoric archaeological resources in New Hanover County was formulated. In two important respects, this initial model was similar to many of the predictive models currently being developed by archaeologists. First, the model was derived from site location data resulting from the CETA survey. Second, it relied upon descriptive procedures to summarize this data and to indicate the types of localities most likely to contain sites. Because the model developed here conformed in these respects to current procedures within the field of predictive modeling, it served as a partial test of these procedures.

The second phase of the project involved evaluating the model developed during phase one. This was achieved through the use of statistical hypothesis testing techniques. A field survey based on probabilistic sampling procedures was conducted to reveal the true distribution of archaeological resources within the county. The data generated by this survey were then used to evaluate the model; specifically, statistical tests were conducted to assess the degree of congruence between the predictions of the model and the actual distribution of archaeological materials in the county, as documented by the survey.

The third phase of the project involved the exploration of alternative approaches to the problem of archaeological prediction. The data collected during the second survey phase of the project were used to develop alternative predictive models. These latter models were thus derived from more controlled sample data. Furthermore, more rigorous analytical procedures, such as regression analysis, were employed in their formulation. These models should prove to be more reliable in terms of predictive accuracy and more useful for cultural resource management purposes. The procedures used in the development of these alternative models have been proposed as preliminary guidelines for the formulation of predictive models in the future.

The pages which follow contain the specifics of the New Hanover County Predictive Models Archaeological Research Project, as structured by the three phases described above. This project provided an opportunity for an extensive investigation of predictive modeling within the general context of cultural resource management. As a result, this report presents theoretical and methodological discussions of predictive modeling issues, as well as substantive modeling results applicable to the prehistoric archaeological resources of New Hanover County.

CHAPTER TWO

NEW HANOVER COUNTY OVERVIEW

ENVIRONMENT

New Hanover County is within the Atlantic Coastal Plain topographic province; it exhibits typical coastal plain topographic and geographic features. Relic sand ridges, coastal bottomlands and uplands, and some small, shallow sinks are found in the interior; coastal areas consist of beaches, tidal marshes, and shallow sounds between barrier beaches and the mainland (Bain 1970). One-third of the county consists of swamps, marshes, and beaches. Elevations are generally low, ranging from 6 to 12 meters above mean sea level (MSL). The highest elevations occur in the sandhills, which extend in an approximately north-south direction along both the eastern and western boundaries of the county.

Hydrology

New Hanover County is bounded on the west by the Cape Fear River, on the east by the Atlantic Ocean, and on the north by the Northeast Cape Fear River (see Figure 2.1). The county can be divided into two major drainage systems, separated by a line running west and just south from Scotts Hill in Pender County to the Northeast Cape Fear River. North of this line the streams flow northward into the Northeast Cape Fear River; south of it they flow either eastward into the Atlantic Ocean or westward into the Cape Fear and Northeast Cape Fear River (Larsen 1958). The major streams which form this drainage pattern include Island Creek, Prince George Creek, and Smith Creek, all of which flow into the Northeast Cape Fear River. Mott Creek is the only major stream to flow into the Cape Fear River, whereas Futch Creek, Pages Creek, Howe Creek, Bradley Creek, Hewletts Creek, and Whiskey Creek all flow into the Atlantic Ocean. The drainage systems of these creeks generally do not penetrate a great distance into the interior of the county, but are instead limited primarily to its periphery near the three major bodies of water that form its boundaries. As a result, the interior of the county is poorly drained and flat, whereas the periphery is well drained and shows greater topographic relief. The Cape Fear and the Northeast Cape Fear rivers are tidal throughout those portions that bound New Hanover County; the smaller creeks that flow into the Cape Fear, Northeast Cape Fear, and the Atlantic Ocean are also tidal throughout major portions of their lengths (Fish 1968: 223; U.S. Corps of Engineers 1976:B-27). As a result, fresh water is available primarily from small streams, springs, and swamps.

Climate

New Hanover County summers are hot and humid; winters are mild. The average annual temperature is 19 degrees C (66.2 degrees F). Average July and January temperatures are 27 degrees C (80.6 degrees F) and 10 degrees C (50.0 degrees F) respectively. The county receives an average of 136 cm (53.54 in.) of rain per year, most (61%) of which falls between April and September. Both droughts and snowfalls are rare. The first freeze in the fall usually occurs in mid-November, and the last freeze of winter usually occurs at the end of February or in March. Thus, there are approximately eight frost-free months a year (Sharp 1954; Weaver 1977).

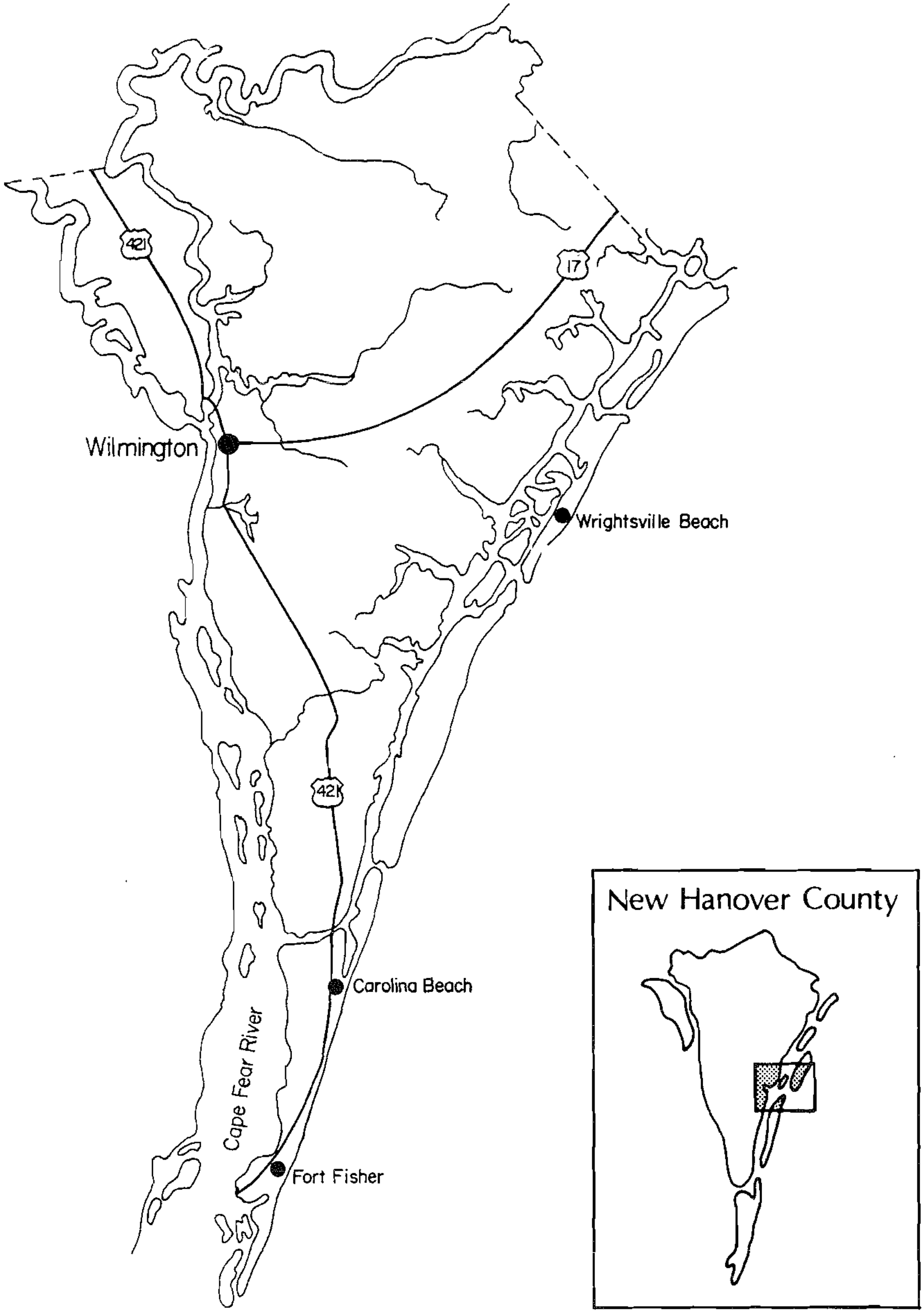


Figure 2.1 Map of New Hanover County

Geology

The underlying geologic strata of the county are typical of the Atlantic coastal plain and consist of 335 to 457 meters (1099 to 1499 feet) of sedimentary strata overlying a metamorphic and igneous basement. This basement is composed of schists, gneisses, granites, and other metamorphosed volcanics of Pre-Cambrian to Mississippian age. Presently these deposits are deeply buried beneath more recent, sedimentary strata, including the Black Creek, Peedee, and Castle Hayne Formations. This sequence is capped by the Pamlico terrace, which is the lowest and most recent of five Pleistocene marine terraces found along the Atlantic coast (Richards 1950).

The soils of New Hanover County are generally sandy but vary somewhat with local drainage patterns. Areas that are elevated relative to the local water table, such as small knolls, sand ridges, low hills, and local upland flats, have well-drained, rapidly permeable soils. In contrast, depressions of all kinds and extensive upland flats tend to have poorly drained soils as a result of the proximity of the water table to the land surface. The six major soil associations that are found in New Hanover County (Table 2.1) have been defined primarily on the basis of these variations in drainage character, and secondarily on the basis of broad, regional similarities. The Dorovan-Johnston association consists of poorly drained soils that are found on the floodplains of the Cape Fear and Northeast Cape Fear rivers. Since these rivers are tidal, most Dorovan-Johnston association soils are flooded daily. The upper layers generally consist of muck, loam, or sand rich in organic materials, and overlie muck or sand. The Kureb-Baymeade-Rimini association, the Wrightsboro-Onslow-Kenansville association, and the Kenansville-Craven-Lakeland association all consist primarily of soils that are well-drained and that have developed on upland areas between streams. These soils range in composition from sands to fine sandy loams. In depressions within these associations occur poorly drained soils belonging to various series. The Murville-Seagate-Leon association consists of poorly drained soils that have developed on broad, poorly drained upland flats and in slight depressions. Surface layers are sands or fine sands; subsurface layers are sands, fine sands, sandy loams, or clay loams. Knolls and ridges within the association exhibit various well-drained soil types. Finally, the Tidal Marsh-Newhan association consists of tidal marshes and associated dune formations found along the seashore (Weaver 1977).

The suitability of New Hanover County's soils for agricultural purposes varies largely with their drainage characteristics. In general, well-drained soils are suitable for the cultivation of crops, while poorly drained soils are not. Exceptions to this generalization exist, because excessively well-drained, sandy soils such as Kureb sand, Newhan fine sand, Rains fine sandy loam, and Rimini sand are not suitable for agriculture. Conversely, certain poorly drained soils, including Lynchburg fine sandy loam, Lynn Haven fine sand, Seagate fine sand, and Stallings fine sand, are suitable for agriculture when artificially drained.

Biota

Eleven ecological communities have been recognized in New Hanover County. Perhaps the most distinctive of these is the Maritime Forest (or Salt spray)

Table 2.1 Soil association and soil series names for New Hanover County, North Carolina (from Weaver 1977)

Soil Association and Soil Series	Series Abbr.	% of County	% of Assoc.
1. Dorovan-Johnston		6	
a. Dorovan	Do		60
b. Johnston	Jo		30
c. Pamlico	Pm		
d. Murville	Mu		
e. Lynn Haven	Ly		10
f. Torhunta	To		
g. Woodington	Wa		
2. Kureb-Baymeade-Rimini		7	
a. Kureb	Kr		25
b. Baymeade	Be		20
c. Rimini	Rm		8
d. Lakeland	La		
e. Leon	Le		
f. Lynn Haven	Ly		47
g. Murville	Mu		
h. Wakulla	Wa		
3. Wrightsboro-Onslow-Kenansville		7	
a. Wrightsboro	Wr		22
b. Onslow	On		20
c. Kenansville	Ke		15
d. Seagate	Se		
e. Stallings	St		
f. Norfolk	No		43
g. Pantego	Pn		
h. Torhunta	To		
4. Kenansville-Craven-Lakeland		3	
a. Kenansville	Ke		45
b. Craven	Cr		20
c. Lakeland	La		15
d. Leon	Le		
e. Norfolk	No		
f. Onslow	On		20
g. Wrightsboro	Wr		
5. Tidal Marsh-Newhan		3	
a. Tidal Marsh	TM		53
b. Newhan	Nh		25
c. Urban	Ur		22
6. Murville-Seagate-Leon		46	
a. Murville	Mu		26
b. Seagate	Se		13
c. Leon	Le		13
d. Kureb	Ku		
e. Lynn Haven	Ly		
f. Onslow	On		48
g. Pamlico	Pm		
h. Rimini	Rm		
i. Wakulla	Wa		
7. Urban		10	

community, which consists of salt tolerant scrub vegetation and is found in the immediate vicinity of the coast. Today this community is occupied primarily by small mammals; in the past large mammals such as black bear, wolf, and white-tailed deer were probably also present. Within New Hanover County, maritime forests are found only on Smith and Baldhead Islands and in the Fort Fisher area; however, they were probably more extensive in the past (Cooper 1964; Shelford 1963). Most immediate coastal areas are presently dominated instead by the Beach Dune Scrub community, which consists of grasses and low shrubs. Few animal species use this community exclusively, although sea turtles lay eggs within it and small mammals exploit it to a limited extent (Shelford 1963). Moving inland, the Tidal Marsh community dominates the sound areas between outer barrier islands and the mainland. It is also found along the tidal portions of major rivers and streams. The dominant vegetation again consists of grasses and shrubs. These areas are rich in shellfish, small mammals, and reptiles. Avian fauna, especially wintering migratory birds (Von Oesen 1976) are also abundant in these areas of tidal marsh. The Fresh Water Marsh community shares many of these characteristics, but is found in inland depressions and along fresh water streams. Swamp Forest, another wetlands community, is found in continuously flooded areas along the Cape Fear and Northeast Cape Fear rivers. This community is dominated by water tolerant tree species such as the bald cypress, but an understory of shrubs is also present. Swamp forests, like marsh areas, are important spawning and nesting areas and have many of the same aquatic, reptilian, and avian species. In addition, large mammals such as the white-tailed deer, black bear, and bobcat utilize the swamp forest biome (Funderburg 1955; N.C.D.O.T. 1976). The Pocosin community is also associated with wet areas and is found within flat upland areas with poorly drained soils, as well as in drainage ravines between sand ridges. The dominant tree species is the pond pine, but abundant and dense sbrub vegetation distinguishes the pocosin community. Pocosins do not serve as permannent homes for many animal species, but are important refuge areas for white-tailed deer, small mammals, and birds (N.C.D.O.T. 1976).

Well-drained upland flats in New Hanover County are dominated by Pine Savannah and Longleaf Pine-Turkey Oak communities. In the former, widely spaced longleaf and loblolly pines are separated by extensive areas of low shrubs and grasses. In the latter, the vegetation is dominated by mixed stands of longleaf pine and turkey oak. Within both communities, white-tailed deer and numerous small mammals are present, as are various reptilian and avian species. The Mixed Pine-Hardwood Forest is a transitional community also found in well-drained upland areas. It consists of mixed stands of loblolly pine, oaks, gums, and poplars (N.C.D.O.T. 1976). It shares many mammalian species, including the white-tailed deer, with the pine savannah and the longleaf pine-turkey oak communities. In addition, it is heavily used by migratory land birds (NC D.O.T. 1976). Two additional plant communities found within New Hanover County are maintained by the silviculture industry. Hardwood Forests consist of mixed stands of oaks, dogwoods, hickories, maples, and other hardwoods. These forests share many animal species with the pine savannah, longleaf pine-turkey oak, and the mixed pine-hardwoods communities. The nearly pure stands of pine maintained by the silviculture industry are known as Pine Plantations and consist primarily of longleaf, loblolly, and slash pine. These plantations are exploited by few animal species.

New Hanover County is exceptionally rich in aquatic resources. The Cape Fear River and the Northeast Cape Fear River are important spawning grounds for many marine fish, including anchovies, bluefish, mullet, gray trout, silver perch, American shad, Atlantic sturgeon, and tarpon (Wilde-Ramsing 1978). Fresh water fish are numerous in streams and ponds. The county is also exceptionally rich in migratory birdlife: Canadian geese, snow geese, various duck species, egrets, cormorants, and teal all frequent the area in season (Pearson et al. 1959). A more detailed listing of plant and animal components of the various ecological communities is provided in Appendices A and B.

Prehistory

Attempts to predict the locations of archaeological sites are predicated upon the assumption that such sites are not scattered randomly throughout the landscape. It is assumed that the prehistoric inhabitants of New Hanover deliberately patterned their behavior in response to the characteristics of their environment. An outline and a brief discussion of the salient features of these behavioral responses, as interpreted from the archaeological record of New Hanover County and surrounding areas, follows. Griffin (1952, 1967) established four broad chronological periods of Eastern United States prehistory: (1) Paleo-Indian, (2) Archaic, (3) Woodland and (4) Mississippian. The Archaic and Woodland periods are further divided into Early, Middle, and Late stages. Each of these cultural periods is represented to some extent in the archaeological record of coastal North Carolina. In the following discussion, the dates presented by Griffin for these periods have been altered to accomodate recent data applicable to the North Carolina coastal region.

Paleo-Indian

The earliest undisputed evidence of human occupation in North Carolina has been attributed to small nomadic groups of hunters and gatherers, who are collectively labeled Paleo-Indian. Grouped into bands of perhaps 25-30 people, they occupied the region during the final stages of the last North American glacial sequence, the Wisconsin. Associated dates range from 10,000 B.C. to 8000 B.C. Artifactual remains and interpretive data from the Paleo-Indian period are meager in North Carolina. Therefore, behavioral models of Paleo-Indian adaptations must be derived from data found in other areas of the United States (see Newman and Salwen 1977 for a comprehensive overview). These data suggest that subsistence strategies often focused on the Pleistocene megafauna, such as bison, mammoth, mastodon and caribou which were associated with the cooler, moist post-glacial environmental conditions. This traditional model has been further refined by recent data (e.g., Clausen et al. 1979) documenting the exploitation of smaller game and plants.

Archaeologically, the Paleo-Indian period is most readily identified by a distinctive form of projectile point, and it is the occurrence of these specimens which documents a Paleo-Indian presence in North Carolina. These lanceolate-shaped points are usually made from "exotic" lithic material, i.e., high-quality microcrystalline stone which is often from a non-local source. The points frequently exhibit a basal longitudinal flute on both faces, occasionally running the full length of the point. The distribution of Paleo-Indian projectile points in North Carolina has been documented by Perkinson (1971,1973), several of which have been recovered from the coastal counties.

The exposure of the Atlantic continental shelf during later Wisconsin times may have significantly influenced Paleo-Indian settlement patterns. The Paleo-Indian period closed with the advent of Holocene climatic conditions, which were accompanied by changes in floral and faunal communities. In response to this warmer, drier environmental situation, small bands of Paleo-Indian hunters and gatherers appear to have increased their reliance on a more diversified resource base (Gardner 1977). The transitional period between the Paleo-Indian emphasis on hunting and the later Archaic pattern of more extensive localized subsistence is represented in the archaeological record of North Carolina by Hardaway and Hardaway-Dalton projectile points. These broad-bladed points with narrow side notches and recurved, concave bases have been described in detail by Coe (1964) from excavated contexts at the Hardaway Site. Coe and McCormick (1970) also report a concentration of Hardaway materials from 31Ch159, near the confluence of the Haw and New Hope rivers. Both 31Ch159 and the Hardaway Site are on high ground overlooking major piedmont rivers. Comparable settlement data have been observed by Williams and Stoltman (1965) for other areas of the Southeast; settlement data have not been reported for the North Carolina coastal plain.

Archaic

The second major division of Eastern United States prehistory is the Archaic Period, developing from the Paleo-Indian Period around 8000 B.C. and lasting until approximately 1000 B.C. The dominant theme of the Archaic has been succinctly expressed as an ". . .increasing efficiency and success in exploiting the resources of the forest" (Caldwell 1958:6). The period is characterized by a slow steady growth in population, refinement of a seasonal economic pattern, emphasis on locally available resources, and increased reliance on food gathering activities. Associated with these adaptive responses is the development of regional variability among cultural systems (e.g., Ford 1974).

Early Archaic

The Early Archaic (8000 B.C. - 5500 B.C.) in North Carolina was a time of continued cultural adjustment to post-Pleistocene environmental conditions. This stage has been defined in North Carolina (Coe 1964), in Tennessee, (Chapman 1973, 1975, 1977), and from West Virginia (Broyles 1971). Data from these investigations have established a sequence of projectile point forms which are seen to constitute an evolving morphological continuum. The integrity of this continuum implies a stable behavioral and cultural pattern of adaptation. The following broad typological divisions have been created for North Carolina Early Archaic projectile points (in general chronological order): Palmer; Kirk Corner-Notched; the bifurcate tradition, including MacCorkle, St. Albans, and LeCroy; and Kirk Stemmed/Serrated. Some researchers also include the Stanly/Kanahwa forms (ca. 6000 B.C. - 5500 B.C.) (e.g., Keel (1976). These categories are defined primarily on the basis of morphological variability.

Little has been reported concerning the subsistence economies of Early Archaic peoples in North Carolina. Mathis (1979:31-33) has discussed theoretical models relevant to this concern; unfortunately, these models remain untested by field data. Thomas et al. (1977) have developed general subsistence models for the Delaware coastal plain which may be more directly applicable to the New Hanover County area. It should be noted that the models developed by Thomas et al. use resource procurement strategies to predict zones of site location. A reservoir study by Smith (1965) in Piedmont North Carolina indicates that Early Archaic occupations were situated on hills and knolls. Data from eastern Tennessee documenting extensive Early Archaic occupations in floodplains seem to support this model (Chapman 1977).

Middle Archaic

Projectile point forms serve as the chronological indicators for the Middle Archaic Period and its associated complexes: the contracting-stemmed Morrow Mountain point, the lanceolate Guilford point, and the side-notched points of the Big Sandy (II)/Halifax continuum span the Middle Archaic period. Cultural behavior during the Middle Archaic is still poorly understood, especially in the North Carolina coastal area. Coe (1964) suggests that the earliest Middle Archaic cultures may be intrusive into the Piedmont area, with ancestral affinities lying to the west. Cridlebaugh's (1977) analysis of the Morrow Mountain phenomenon, however, indicates that further study is required before adequate behavioral models can be developed for this period. Coe (1964) suggests that the later Guilford complex is focused in the North Carolina Piedmont, and Smith (1965) notes a subtle shift in site location from upland knolls to lower terrace and ridge formations. If these data reflect increased regionalization, comparable settlement shifts may be plausibly predicted for the coastal area. These anticipated shifts may correspond with settlement patterns to the north; for example, Coe (1964) suggests a northern influence for the Halifax subperiod in coastal North Carolina. The Middle Archaic terminates with evidence of increased sedentism and new technologies.

Late Archaic

Late Archaic cultural adaptations (2500 B.C. - 1000 B.C.) set the stage for the settled villages and maize horticulture of the subsequent Woodland period. The general trends of the Archaic -- increased resource specialization, population growth, and cultural regionalization -- are clearly reflected in Late Archaic artifacts and settlement patterns. Coe, in reference to the Late Archaic habitation at the Gaston Site (northern N.C. coastal plain), states that "Every indication suggested a larger group occupying the site over a longer continuous period than had been true of the earlier periods" (1964:119). Recent studies in the piedmont and mountain regions of North Carolina (e.g., Bass 1977) indicate that this behavior may be modeled as a consequence of social circumscription (Carneiro 1970).

Late Archaic occupations in the southeast are recognized by the presence of larger, rather crudely manufactured, Savannah River projectile points. Steatite (soapstone) vessels also appear in the archaeological record during the Late Archaic, and are easily viewed as indicators of increased permanence in habitation. Perhaps the most significant artifact appearing during the Late Archaic is pottery. Fiber-tempered pottery sherds from the South Carolina coast have been dated to 2000 B.C. The behavioral implications of this important technological innovation are not yet understood. For example, some cultures adjacent to the early pottery-using societies apparently found little immediate adaptive value in this technology (cf. Coe 1964 for the North Carolina Piedmont area), although ceramic pottery becomes one of the hallmarks of the following Woodland period.

Woodland

The Woodland period of prehistoric occupation in coastal North Carolina (1000 B.C. - A.D. 1650) contrasts markedly with earlier periods. Whereas Archaic subsistence strategies focused primarily on hunting and gathering, Woodland populations increasingly depended on cultivated crops, especially maize. It is not yet resolved whether this economic shift is the cause or the effect of greater population densities and more formally structured societies. Nevertheless, the appearance of farming hamlets and/or villages is well established. John White, who visited coastal North Carolina at the end of the sixteenth century, documented crops of maize in three stages of growth at one village (Rights 1957). Coe (1964:94) reports a stockaded village dating to post A.D. 1700 from archaeological excavations at the Gaston Site. This site, which is located in the northern interior coastal plain of North Carolina, also produced abundant Woodland artifacts, including woodworking tools (e.g., celts), smoking pipes, bone needles, fishhooks, and numerous pottery fragments. Data from the Gaston Site also provide a chronological framework for Woodland societies in the North Carolina coastal region. Coe (1964) dates the initial Gaston Site Woodland occupation at A.D. 500 (the Vincent Phase); the subsequent Clements phase is dated at A.D. 1200 and lasts until approximately A.D. 1600. The protohistoric/historic Gaston occupation appeared during the first half of the eighteenth century. Phelps (1978) and Ferguson and Widmer (1976) provide alternative Woodland chronologies for coastal North Carolina and South Carolina, respectively. The latter researchers further correlate this sequence with changing Woodland subsistence patterns. During the Early Woodland subperiod (ca. 1000 B.C. - 300 B.C.) incipient horticultural practices began, but hunting and gathering activities were still emphasized. During the Middle Woodland (ca. 300 B.C. - A.D. 1000) horticulture more effectively complemented hunting and gathering efforts. Fully developed horticulture characterized the Late Woodland (A.D. 1000 - European contact), although protein was still obtained from hunting and fishing.

The Woodland phases of North Carolina are most commonly recognized by variations in ceramic artifacts. These variations are frequently stylistic, although technological aspects such as tempering agents and firing techniques can also serve as temporal indicators. South (1976), for example, has established the following chronological pottery sequence for the North Carolina coastal plain: Hanover Sherd Tempered Series; Cape Fear Sand Tempered Series; Oak Island Shell Tempered Series; Tooled Interiors Series; Sand Tempered Plain Series; and Historic Brunswick Series. Coe (1952, 1964), Loftfield (1976), Haag (1958), Binford (1964) and Phelps (1980) have established comparable though slightly different sequences. Woodland chronologies can also be derived from a continuum of triangular projectile points, which show a general decrease in size as the Woodland period progresses.

Mississippian

The final major stage of cultural evolution in the eastern woodlands is the Mississippian. The chiefdom and state-level sociopolitical systems of the Mississippian period appeared in North Carolina among the mountain Cherokee (see Dickens 1976; Keel 1976), and minimally in the Piedmont in Richmond and

Montgomery counties during the sixteenth and seventeenth centuries (Coe 1952). The magnitude and nature of impact by these societies upon indigenous populations have not been adequately studied. However, the Piedmont Pee Dee cultural complex produced certain diagnostic artifacts (particularly pottery with distinctive design motifs) which have been encountered throughout coastal North Carolina.

CHAPTER THREE

INITIAL PREDICTIVE MODEL DEVELOPMENT

Background

The approaches to predictive modeling employed by archaeologists can be roughly divided into three categories: (1) descriptive approaches, (2) behavioral approaches, and (3) statistical approaches. Each has specific data requirements, and each has distinctive advantages and disadvantages.

Descriptive Approaches

Descriptive predictive models consist of summaries of previously collected archaeological data, and indicate which areas, or kinds of areas, have produced archaeological materials. Most commonly, all known sites within an area are located on topographic maps, and the geographic or topographic features exhibited by all or most of these site localities are identified. Other localities exhibiting the same topographic or geographic characteristics are then identified as having a high probability of producing archaeological remains. A qualitative definition of high probability areas is thus provided. A second and more sophisticated approach to descriptive modeling is based on quantitative, rather than qualitative, parameter definitions. For example, descriptive statistics may be used to identify those types of localities which have produced the majority of archaeological remains, and to define quantitative values for the geographic variables that delineate such areas.

The advantage of descriptive models lie in their flexibility and simplicity. The only data required in their formulation are information on which types of localities have produced archaeological materials. Descriptive models can also be tailored to meet the requirements of any type of data. When only anecdotal evidence concerning the locations of sites is available, a simple narrative description summarizing these data can easily be formulated. On the other hand, if a large and representative body of site location data exists, a more sophisticated model employing descriptive statistics can be developed. Because they rely on extant site location data, and because they are relatively simple to formulate, many archaeological predictive models fall into the descriptive category.

Regardless of their complexity, however, all descriptive predictive models have several weaknesses. First, since they rely largely on qualitative data analyses and summaries, it it difficult to weight variables according to their predictive power. For example, evidence may indicate that soil type is a better predictor of archaeological materials than slope direction, but that both had an influence on prehistoric settlement choices. Since both of these variables are qualitative, however, it is difficult to analytically weight them. For similar reasons, variable interactions, like variable weightings, are difficult to define. The predictive precision of descriptive models may be limited by these characteristics.

Another weakness in descriptive modeling procedures derives from the nature of the data upon which they are often based. In many areas, existing site location data have been collected from a combination of sources, including collector interviews and conventional surface surveys. In general, these techniques tend to reveal only those sites which are either especially prominent, or which occur in areas with exposed ground surfaces, such as cultivated fields. Under these circumstances site location data are likely to be biased, since these areas often do not constitute a representative sample of the region within which they occur. As a result, predictive models based on such data need to be validated through additional field testing.

Behavioral Approaches

The predictive models that fall into this category are based on ecological and economic reconstructions of prehistoric lifeways. If the interactions between a cultural system and its environment can be reconstructed, the investigator can specify which microenvironmental zones within that environment were exploited, and for what purposes. The areas which contain these zones can then be designated as having a high probability of producing specified types of archaeological remains.

The primary advantage of models of this type lies in their behavioral content. The goals of archaeology as a science involve reconstructing and explaining past human behavior; behavioral models make a direct contribution to achieving these goals. To be successful, however, behavioral models must be based on high-quality archaeological, ethnographic, and ecological information. A detailed model of this type can only be formulated at the completion of intensive, long-term research. These research efforts should, however, provide detailed and accurate predictions concerning the distribution of archaeological remains within an area of interest.

The primary disadvantage of the behavioral approach is the large research commitment that a comprehensive model would require. Although a generalized model can be formulated for most areas based simply on preliminary ethnographic research and on the distribution of a few key natural resources, the predictive discrimination of such a model will be low, and its accuracy questionable. General behavioral models of this kind are best used to interpret the site distributions revealed by other modeling procedures, and to extend predictions of models to specific categories of sites (e.g. small hunting camps) or to specific types of areas (e.g., those with heavy ground cover) for which direct data may be missing. Because they require detailed environmental and behavioral information not generally available, comprehensive behavioral models are not usually a feasible management tool.

Statistical Approaches

Statistical models consist of equations that express relationships among a specified set of variables. For example, a value for a dependent variable (such as the amount of archaeological material at a locality) may be predicted from values for a set of independent variables (such as the soil fertility, the distance from water, etc., of that same locality). Variable interactions and weightings are expressed in the equation(s) comprising the model, and thus are automatically included in the computational procedure. Furthermore, the statistical reliability of the predictions can generally be assessed from the variance of the sample data used to formulate the model. Finally, such models do not require detailed environmental and behavioral reconstructions. Because of these strengths, statistical models are generally more powerful than either behavioral or descriptive models; they should thus provide the most reliable and cost-effective approach to the problem of predicting the distributions of archaeological resources.

However, a key requirement of statistical models has limited their use for the purposes of archaeological predictive modeling. To formulate a statistical model of site distribution, a representative sample of localities from within the universe of all the localities for which predictions are desired must be available for analysis. All of the variables included in the model can then be measured for each of the localities in the sample. These variables can be correlated with the presence-absence or amount of archaeological material at the same localities, and these correlations can be used to construct a statistical model predicting the distribution of archaeological materials in the remaining, unsampled portion of the universe of interest. Unfortunately, this procedure cannot be conducted using conventional archaeological data. Such data are usually recorded and filed by sites, rather than by arbitrarily defined units of land surface such as square meters or square kilometers. These data provide information only for localities that contain archaeological material. It follows that these data do not constitute a representative sample of the total range of localities in the study area. Furthermore, it is the locality, not the site, which is the unit of interest for predictive modeling purposes. Since the unit of analysis in the data set (the archaeological site) is not the same as the unit of interest (a locality or a unit of land surface), the conventional data set is not appropriate to address the problem at hand.

Because of this difficulty, conventional archaeological data are generally unsuitable as a basis for formulating statistical models. Unless environmental data are recorded for all the units of land surface that are surveyed, rather than for only that subset of units containing archaeological material, predictive modeling for archaeological purposes must rely largely on descriptive models, as described above.

A Preliminary Predictive Model

The initial phase of the New Hanover County Predictive Models Project involved developing a preliminary predictive model. This model was designed to meet several requirements. First, to reflect current practice in the field of predictive modeling, the preliminary model was based on descriptive analyses of site location data. Second, to evaluate the adequacy of conventional site data for predictive modeling purposes, the preliminary model was derived from the site file data available for New Hanover County. Third, to insure replicability and to allow the use of computerized data manipulation, the model was formulated using descriptive statistics. As a result, this first model represented a statistical summary of the types of localities that have produced the majority of prehistoric archaeological sites within New Hanover County.

Data

Data for 463 archaeological sites in New Hanover County are on file at the Archaeology Branch, North Carolina Department of Cultural Resources. These data were recorded on North Carolina Prehistoric Site Forms and on magnetic tape to allow efficient computerized data manipulations. Information regarding the environmental setting of each site is recorded, and provides the basic data necessary to a predictive modeling exercise. The environmental variables suitable for inclusion in a preliminary predictive model included topographic situation, elevation, slope, slope face direction, distance to nearest water, type of nearest water, distance to and type of second nearest water, soil composition, and soil type. Time and cost limitations prevented the inclusion of all of these variables in the preliminary predictive model. Four variables were therefore selected which minimized information redundancy while maximizing probable predictive power. The four variables selected were elevation, distance to nearest water, type of nearest water, and soil type. The data used in the development of the preliminary model thus consisted of values for each of these four variables for each of the 463 sites in the data set.

The site location data available for New Hanover County were collected primarily during the CETA survey using conventional exposed ground surface survey techniques and thus were subject to the biases inherent in such data. In this respect, however, they were comparable to the majority of site location data sets presently available for other parts of North Carolina and for much of the Eastern United States. Developing and testing a preliminary predictive model using this type of data thus provided an opportunity to investigate the reliability of such data for predictive modeling purposes.

Analysis

The analytical procedure used to develop a preliminary predictive model for New Hanover County involved three basic steps. First, univariate distributions of all the sites in the data set were produced for each of the four variables included in the model. Second, zones having various probabilities of site occurrence were defined for each variable using the central tendency of its distribution as a guide. Third, probability zones (or areas) defined on the basis of all variables in combination were identified by locating the areas of intersection of the separate, single variable probability zones.

The first step in generating a preliminary predictive model thus focused on frequency distributions for each environmental variable included in the model. Four graphs were produced, each showing the distribution of archaeological sites by variable measure or value (e.g., Figure 3.1). This was accomplished using procedures available in the SPSS library of computer programs (Nie *et al.* 1975).

In the second analytical stage, these variable distributions were analyzed to identify zones having various probabilities of site occurrence. For the metric variables in the model (i.e., elevation and distance to nearest water) the standard deviation of the distribution was used. Specifically, multiples of the standard deviation were used to produce three probability zones. For any variable i, localities exhibiting values that fell between $+0.5s_i$ and $-0.5s_i$ (s_i = standard deviation for variable i) were designated as high probability areas; those between $+0.5s_i$ and $+1.5s_i$ and between $-0.5s_i$ and $-1.5s_i$ were designated as medium probability areas; and those localities exhibiting values

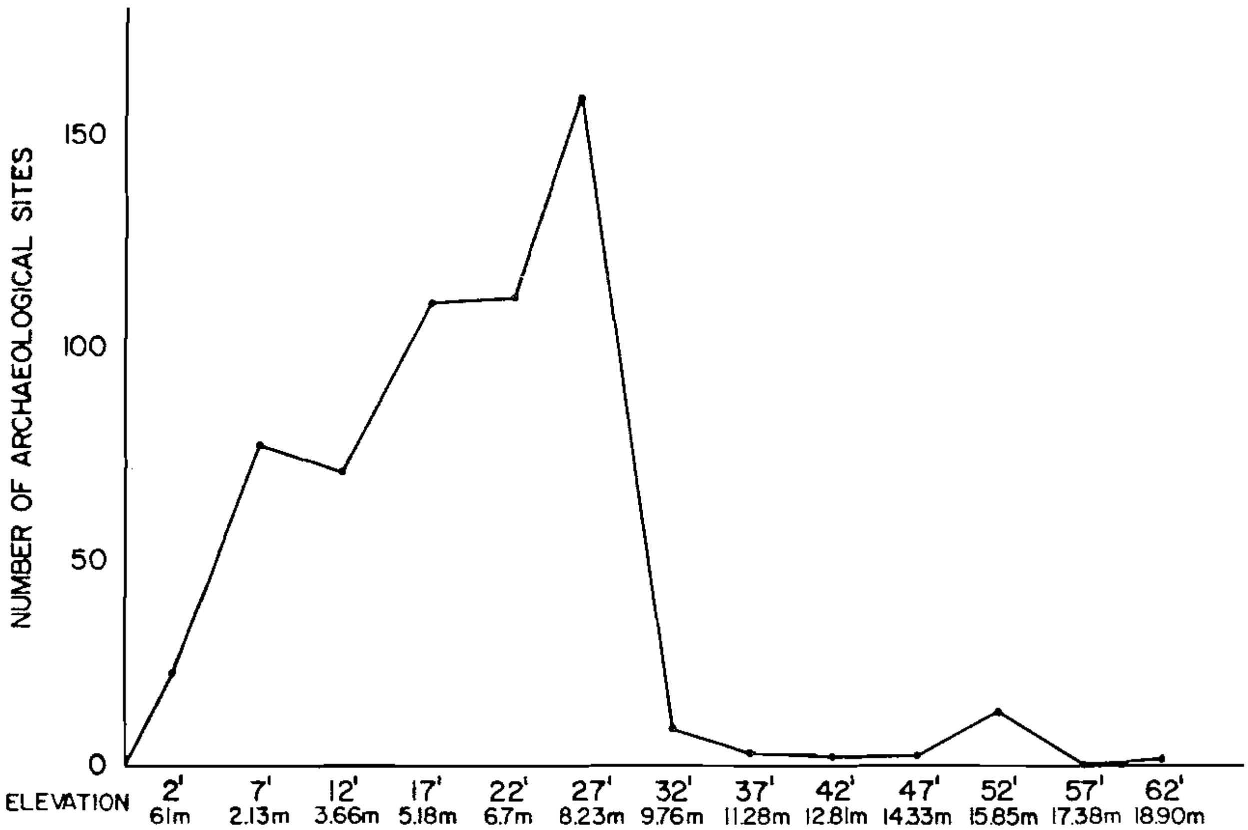


Figure 3.1 Distribution of archaeological sites by elevation

greater than $+1.5s_i$ and less than $-1.5s_i$ were designated as low probability areas.

As an illustration of this procedure, a normal curve approximation of the distribution of elevations for all components in the data set is shown in Figure 3.2. Using the cutoff points defined above, the high probability zone for this variable was defined as that portion of its distribution that fell between $+0.5s_e$ and $-0.5s_e$ (s_e = standard deviation for elevation). Translated into actual elevation). Translated into actual elevations, $+0.5s_e$ to $-0.5s_e$ represents an elevation range of 4 meters to 7 meters (13 to 23 feet) above mean sea level (MSL). Thirty-eight percent of the archaeological sites in the data set fell within this zone; a larger percentage than is contained by any other equivalent elevation range within the county. Medium probability areas for this same distribution were defined as those that exhibit values that are $+0.5s_e$ to $+1.5s_e$ and $-0.5s_e$ to $-1.5s_e$ from the mean. The actual elevations that correspond to these cutoff points are 1.2 meters to 4 meters (4 to 13 feet) and 7 meters to 10 meters (23 to 33 feet) above (MSL). Each of these zones contains 24% of the sites in the data set. Low probability areas were represented by the tails of the distribution, which for the purposes of the present analysis were defined as being greater than $+1.5s_e$ (10 meters or 33 feet above MSL) or less than $-1.5s_e$ (1.2 meters or 4 feet above MSL). Only 12% of the sites in the data set occur in these zone ranges. Using these procedures, high probability, medium probability, and low probability zones were defined for both elevation and distance to water.

Two of the four variables included in the model (i.e., soil type and type of nearest water) were non-metric. Since standard deviations can only be computed for metric measures, the distributions of these non-metric variables could not be subdivided into probability zones in the same manner. A different procedure for defining cutoff points for these non-metric variables was thus required. Since the high probability zone for metric variables ($\pm 0.5s_i$) includes approximately 38% of the normal distribution, an equivalent zone for a non-metric variable would include a subset of the values for that variable which accounted for approximately 38% of the total distribution of the variable. To represent the central tendency of the distribution, these values would have to consist of that subset with larger numbers of cases than any other subset of values.

To identify the portion of a non-metric distribution meeting these two criteria, bar graphs were constructed for each non-metric variable (i.e., soil type and type of nearest water). On each graph, variable values were arranged in descending order, as defined by the number of sites for each variable value. Those values which 1) exhibited the largest numbers of cases, and 2) together accounted for approximately 38% of the distribution in question were then identified as representing the high probability zone. Medium and Low probability zones were identified in a similar manner (see Figure 3.3).

A final problem in this step of preliminary predictive model development was posed by the absence of data for certain variable values. For example, certain soil types that are present in the county produced no archaeological sites. These soil types did not appear on the distributions used to define high, medium, and low probability zones. Therefore, they could not be assigned to any one zone. To produce as complete a model as possible, values with missing

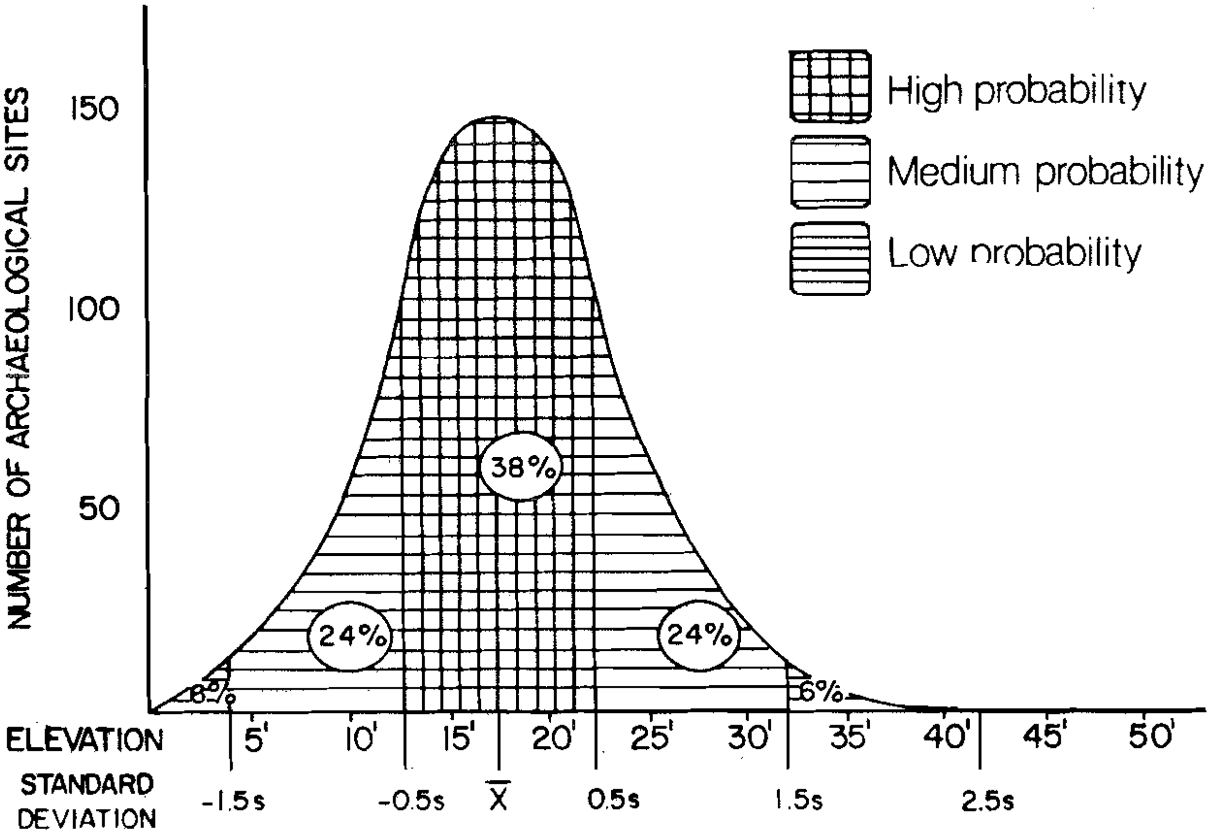


Figure 3.2 Normal curve approximation of the distribution of archaeological sites by elevation

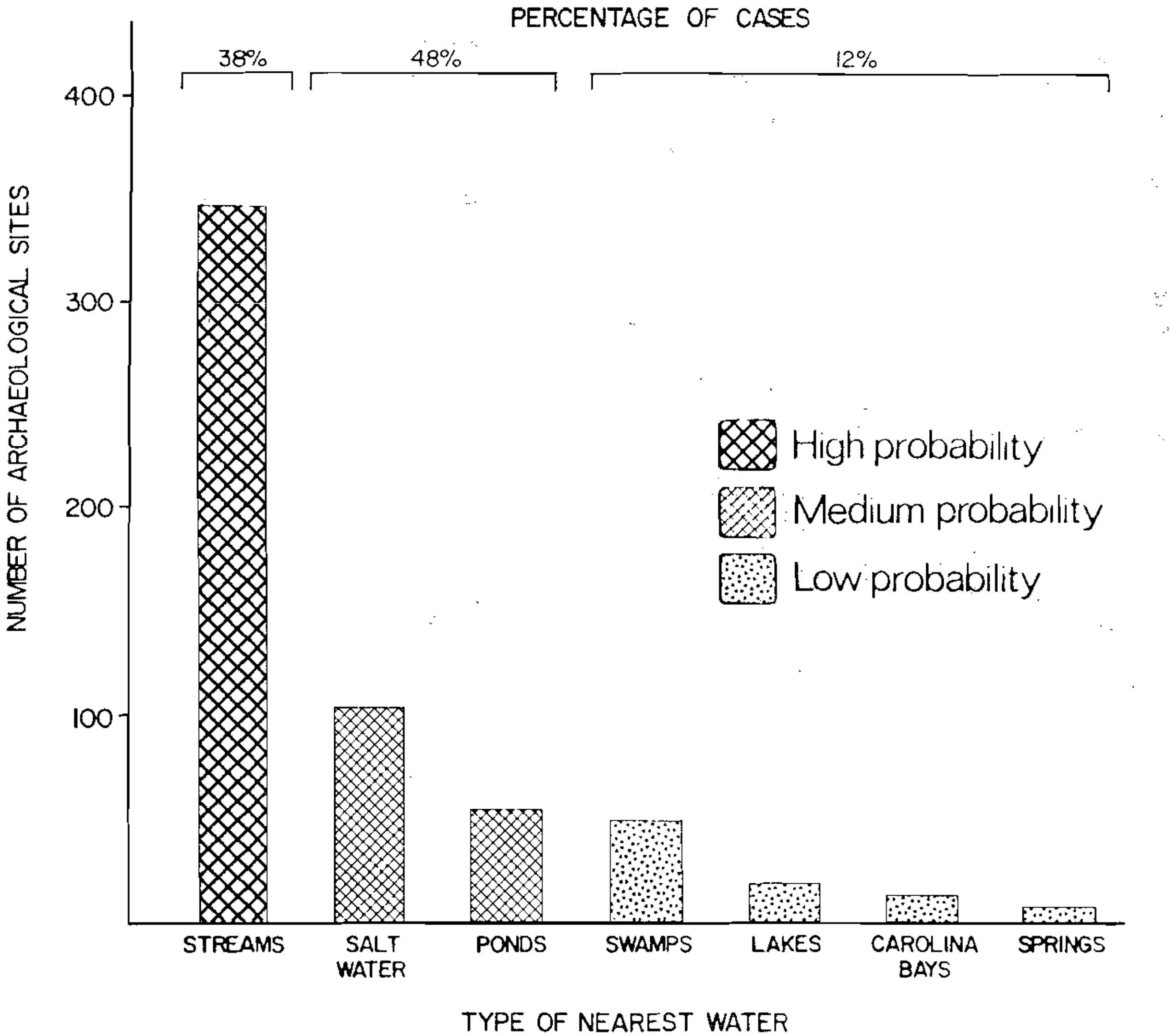


Figure 3.3 The distribution of archaeological sites for various types of nearest water

data were assigned to a fourth probability zone. Since the areas included within this zone had produced no archaeological materials, they were termed nonprobability areas.

Completion of this second phase of model development resulted in the definition of probability areas for each of the four environmental variables included in the model. These probability area definitions are presented in Table 3.1 and can be considered a preliminary predictive model for New Hanover County. Before this model could be tested in the field, however, probability zones defined by the interaction of all four variables in combination had to be delineated.

Initially, the mapping capabilities of a computer graphics system, LRIS (the Land Resources Information Service, North Carolina Department of Natural Resources and Community Development) were employed to define the areas of probability zone intersection. These procedures are described in Appendix C. This system is designed to store and manipulate geographic data and can produce maps of any stored data set to any specified scale. LRIS can also produce maps of combined data sets in overlay fashion.

Unfortunately, these attractive system characteristics were partially negated by a series of data processing time overruns. As a result, the preliminary predictive model was mapped in part by hand and in part by the LRIS computer facility. One unfortunate byproduct of these difficulties involved the partial loss of soil map data; as a result, soils data were available for only 17 of the 30 soils maps for New Hanover County (cf. Weaver 1977). The preliminary model was therefore mapped for the area covered by these 17 maps. However, the coverage provided by these maps was considered adequate for a field test of the preliminary predictive model (see Figure 3.4). With the completion of this final phase of preliminary model development, a field test of the preliminary model was possible.

Table 3.1 A descriptive predictive model for New Hanover County

Probability	*Soil Type	Type of Nearest Water	Elevation Meters (Feet)	Distance to Water (Meters)
High	LA, NH, CR	Streams	4-7 (13-23)	98-389
Medium	KE, ST, LE, RM BE, LY KR, WA	Saltwater Ponds	1.2-4 (4-13) 7-10 (23-33)	0-98 389-680
Low	DO, NO JO, MU SE, LS TM, WR ON, WO PN, TO	Springs Lakes Swamps	0-1.2 (0-4) 10-12.8 (33-42)	680-971
Non	BA, PM, RA	Sloughs Other	12.8 (42)	971

* See Table 2.1 for soil series abbreviations.

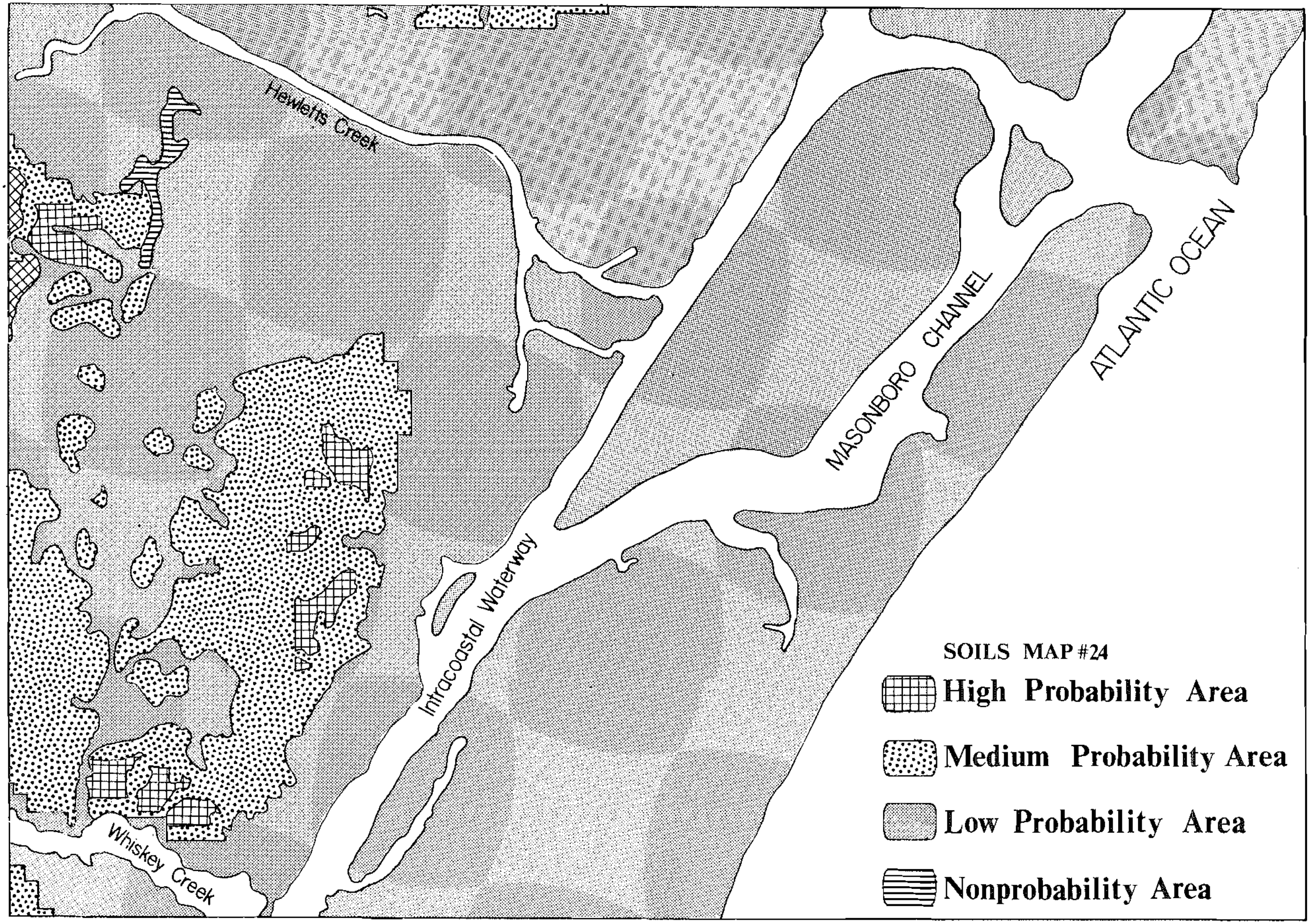


Figure 3.4 High, medium, low and nonprobability zones for soil map #24, New Hanover County (cf. Weaver 1977)

CHAPTER FOUR

MODEL EVALUATION

Field Survey

Survey Design

The preliminary predictive model developed in the previous section can be considered a statistical summary of the extant site location data available for New Hanover County. As such, its accuracy depends on two factors, the reliability of the data upon which it was based and the adequacy of the analytical procedures used in its formulation. To assess either of these factors, however, an entirely independent body of site location data for New Hanover County was required. Furthermore, these new data had to accurately reflect the true distribution of archaeological materials in the county, i.e., they had to constitute a representative sample of the population of archaeological sites in the study area. Only then could the predictions of the model be compared to the estimated actual distribution of archaeological materials, and the accuracy of those predictions assessed.

Before an appropriate body of data could be provided to test the model, however, two methodological issues had to be addressed. The first involved the sampling procedures used to generate a representative sample of sites from the population of sites within New Hanover County. The second involved the choice of the field methods that would be used to collect the data in question.

Sampling - Theoretical Considerations

Despite New Hanover County's rank as second smallest in areal extent for North Carolina counties it was obvious that a field test of the predictive model could involve only a fraction of its land surface. To insure that this fraction would constitute a representative sample of the county, probabilistic sampling techniques were employed.

Four general categories of probabilistic sampling strategies are commonly used by archaeologists: simple random sampling, stratified sampling, cluster sampling, and systematic sampling. For the purposes of the present study, a multistage hierarchical sampling design was developed, incorporating all four of these sampling strategies.

At the most general sampling level, the county was subdivided into four zones. These zones, or strata, corresponded to the four probability zones produced by the predictive model. This stratification of the county was implemented for several reasons. First, stratification insured that data would be collected from each probability zone. Second, stratified sampling groups homogeneous data in order to reduce variability, hence enhancing the reliability of statistical estimates. Third, a prior designation of subpopulations improves administrative convenience, such as in the revision of field sampling strategies.

A second sampling stage defined the population, or frame, of sampling units. These units were defined by the nature of archaeological predictive modeling--a unit of land surface is measured for the presence-absence or amount of archaeological material. Practical considerations of field survey (discussed in the next section) suggested that a unit measuring 200 meters x 200 meters (4 hectares) was an appropriate survey unit size. A grid of these units was therefore superimposed over the New Hanover County probability zone strata. Within each of the strata, a random sample of grid units was selected for actual field inspection. Three factors - budgetary constraints, rate of field survey coverage, and inequality of stratum size - dictated the number of survey units selected within each stratum. As detailed below, these three concerns suggested that disproportionate random sampling was most appropriate for the selection of grid survey units within strata. In other words, the percentage of survey coverage in each stratum would differ.

The extensive vegetation cover over much of New Hanover County prevented complete survey of the grid units selected in each of the four probability zones, or strata. Additional sampling strategies were therefore necessary to recover data within any given grid unit. This problem was addressed by considering each survey (grid) unit to be a collection, or cluster, of discrete sampling points (i.e., shovel tests). Effective probabilistic sampling within a survey unit thus depended upon selection of a subset of these sampling points. As discussed below, practical considerations of field survey indicated that a systematic arrangement of individual sampling points (shovel tests) within each cluster (survey unit) constituted an adequate sampling strategy. As such, the final analytic sampling units defined for this archaeological survey of New Hanover County were systematically arranged points (or shovel tests), clustered into grid survey units. These units were in turn randomly selected (with disproportionate weighting) from each of four strata, which represented discrete zones of predicted archaeological site density.

Sampling - Practical Considerations

Several decisions were required to implement these sampling schemes. First, an appropriate overall survey area size had to be determined. It was estimated that approximately 2 hectares (5 acres) per person per day was a reasonable coverage estimate, assuming that shovel testing was the major survey technique employed. Project cost constraints dictated that the field crew would consist of five persons, and that the field season would cover a four-week period. From these figures, it was estimated that the total survey area would consist of 160 hectares (400 acres). It was then necessary to determine the optimum size for each survey unit, or cluster. For samples of the same overall size, those composed of a large number of small clusters (or survey units) are generally more statistically efficient than those composed of a small number of large clusters (or survey units). Statistical considerations thus favor small, numerous clusters (Cochran 1977; Blalock 1972). Practical considerations, however, have the reverse effect. As clusters (survey units) increase in number and decrease in size, the costs of moving from cluster to cluster and of locating clusters in the field increases, reducing overall survey coverage. To determine the optimum size and number of clusters (survey units) one must seek a balance between these conflicting requirements of statistical efficiency and practical efficiency. Since the total acreage to be surveyed was small relative to the overall size of New Hanover County, it was anticipated that survey units (or clusters) would be widely scattered throughout the county and that the time consumed by moving from unit to unit would be considerable. In view of this, it was decided that one such move per day on the average was acceptable, but that more moves would seriously reduce survey coverage. An average coverage rate of two survey units (or clusters) per day was therefore selected, and a total of 40 units (or clusters) was to be surveyed during the 20 days allocated for fieldwork. Dividing the total number of hectares to be surveyed by the total number of units to be surveyed gave a size of 4 hectares (approximately 10 acres) for each survey unit. A 200m x 200m grid unit was therefore selected as the survey unit size.

Finally, the number of survey units (or clusters) to be drawn from each probability zone (or strata) had to be determined. Two standard options are available for this procedure; to select a sample from each stratum that is proportional in size to the size of that stratum, or to select equal sized samples from each strata regardless of variations in strata size. Because of the nature of the preliminary predictive model, the four probability zones varied enormously in areal extent. High probability zones were quite limited in area, while low and nonprobability zones were large and widely distributed. This variation precluded a proportional sampling scheme. The 160 hectares (400 acres) to be surveyed during the field test constituted 0.3% of that portion of New Hanover County available for study. With proportional sampling, this fraction of each probability zone would have to be surveyed. Since approximately 8,932 hectares (22,070 acres) of the study area fell into the nonprobability zone, a total of 27 hectares (66 acres), or approximately 7 units of nonprobability surface would have to be surveyed using a proportional sampling scheme. In like manner, 100 hectares (247 acres) or 25 units of low probability, 16 hectares (40 acres) or 4 units of medium probability, and 0.8 hectares (2 acres) or 0.2 units of high probability would be surveyed. These figures were clearly unacceptable, since a survey of only 0.8 hectares (2 acres) of high probability land surface could hardly be expected to produce a reliable picture of the distribution of the archaeological resources within that zone. Equal sampling of all strata was thus a preferable approach, and involved the random selection of 10 survey units, or 40 hectares (100 acres), from within each of the four probability zones.

Archaeologists often encounter problems when implementing probabilistic sampling schemes, since randomly selected units may not be accessible for on-the-ground inspection. Land development and landowner permission are among the key factors that cause such problems. In order to avoid these difficulties in the present study, a second sample of 40 units was drawn in the same fashion as the first sample. When field access to one of the units in the first (primary) sample was restricted, an alternate unit was randomly selected from the second sample to replace it. In this fashion, the probabilistic nature of the final sample was not violated.

Field Methods

More than 60% of the land surface presently available for archaeological survey in New Hanover County is wooded (Weaver 1977). Since the field test phase of the present study was based on a probabilistic sample, a high percentage of the survey would therefore be performed in wooded areas. A survey methodology appropriate for such areas was thus required.

Locating archaeological remains in areas with heavy ground cover is a key problem for archaeologists working in the Eastern Woodlands. Numerous strategies for locating sites in such areas have been used: augering, shovel testing, and surface clearing are some of the more common techniques. Augering involves sinking deep, narrow holes with a soil auger or posthole digger. The soil removed is examined to determine the presence or absence of artifacts or soil anomalies, and may or may not be screened. In general, this method is used in areas where deeply buried remains are suspected. Shovel testing involves excavating relatively small, shallow holes and examining the soil for archaeological remains, either by screening or by simple visual examination. Surface clearing involves removing the leaf litter in wooded areas to expose patches of ground surface, which are visually examined for evidence of archaeological material.

The purpose of the present field survey was to recover a representative sample of archaeological sites from each probability zone, thus providing adequate data with which to test the predictive model. Various survey methods were evaluated relative to this goal. Shovel tests were selected rather than surface clearing tests or auger tests, because they can provide larger volumes of earth for examination at each testing point. In general, only large artifacts are revealed by simple visual inspection of the soil from a test; therefore, screened shovel tests (¼" mesh) were selected. Relatively large shovel test dimensions (50cm x 50cm x 50cm) were selected in order to maximize the probability of finding archaeological material within each test. Finally, a test point interval that was sufficiently small to intersect small sites as well as large ones was considered desirable. Given the small size of prehistoric hunting camps of the Eastern Woodlands, test points were spaced at 30 meter intervals. This resulted in a total of 49 tests per 200m x 200m survey unit. To define site boundaries, shovel tests were to be excavated at 5 meter intervals starting at each find point and proceeding in the 4 cardinal directions until 2 consecutive sterile tests occurred.

Survey Implementation

In order to use the sampling strategies and field methods that had been selected for the survey phase of the project, 49 shovel tests had to be located within each of 40 survey units. This process involved three steps: 1) locating survey units on the ground, 2) locating 49 test points within each survey unit, and 3) excavating a shovel test at each such testing point.

Survey units were located by first transferring their boundaries from the probability zone maps to high resolution (1 inch = 1000 feet) aerial photographs. An easily recognizable locality in the immediate vicinity of a survey unit was then selected, and compass bearings and distances from that point to the nearest boundary of the survey unit were calculated. These measurements were then laid off in the field, and a field crew thus arrived at a known point on the boundary of a survey unit.

Using this point as a reference, a baseline consisting of 7 test points located at 30 meter intervals was laid out along one boundary of the survey unit. A shovel test was excavated at each of these 7 baseline test points. Field crews then moved through the survey unit, starting at each baseline test point and proceeding in a direction perpendicular to the baseline. Each crew excavated a shovel test every 30 meters. After 6 tests had been completed, each crew began a new transect starting from another baseline test point. In this fashion, 49 shovel tests were located and excavated in a survey unit.

When establishing baselines and when moving from shovel test to shovel test, field crews used Brunton compasses to maintain their directional orientation. Thirty-meter lengths of non-stretch cord were used to measure distances between test points. These proved more useful than metal tapes, because the cord could be thrown over or through dense underbrush. In areas of heavy vegetation, these techniques were essential, since it was often impossible to maintain a compass orientation or estimate distances. In areas of open vegetation, these methods proved to be no more time consuming than less precise methods such as pacing the distances between shovel tests. Two member field crews were generally used during field survey; one crew member maintained compass orientation for the other who proceeded to the next testing point.

At each testing point, a 50cm x 50cm x 50cm shovel test was excavated. In general, one crew member dug while the other screened. Small, hand-held screens were used. Standard field recording techniques were employed; notes were taken concerning soil profiles, vegetation, topography, and the nature of any cultural material recovered. Soil type identifications were made using the descriptions of soil profiles in association with the general soil maps of the county (Weaver 1977). When appropriate, soil samples were collected. Dominant vegetation types, site locations, and archaeological features were photographed.

After all of the tests located at 30 meter intervals had been completed, the distribution of archaeological remains within the survey unit was reviewed, and appropriate procedures were taken to define site boundaries; shovel tests were located at 5 meter intervals between all pairs of adjacent sterile and nonsterile shovel tests. To locate the approximate boundaries of sites that extended beyond the limits of the survey unit, lines of shovel tests at 30 meter intervals were extended from the unit oriented in the four cardinal directions until a sterile test occurred.

Several problems with these procedures were encountered during the initial phases of the field survey. In sections of several units, impenetrable underbrush was encountered. The only effective way to move from testing point to testing point in these areas was to cut a path through the dense vegetation. However, this procedure was so time consuming it was incompatible with reasonable survey coverage rates. As a result, testing points were declared inaccessible when the underbrush was so dense that it required cutting a path.

Low survey coverage rates posed a more important problem. Originally, it was estimated that an average of 2 survey units per day would be completed. However, surveying the first unit required 3 full days of field work, or 6 times the original estimate. Although field crew efficiency was characteristically sub-optimal during the survey start-up, it was clear that rather radical changes in field methods and survey design would be required if an adequate test of the model was to be achieved. If the original design had been maintained, low coverage rates would have allowed a maximum of only 2 units per week to be surveyed. Total coverage would have equalled 8 units, or 2 units per probability zone. The results of the field test could thus have been heavily biased by the random selection of a single non-representative survey unit. To guard against this, it was necessary to survey a larger number of smaller survey units. Using the preliminary coverage rates achieved during the initial phases of the survey, it was estimated that approximately 40 hectares (100 acres) could be surveyed in four to six weeks. The following procedures were used to provide a sample of this magnitude which would also serve as an adequate test of the preliminary predictive model. First, a random sample of 5 survey units was selected from the original 10 units within each probability zone. The total sample of survey units was thus reduced from 40 units to 20 units; as a result, the survey units finally selected are not numbered consecutively (see Appendix D). Second, each survey unit was halved, producing a unit measuring 100m x 200m and consisting of approximately 2 hectares (5 acres). In combination, these alterations of the survey design reduced the total area to be surveyed from 160 hectares (400 acres) to 40 hectares (100 acres). At the same time, however, the probabilistic nature of the sample was not violated. Finally, to increase survey coverage rates, the interval between shovel tests used to define site boundaries was increased initially from 5 meters to 10 meters, and finally to 15 meters.

Several additional modifications of the survey design would have further improved coverage rates but were not implemented. It is likely, for example, that test point interval could have been increased and shovel test size decreased to some extent. A lower recovery rate of small, sparse sites would have resulted; however, this loss was offset by the greater number of larger and denser sites that would have been discovered with the resulting increase in overall survey coverage. It was nonetheless feared that these changes in survey methodology would result in two sets of non-comparable data: one produced by the initial methods, the other by the revised methods. Since using both data sets in a single test of the predictive model would have constituted a complex analytical procedure, these revisions were not implemented.

A final modification which would have improved survey coverage rates was precluded by available equipment and personnel. Considerations of field logistics provided an opportunity to assess the relative effectiveness of field crews of varying size, since crews of 4, 3, 2, and 1 surveyors each were used at different times during the project. As Figure 4.1 indicates, one-member field crews proved substantially more efficient than larger crews. Dividing the original two-member crews into one-member crews would thus have improved coverage rates considerably. However, working in one-member crews was

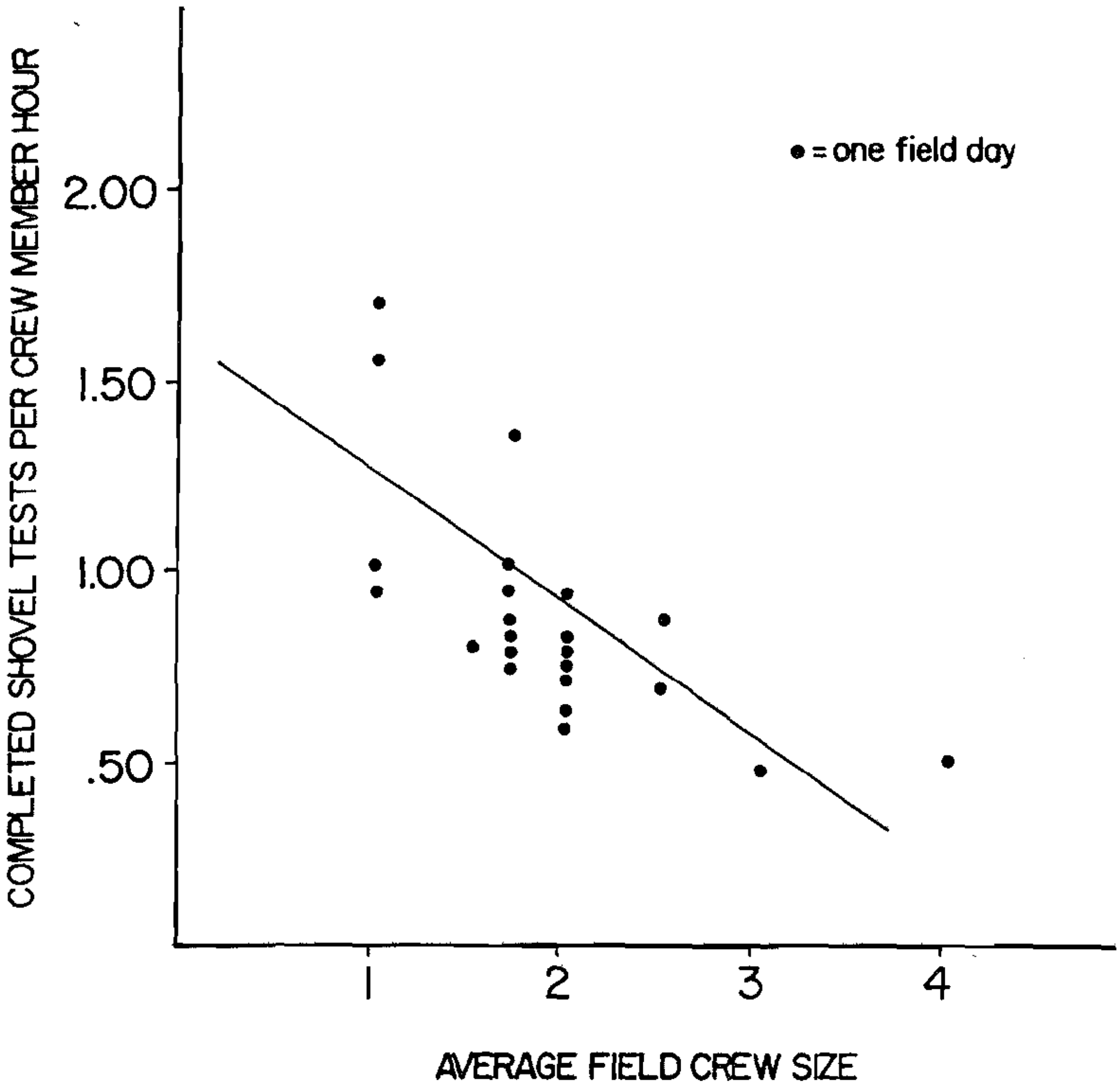


Figure 4.1 Survey effectiveness, expressed as completed shovel tests per crew member per hour, for various crew sizes

strenuous, and the costs of the additional equipment required by 6 crews as opposed to 3 crews were significantly higher. As a result, one-member crews were used only when insufficient personnel was available for 3 crews of two people each.

To summarize, the revised survey methodology involved surveying each of 20 randomly selected 2 hectare (5 acre) units, or 5 units from each probability zone (Figure 4.2). Each of these units required the excavation of at least 28 shovel tests with dimensions of 50cm x 50cm x 50cm at 30 meter intervals. If archaeological material was discovered in any subset of these original 28 shovel tests, additional test were excavated to determine site boundaries. Within survey units these additional tests were placed at 15 meter intervals; outside units, tests were placed at 30 meter intervals. With these modifications, a maximum coverage rate of approximately 1 hectare (2.5 acres) per person per day was achieved. The average coverage rate for the project was approximately 0.6 hectare (1.5 acres) per person per day.

Survey Results

Revisions in the survey design implemented during the course of the survey produced an initial survey unit measuring 200m x 200m, while the remaining units all measured 200m x 100m. Five survey units each were located in high, medium, and nonprobability areas. Since the first unit, which was twice the size of all the remaining units, was located in a low probability area, only 3 additional units were placed within the low probability zone. Approximately equal survey coverage in each probability zone was thus maintained. A total of 19 survey units was consequently examined during the course of this survey.

Seven-hundred and three (703) testing points were located within these 19 units. At 133 (20%) of these, shovel tests were not excavated because they were inaccessible and/or were located in standing water. Of the 570 tests that were excavated, 134 (24%) produced archaeological material. These totals include the tests that were located within survey units at 30 meter intervals (base grid tests). They also include the tests which were used to identify site boundaries both inside and outside the confines of the survey units. A total of 553 base grid tests were located, 406 (73%) of these were actually excavated, and 63 (16%) produced artifactual material.

A total of 24 archaeological sites were identified during the course of the survey. When conventional surface survey methods are employed, sites are generally defined as discernable scatters of artifacts or as areas within such a scatter which exhibit especially high artifact concentrations. In areas with heavy ground cover, however, a continuous surface is not available for visual inspection, and artifact concentrations cannot be directly discerned. Instead, they must be inferred from subsurface test data. The present study defined an archaeological site as an area containing a subsurface test which produced prehistoric artifactual material, or as any set of such tests adjacent to one another.

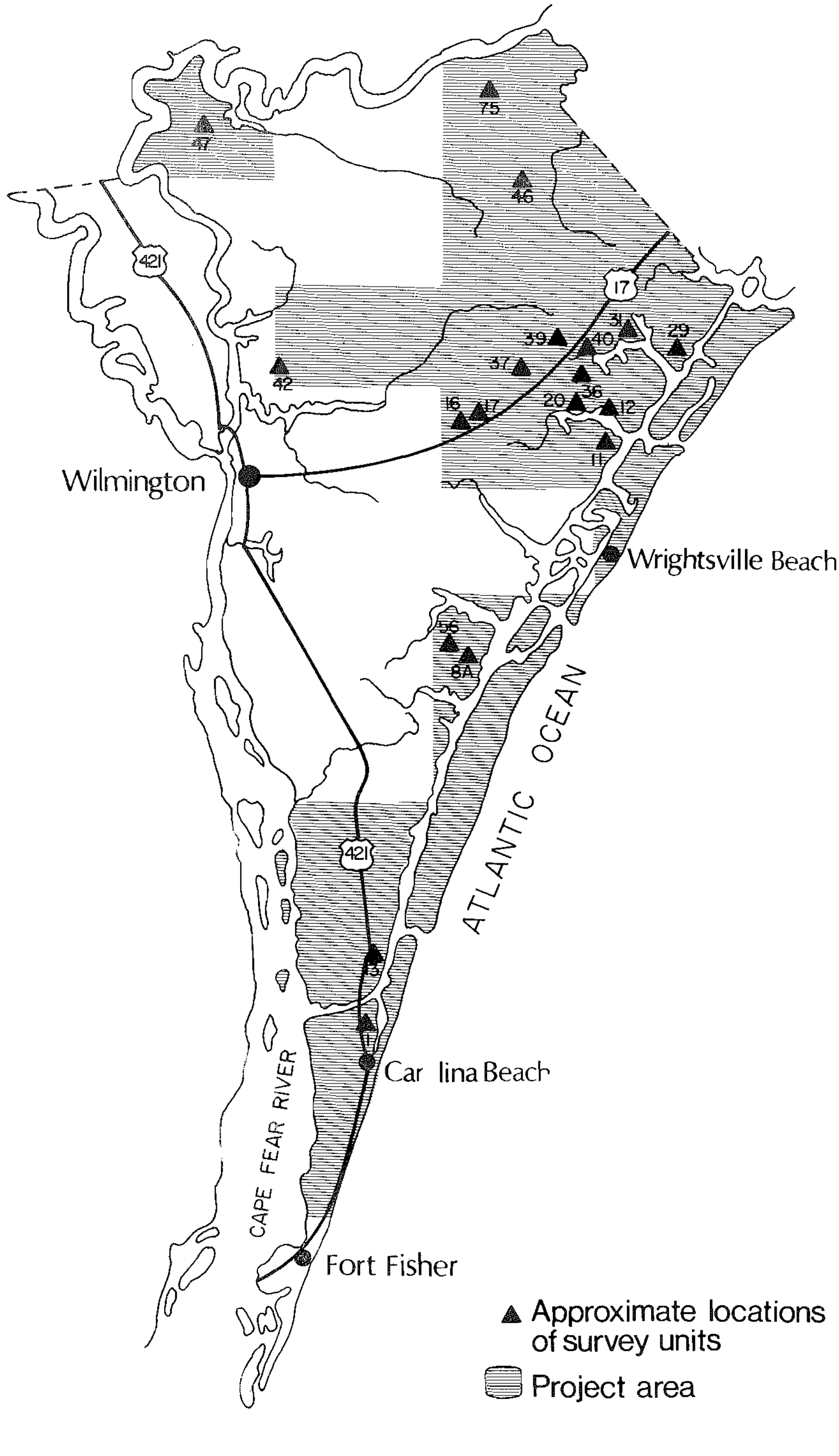


Figure 4.2 Survey unit locations, New Hanover County, North Carolina

From the survey data, an overall site density of 0.75 sites per hectare (0.30 sites/acre) for surveyed areas can be computed (Table 4.1). This estimate cannot be taken as the true site recovery rate of the survey, however, because the survey was based on a weighted sampling design. Approximately equal amounts of each probability zone were surveyed, but since the high and medium probability zones were far less extensive than were the low and non-probability zones, the sampling fractions for the former areas were considerably higher. As a result, they have contributed disproportionately to the site recovery rate. When this effect is corrected, an overall site recovery rate of 0.55 sites per hectare (0.22 sites per acre) can be estimated. This can be taken as the recovery rate that would have been achieved had each probability zone been sampled proportionally (Table 4.1).

The 24 sites discovered during the survey range in area from approximately 150 square meters to 10,000 square meters, and exhibit artifact densities of 8 cubic meters to 64 cubic meters. Two of these sites are represented by single artifact discoveries in isolated shovel tests. The remaining sites produced more extensive artifactual material, although the survey techniques employed generally produced relatively small artifact samples (Appendices E, F and G).

The prehistoric artifactual materials recovered consisted entirely of lithic and ceramic artifacts. Lithic materials were relatively rare, and all of those recovered in shovel tests were temporally nondiagnostic flakes, cores, or preforms (Appendix G). One projectile point, probably Morrow Mountain type, was recovered from the surface of Unit 36. Ceramic artifacts were relatively abundant (Appendix F). Fabric impressing and cord marking were by far the most common surface decorations; together they account for 93% of all sherds with non-eroded surfaces. Fabric impressing (65%) was more common than cord marking (28%). Sherd-tempered (37%), sand-tempered (58%), and shell-tempered sherds (5%) were identified.

Although sample sizes were low, features were discovered at 5 sites. A postmold was discovered at one site (31NH605), and dark, midden-like lenses were observed in shovel tests at four sites (31NH610, 31NH611, 31NH612 and 31NH613)

The relatively limited scale of the present survey, and the small size of the artifactual samples recovered, limit the culture-historical and behavioral interpretations that can be derived. However, 88% of all sites with prehistoric ceramics produced sherds with at least two out of the three varieties of tempering material that occur in the overall sample (Appendix F). If South's (1976) arguments concerning the chronological sensitivity of tempering materials in the coastal areas of the Carolinas are correct, the majority of the sites recovered during the present survey must thus have several Woodland period components. This, in conjunction with the relatively high density of archaeological sites revealed by the survey, suggests that the coastal areas of southern North Carolina were intensively utilized throughout most of the Woodland period. During the field survey, an especially high frequency of archaeological sites was noted on minor knolls and ridges. In an area that is largely flat and poorly drained, every slight elevation seems to have been inhabited

Table 4.1 Site recoveries per hectare (acre).

Prob. Zone	Total Hectares (Acres) in Sampling Universe	Hectares (Acres) Surveyed	Sampling Fraction	No. Sites Disc.	Est. Total # of Sites in Zone
High	73 (180)	9.5 (23.50)	.131	11	84
Medium	1891 (4670)	9.0 (22.25)	.005	8	1679
Low	10,259 (25,340)	8.0 (19.50)	.001	5	6497
Non	2793 (6900)	5.6 (13.75)	.002	0	0
			Estimated Total Sites		8260

prehistorically. A settlement system composed of small, temporary villages occupied by swidden horticulturalists would produce such a pattern. Short-term village occupations would produce a large number of sites over a period of 2,000 years. Virtually every suitable locality would likely be selected as a village site sooner or later, and many would see repeated occupations. Such a system may thus have characterized the coastal portions of southern North Carolina throughout much of the Woodland period.

Statistical Assessment of the Preliminary Model

As discussed above, the preliminary predictive model was developed from existing site location data for New Hanover County, with these data being recorded in the conventional archaeological format, i.e., the site was the recording unit. The procedures used to generate the predictive model were designed to provide a statistical summary of these data. Four topographic variables were selected, and parameter values bracketing a zone containing the majority of the archaeological sites in the data set were determined for each of these four variables. The areas of intersection among these zones were then identified, and high, medium, low, and nonprobability areas of site occurrence were defined and mapped. Descriptive statistics and quantitative parameter definitions thus provided a systematic description of the types of localities within New Hanover County that had yielded the majority of the previously discovered archaeological materials.

To be an effective review and planning tool, a predictive model of this kind must identify localities that contain especially high densities of archaeological sites. Assuming that the predictions of the model are correct, a representative sample of localities from within the area of interest should thus document a non-random association between the distribution of archaeological materials and the probability zones defined in the model. Specifically, the present model predicts that high probability zones should contain more sites than medium probability areas, medium areas more than low, and so forth. Using the data collected during the field survey phase of the project, the accuracy of the preliminary predictive model was assessed by testing this prediction.

Data

As indicated above, a stratified, clustered, systematic sample of 406 base grid shovel tests was provided by the survey phase of the project. Sixty-three of these tests produced archaeological material. These tests provided the data used to evaluate the preliminary model. However, sampling schemes vary in efficiency, and statistical procedures must take these variations in efficiency into account. For example, a standard chi-square test conducted using data derived from a clustered sample may give erroneous results, because the chi-square formula assumes that the more efficient, simple random sampling process has been used (Blalock 1972). The sampling design of the present survey was hierarchical and used different sampling strategies at different levels of the hierarchical design. Before a statistical test of the model could be conducted, the effects of these sampling procedures had to be evaluated.

The sampling procedures used in the design of the present survey included stratifying the universe of interest, simple random sampling within the strata, clustering the shovel test unit of analysis, and systematic sampling within these clusters. No correction factors need to be introduced for simple random sampling, since formulas for computing inferential statistics generally assume this sampling strategy. However, the remaining procedures (i.e., stratification, clustering, and systematic sampling) frequently differ in statistical efficiency from simple random sampling. The effects of these latter procedures must therefore be assessed.

Formulas for computing inferential statistics are generally not appropriate for stratified samples (Cochran 1977). If well designed, however, stratified samples are almost always more efficient than simple random samples (Cochran 1977); as a result, hypothesis tests using simple random sample formulas should generally be more conservative than those conducted with the correct formulas. Since stratified sampling formulas that met the requirements of the present study have not yet been developed (Blalock 1972), it was necessary to use simple random sampling formulas in their place. The levels of statistical significance indicated by these tests should therefore be considered conservative estimates of true significance levels. They provide approximate, rather than exact indications of the levels at which hypotheses should be rejected.

In contrast to stratified samples, cluster samples are less efficient than random samples of the same size. As a result, simple random sample formulas do not provide approximations of the correct rejection levels, and corrective measures must be utilized. An estimate of the relative efficiencies of a cluster sample and a simple random sample of the same size is given by the ratio of sample variances, or

$$\frac{V_{bin(p)}}{V_p} = \frac{N\bar{M}PQ}{\sum_{i=1}^{N} M_i^2 (p_i - P)^2} \qquad \text{(Eq. 4.1)}$$

where:

- V_p = the variance of p (in this case, the proportion of culturally nonsterile shovel tests in a survey unit) when cluster sampling is used
- $V_{bin(p)}$ = the variance of p assuming simple random sampling
- M = the number of elements in a cluster (in this case, excavated shovel tests)
- N = the number of clusters (in this case, survey units in the sample)
- P = the mean of p (in this case, the mean proportion of culturally nonsterile shovel tests per survey unit
- Q = 1 - P

(Cochran 1977: Mendenhall *et al.* 1971).

This ratio can be used to correct the sample sizes of cluster samples for use with simple random sampling formulas.

For the present study,

$$\frac{V_{bin(p)}}{V_p} = 0.14 \qquad \text{(Eq. 4.2)}$$

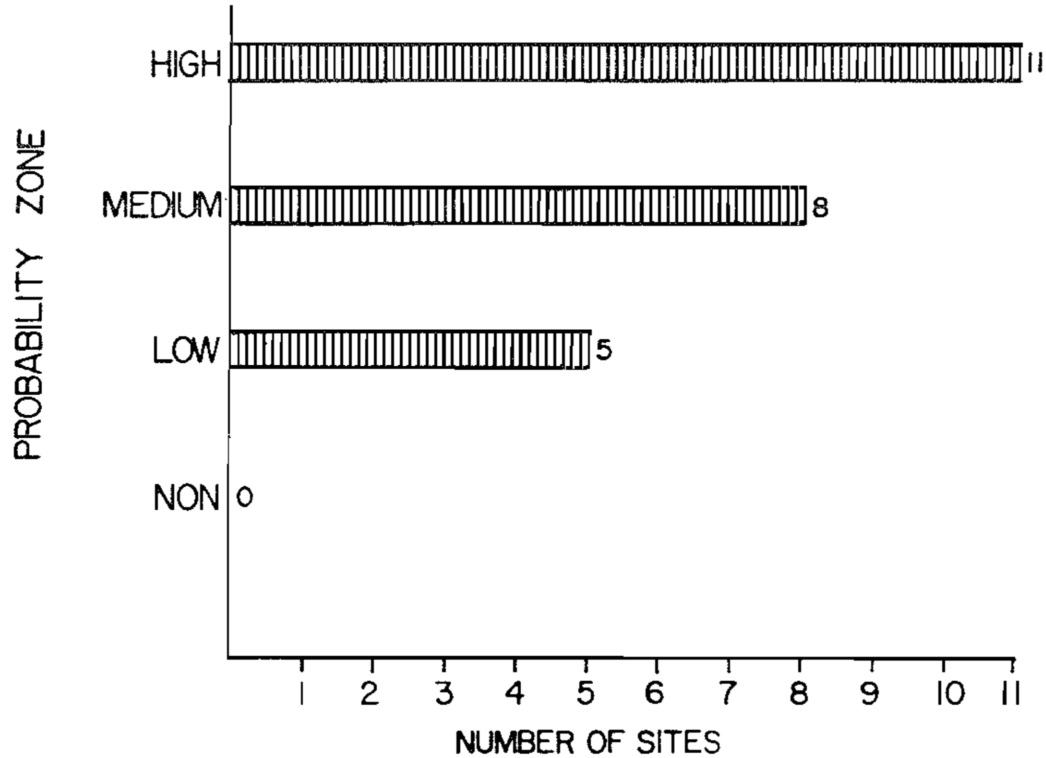
indicating that the actual sample should be reduced to approximately 14% of its true size if random sampling formulas are to be used.

Systematic sampling has adverse effects on sample reliability in two situations. First, if the items sampled are ordered, a biased sample may result. Second, if the items sampled exhibit cyclical or periodic values, the sampling interval may coincide with this periodicity. Again, a biased sample may result. In the present case, the items selected consisted of shovel tests, and the variable of interest was the presence-absence of archaeological material. Since it was very unlikely that this variable exhibited any ordering or periodicity vis-a-vis the 30 meter interval between shovel tests, the possibility of such biases was discounted. When this is the case, systematic samples are analytically equivalent to simple random samples (Cochran 1977; Blalock 1972).

For all of the statistical tests conducted to evaluate the predictive model, sample size was altered to correct for the effects of cluster sampling. Since stratification generally increases the efficiency of a sample, and since systematic samples are equivalent to random samples under the conditions of this study, no corrections for these sampling methods were implemented. The statistical tests used in this study represent approximations of more exact tests that would ideally have been used. They are conservative approximations, however, and should therefore provide reliable indications of hypothesis rejection levels.

Statistical Test

Several alternative methods for testing the predictions of the preliminary model were explored. First, the number of archaeological sites discovered in each probability zone was examined (Figure 4.3). On the basis of these data alone, the predictions of the model were correct. Eleven sites were identified in high probability areas, eight in medium probability areas, five in low probability areas, and zero in nonprobability areas. However, several problems with these test results exist. First, it treats all archaeological sites, regardless of size or significance, as equivalent units. For example, a single flake recovered from an isolated shovel test (e.g., 31NH615) is treated as equivalent to a large, dense, village site (e.g., 31NH613). With an overall sample of only 24 sites, it cannot be assumed that such discrepancies will be distributed evenly among probability zones; thus, certain zones may exhibit high site frequencies simply because they produced a relatively large number of isolated artifact finds. To control for this effect, the area within each probability zone covered by archaeological sites can be computed among the four probability zones (Figure 4.3). Again, the result is consistent with the predictions of the



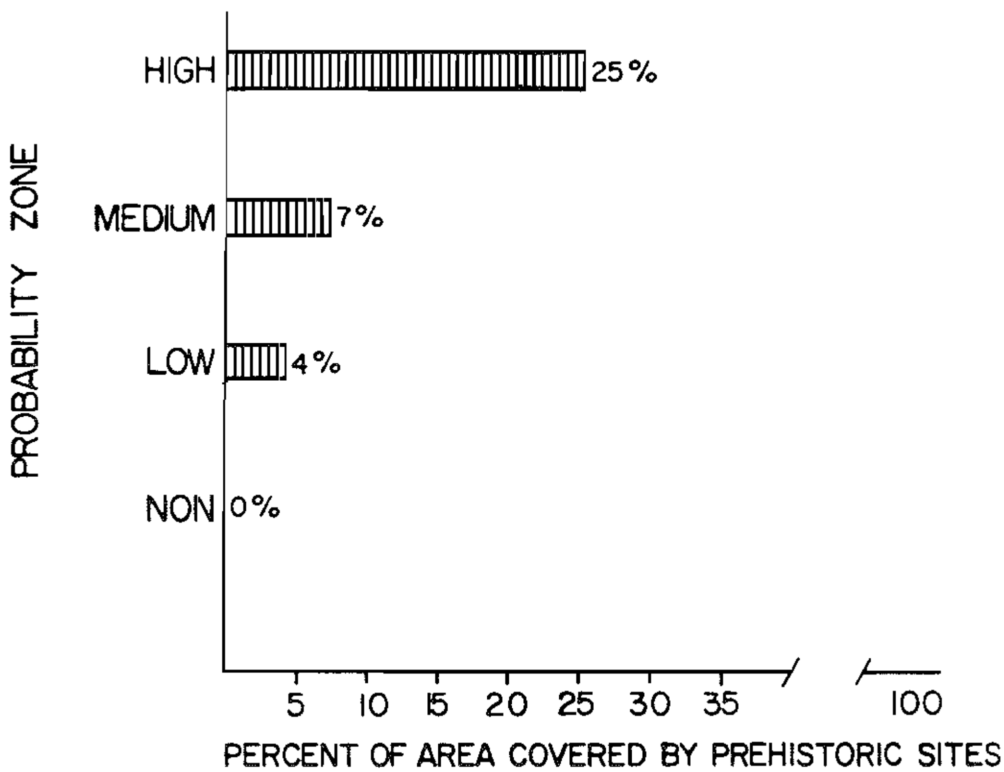


Figure 4.3 Archaeological resources per probability zone

model, since 25% of high probability land surfaces contained archaeological material, while only 7% of medium probability surfaces, 4% of low probability surfaces, and 0% of nonprobability surfaces produced archaeological material. However, this second result, as well as the first, suffers from a further difficulty. For both it is assumed that each probability zone received equal survey coverage, thus having an equal chance of producing archaeological material. Although equal amounts (10 hectares or 25 acres) of each zone were selected for survey, effective survey coverage for the four zones was not equal. High probability areas were generally located in well-drained upland areas, and were either cultivated or covered by mature, open forests. As a result, most of the testing points located in high probability areas were accessible, and shovel tests were actually excavated at 94% of all testing points. In contrast, nonprobability units were frequently located in low-lying, marshy areas with dense pocosin vegetation, and only 55% of all testing points were accessible and actually sampled. Low and medium probability areas were intermediate between these two extremes; 78% of all shovel tests were accessible in the former, 89% in the latter. Because of these differences in effective survey coverage, the sample used to test the model is weighted in favor of the higher probability zones. As a result, uncorrected counts of numbers of sites per probability zone and uncorrected calculations of percentages of probability zones covered by sites are biased.

A measure which corrects for differential survey coverage is the proportion (per probability zone) of excavated shovel tests which yielded artifactual material. This measure can be taken as an unbiased estimator of the relative amounts of site covered land surface in different probability zones. It reveals that high probability areas produced the highest percentages (32%) of culturally nonsterile tests, and that nonprobability areas produced the lowest (0%). However, medium (11%) and low (12%) probability areas produced nearly identical percentages of culturally nonsterile tests (Table 4.2). This latter result is not consistent with the predictions of the model.

Despite this discrepancy, the test of the model appeared to indicate a general association between higher probability areas and greater numbers of nonsterile tests. This apparent association was further evaluated on statistical grounds. The most commonly used test of association employs the chi-square statistic, and the null hypothesis is rejected if chi-square exceeds the value specified at a preselected level of significance. The standard significance level for most statistical testing purposes is $\alpha = 0.05$. In the present analyses, however, sample sizes are small. When the shovel test is considered as the unit of analysis, the effective sample for testing purposes must be reduced to 14% of its uncorrected size. Thus the original sample of 63 nonsterile tests constituted an effective sample of only 9 tests. With a sample of this size, only very strong associations will achieve statistical significance at the 0.05 level. A significance level of 0.10 was therefore treated as an indication that the null hypothesis should probably be rejected, while a level of 0.05 was taken as an indication that the null hypothesis should definitely be rejected.

Table 4.2 Test results by unit and probability zone

Prob. Zone	Individual Units: Unit	No. Tests Exc.	# Tests With Preh. Mat.	% Tests With Preh. Mat.	Totals: No. Tests Exc.	# Tests With Preh. Mat.	% Tests With Preh. Mat.
High	8A	27	10	37	130	40	31
	12	23	6	26			
	20	27	4	15			
	31	25	19	76			
	36	28	1	4			
Medium	3	25	2	8	118	12	11
	11	26	2	8			
	39	18	1	6			
	40	23	7	30			
	56	26	0	0			
Low	1	30	2	7	96	11	12
	16	23	1	4			
	29	28	8	29			
	42	15	0	0			
Non	27	27	0	0	62	0	0
	37	3	0	0			
	46	11	0	0			
	47	0	0	0			
	75	21	0	0			

Taking the effects of cluster sampling into account, and correcting for the differential coverage of probability zones, a chi-square statistic was computed for the distribution of culturally nonsterile shovel tests by probability zone (Table 4.3). The result failed to achieve statistical significance at either the 0.05 or the 0.10 level. Thus, the null hypothesis (i.e., that there is no association between probability zones and the distribution of archaeological material) could not be rejected, and the observed association between such material and probability zones (Figure 4.3) was not confirmed.

Summary and Discussion

The results of the field test were found to be in general conformity with the predictions of the model, since the higher probability zones produced greater amounts of archaeological material. However, deviations from the model's predictions also occurred. In particular, the medium and low probability zones produced nearly identical amounts of archaeological material. Furthermore, the statistical significance of the apparent association between the higher probability zones and the occurrence of archaeological materials could not be established

In two respects, these results damage the utility of the preliminary model for planning purposes. First, in the absence of statistical validation, one must entertain the possibility that the apparent association between probability zones and archaeological resources occurred by chance alone. The rather clear pattern exhibited by the date suggests that this was not the case; i.e., that the failure of the test to achieve statistical significance probably relates to the small size of the sample used in the test. Until the significance of the association is established by more extensive field tests, however, the use of the preliminary model for planning or review purposes would constitute a questionable procedure. A second and more damaging weakness revealed by the test of the preliminary model suggested that further field tests of the model were not warranted and that developing alternative predictive models for New Hanover County would constitute a more productive approach.

The primary weakness of the preliminary model was its failure to clearly distinguish those areas most likely to contain the majority of the county's archaeological sites from areas unlikely to contain sites. As indicated above, the field survey revealed nearly identical densities of sites in both medium and low probability areas. On the basis of this result, these two zones must be treated as equivalent from a management perspective. As a result, they must be collapsed into a single zone characterized by an intermediate site occurrence probability. Although such a procedure would seem to be valid, it in fact largely obviates the utility of the model as a planning tool. The resulting collapsed zone would include approximately 81% of New Hanover County's land surface and approximately 99% of its archaeological sites. In contrast, less than 1% of the county's area and approximately 1% of its sites would be included in the unaltered high probability zone, whereas 19% of the county and 0% of its sites would remain in the nonprobability zone. Thus most of the

Table 4.3 Chi-square test of the association between nonsterile shovel tests and probability zones

Prob.	# Tests Obs. f_o	a. # Tests Obs. Cluster Correction f_{oc}	b. # Tests Exp. f_t	c. # Tests Exp. Cluster Correction f_{tc}	$\frac{(f_{oc} - f_{tc})^2}{f_t}$
High	40	5.6	20.8	2.9	2.5
Medium	12	1.68	18.9	2.7	.4
Low	11	1.54	15.4	2.2	.2
Non	0	0.0	9.9	1.4	1.4
					X = 4.5
					df = 3

a. Frequencies have been corrected for the effects of cluster sampling using a 0.14 correction factor. Small sample corrections have not been used because corrected frequencies are not integers.

b. The expected frequencies used here were calculated to account for the differential survey coverage of probability zones. For example, fewer shovel tests were excavated in nonprobability areas; thus, one would expect fewer sites to have been discovered there.

c. To be comparable to the observed frequencies, the expected frequencies must also be corrected for the effects of cluster sampling.

county's land surface and the vast majority of its sites would be included within one very large probability zone. Such a model has little planning value, because of its inability to discriminate areas which contain archaeological sites from areas which do not contain sites. Thus, although the probability zones distinguished in the preliminary model may reflect actual site distributions, the manner in which they do so is not useful for predictive modeling purposes.

A more detailed analysis of the field results was conducted in order to identify the source of this latter difficulty with the preliminary model. The distribution of archaeological resources identified by the survey was not compared to the overall predictions of the preliminary model. Instead, comparisons were made for each of the four variables included in the model. The results (Figure 4.4) indicated that the problem encountered in validating the model cannot be attributed to sample size alone. Instead, they reveal major discrepancies between the predictions of the preliminary model and the actual distribution of archaeological materials in New Hanover County. These discrepancies can be traced to problems inherent in the nature of the original data used to generate the model. The results of the comparisons of these four variables have important implications concerning the types of data that are and are not appropriate for predictive modeling purposes.

Type of Nearest Water

During the development of the predictive model, the distribution of all archaeological sites in the original data set for type of nearest water was produced (Figure 3.3). This distribution indicated that the majority of sites were located near streams. Localities nearest to streams were thus assigned a high probability of containing archaeological remains, while localities located nearest to other types of water (e.g., ponds, lakes, swamps, salt water) were assigned to the medium or low probability zones. The results of the field survey conducted to test the model indicated, however, that these probability assignments were largely in error. Shovel tests located near natural ponds and lakes, which were assigned medium and low probability values in the predictive model, produced the highest percentages of archaeological material. In contrast, those located near streams, which were assigned to the high probability zone in the model, had the lowest percentage. Tests near saltwater and swamps (medium and non-probability, respectively) had low percentages also (Figure 4.4). These discrepancies from the predictions of the model can be attributed in part to biases in the data upon which the model was based, and in part to the conventional format used for recording archaeological site data.

Biases in the data used to generate the preliminary model may have resulted from the use of topographic maps to assess the environmental characteristics of site localities. In New Hanover County, small swamps, springs and streams that do not appear on topographic maps (U.S.G.S., 7.5 Series) of the area are plentiful. As a result, the identification of water sources using such maps may introduce significant biases into the data. If, as seems likely, a large proportion of the environmental data recorded for the sites originally on file for the county were generated using

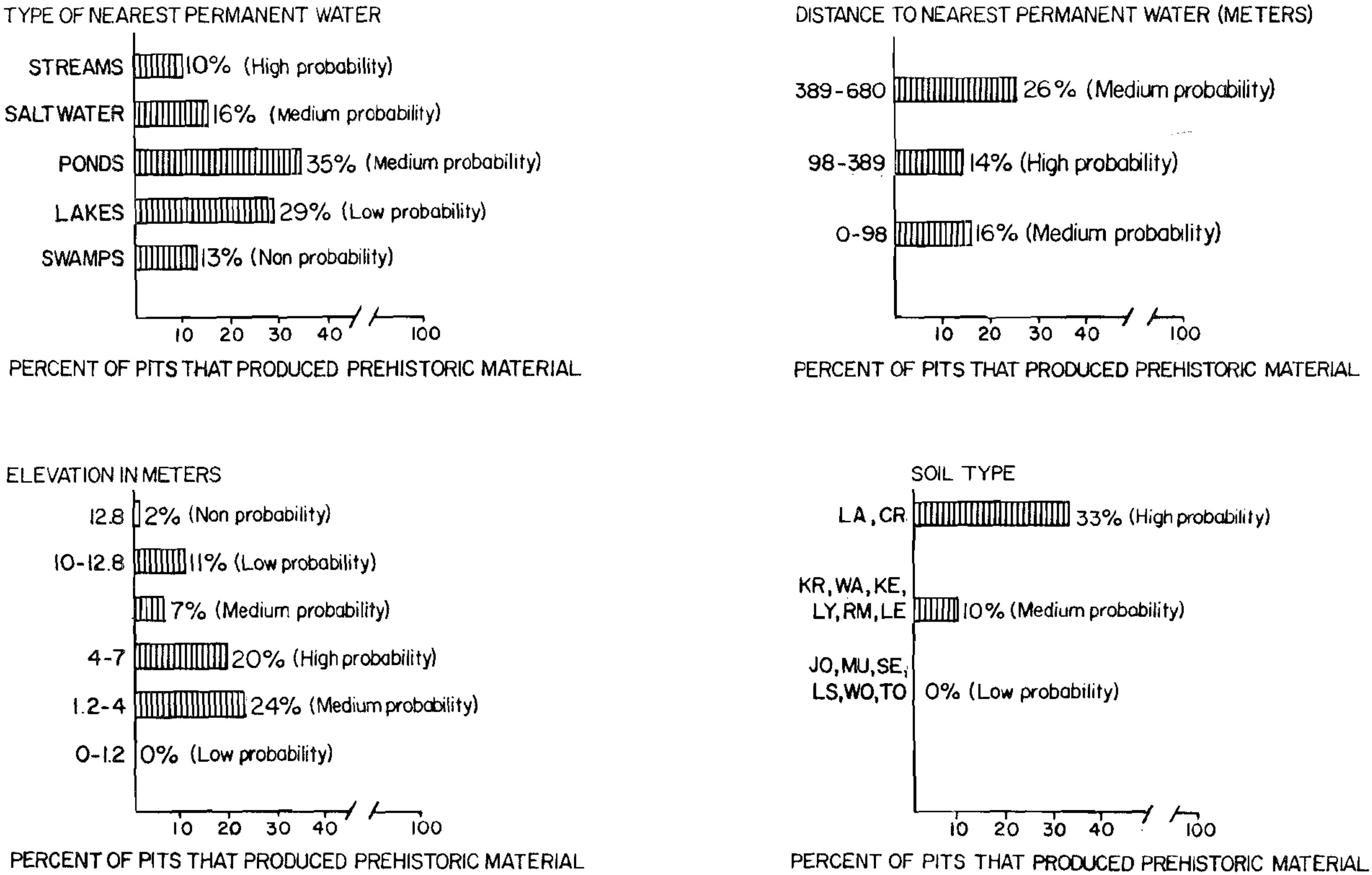


Figure 4.4 A comparison of the distribution of archaeological resources revealed during the field survey with the distribution predicted in the preliminary predictive model

topographic maps, the type of water nearest to many of these sites may have been misidentified. Errors of this kind may thus have contributed to the misleading predictions of the predictive model developed here.

The use of the archaeological site as the basic recording unit in conventional archaeological surveys is potentially a more important problem. Developing the preliminary model involved using site file data to determine which type(s) of water was near the majority of known sites in the county. Unfortunately, these data could provide no information concerning the differential distribution of different types of water sources. For example, if 90% of all the water sources in an area are streams, then it is quite probable that most of the archaeological sites in that area will be located nearer to streams than to any other type of water source, even if the sites are located entirely randomly with respect to water source type. Conventional archaeological site location data can only indicate that most sites are near streams. They cannot also indicate that streams are more prevalent than other types of water sources. As a result, models based on this type of data will assign a high probability of containing archaeological material to localities near streams, despite the fact that such localities may actually exhibit equivalent or lower site densities than localities near other water source types.

An examination of the topographic maps for New Hanover County indicates that streams are the most widespread water source within the county. Many sites should therefore occur near streams simply by chance. Because it was based on conventional site file data, the predictive model developed here could not evaluate the differential distribution of streams within New Hanover County. The predictive errors revealed by the present survey are undoubtedly due at least in part to this factor.

It should be emphasized that this modeling difficulty is inherent to all predictive models that are based solely on conventional site data. Whether they be simple narrative descriptions of the types of localities where sites tend to occur, or more sophisticated statistical summaries of large site location data sets, such models do not systematically control for the differential areal distributions of different topographic settings. As a result, it is generally impossible to determine whether most sites occur in certain types of locations because prehistoric populations actually preferred them or simply because locations of that type are more prevalent than locations of other types.

Distance to Nearest Water

Similar procedures to those described for the previous variable were used to analyze distance to nearest water in the development of the preliminary model. A graph showing the distribution of all sites in the data set for various distances to water was produced. This graph indicated at what distance from water the majority of known sites in New Hanover County were distributed. This zone was thus assigned a high probability of producing archaeological sites; medium and low probability zones were defined in an analagous fashion. However, the survey conducted to test the model indicated that these probability assignments were largely in error (Figure 4.4). These discrepancies between the predictions of the model and the empirical distribution of archaeological sites in New Hanover County can be attributed to the biases in and the format of the original data used to generate the model. As in the case of type of nearest water, the use of topographic maps to measure geographic variables may have introduced unrecognized biases into the data.

A potentially more important problem is related to the conventional archaeological site recording format, which provided no information concerning the differential distribution of topographic settings. In an area where water is abundant, localities within 500 meters of a water source may be more plentiful than localities 500 meters or more away from a water source. Under these circumstances, more sites should be located less than 500 meters from water than are located at greater distances. With conventional archaeological data, it is impossible to determine whether larger numbers of sites in the former zone represent a true concentration of archaeological resources, or whether they simply represent the greater areal extent of that zone. Since water sources are plentiful in New Hanover County, it is likely that a similar problem affected the reliability of the preliminary predictive model.

Elevation

Elevation was analyzed using procedures similar to those described for the previous two variables in the development of the preliminary model. A graph illustrating the distribution of archaeological sites by elevation zones was produced. This graph indicated the elevation zone which produced the majority of known sites. This zone was thus assigned a high probability rating. Medium and low probability zones were defined analagously. However, these probability designations contain discrepancies (Figure 4.4). The discrepancies can be attributed to the same two factors discussed above, i.e., to biases in the original data and to the conventional archaeological data recording format.

The values for elevation contained in the original data provide a clear example of the potential biases that may be encountered in conventional site file data. In the New Hanover case, these biases can be traced to past survey strategies, which focused almost exclusively on exposed ground surfaces. Within the county, such surfaces are encountered primarily in cultivated fields. It follows that most of the sites in the data set used to generate the model were discovered in such localities. Unfortunately, the cultivated portions of New Hanover County constitute a biased sample of the county as a whole. The interior plateau, although at a higher elevation than most of the remaining portions of the county, exhibits little topographic relief. As a result, the water table is close to the surface, and the region is quite marshy. As would be expected, fewer farms are located within this portion of the county. The lowest parts of the county, which are present near the coast and along major streams and rivers, are similarly marshy and are rarely cultivated. The transitional zone between the interior plateau and the low lying coastal and riverine areas is more highly dissected, well drained, and more intensively cultivated. It is this latter topographic zone which is represented primarily in the data set used to construct the model. Interestingly, this zone generally lies between 3 meters and 9 meters (10' and 30') in elevation, which corresponds closely to the zone designated in the model as having the highest probability of producing archaeological material. The designation of this zone can thus be attributed at least in part to biases introduced by the nearly exclusive focus on cultivated areas during previous surveys of the project area.

As with the previously discussed variables, a second factor contributing to the low predictive success of this variable relates to the conventional data recording format, in which the site is the basic analytic unit. Because such data do not control for the areal extent of topographic settings, a larger number of sites in a given elevation zone may reflect the greater areal extent of that zone rather than a true concentration of archaeological resources. Because modeling procedures based only on conventional site data, cannot control for such differences in the areal extent of elevation zones, misleading probability designations may result.

Soil Type

Slightly different procedures were used to analyze soil type during the development of the preliminary predictive model. In contrast with the other three variables, data summarizing the differential areal extent of the soil types found in New Hanover County were available (Weaver 1977). This information was used to convert the raw counts of archaeological sites per soil type into density measures. These density measures were used to identify which soil types exhibited high, medium and low probabilities of containing archaeological sites. By correcting the raw data counts in this fashion, one of the primary difficulties experienced with the other three variables was avoided, i.e., it was possible to control for the differential areal distribution of various soil types. However, these corrections were made possible by the geographic data available through the LRIS (Land Resources Information Service) computer graphics system and could not have been derived from the site file data used to generate the model. In this respect, the procedure used to analyze soil type as a predictive variable differed from a more common approach, which simply summarizes uncorrected site location data. Furthermore, the procedure used here will frequently be unavailable to archaeologists seeking to develop predictive models based on conventional site file data.

In contrast to the other three variables, the distribution of archaeological materials by soil type as indicated by the field survey largely conformed to the predictions of the predictive model (Figure 4.4). The success of this variable as a predictor of archaeological materials can be attributed to the corrective measures that were implemented to control for the differential distribution of soil types. Despite these corrective measures, however, the predictions of the model concerning the distribution of archaeological resources by soil type were not entirely accurate.

The soil types designated as having a high probability of site occurrence (i.e., Lakeland, Craven) yielded high percentages of culturally non-sterile tests during the present survey, a result which is consistent with the predictions of the model. Medium probability soils, however, include some (Kenansville and Wakulla) that produced greater amounts of archaeological material than any high probability soils, some that produced intermediate amounts of archaeological material (Rimini and Kureb), and others that produced virtually no archaeological material (Leon and Lynn Haven). This result is not consistent with the predictions of the model. Low probability soils again conform to the predictions of the model, since none of these soils produced significant amounts of archaeological material.

These discrepancies from the predicted pattern can again be attributed to biases in the data set used to derive the model. For the sites in this data set, soil types were identified using soils maps. However, even detailed soils maps are not sufficiently accurate to show many of the smaller soil units that actually exist within the county (Weaver 1977). For example, broad areas of poorly drained soils such as Leon sand or Murville fine sand contain small knolls or ridges which exhibit other, well drained soil types such as Wakulla sand or Craven fine sandy loam. These latter soil units are frequently too small to appear on soils maps; however, sites are frequently located in just such localities. When soils maps are used to identify soil types for such sites, incorrect assignments will occur, and biases may be introduced. These biases may be reflected in the discrepancies between the predicted distribution of archaeological materials by soil type and the actual distribution as revealed by field survey.

Summary

The problems encountered with the preliminary predictive model developed during this study can be attributed to unrecognized biases in the data used to generate the model and to the conventional archaeological data recording format. The biases that are present in the site file data for New Hanover County may derive from several sources. Among the more important, however, are the use of topographic and soils maps to assess the environmental characteristics of site localities and the nearly exclusive focus during previous surveys on exposed ground surfaces, especially cultivated fields. To a greater or lesser extent, these same biases probably characterize the bulk of currently available site location data for the eastern United States, and perhaps for other regions as well. The results presented here indicate that such biases may result in misleading predictions concerning which areas are most likely to contain archaeological sites. Unless currently available site file data is validated by field surveys based on probabilistic sampling procedures, it can provide an unreliable basis for the generation of predictive models.

As revealed during the present study, a more important problem with conventional archaeological site file data sets results from the use of the archaeological site as the basic recording and analytical unit. Such data can provide no information concerning the areal extent of the various topographic settings of interest. Consequently, they can only indicate which settings have produced greater or lesser numbers of previously discovered sites. Such data cannot indicate whether a given feature exhibits a large number of sites because it is relatively widespread, because it actually exhibits a higher concentration of archaeological resources, or because both of these factors are operating in combination. This difficulty may result in misleading predictions concerning which areas are most likely to contain archaeological materials.

This latter difficulty arises from a discrepancy between the nature of conventional archaeological data and the data needs of predictive models. Unfortunately, the primary unit of interest for predictive modeling purposes is <u>not</u> the archaeological site, but is instead some unit or units of land surface. The goal of such a model is to predict whether or not a unit or units is likely to contain archaeological material. To make such a prediction, information must be available for localities which do not contain sites, as well as for those that do.

CHAPTER FIVE

ALTERNATIVE PREDICTIVE MODELS

The preliminary predictive model generated and tested during the first two phases of the New Hanover County Project proved to be an informative first step in model developments. At a general level, the model appeared to successfully predict archaeological site distribution within New Hanover County. However, problems were encountered in attempting to validate the model's accuracy. These problems reduced the utility of the preliminary model as a planning and management tool.

The third phase of the New Hanover County Project involved an exploration of alternative approaches to the problem of predictive modeling that might avoid these difficulties. These approaches differed in several respects from the approach used in the development of the preliminary model. First, and most important, the data that were used were more appropriate for predictive modeling purposes. The data generated during the survey phase of the project were collected using probabilistic sampling techniques, thus assuring a more representative sample. In addition, these latter data were collected for every field test regardless of whether archaeological materials were present or not. Thus, archaeological sites were not the units of analysis. Instead, survey units or shovel tests, arbitrarily defined units of land surfaces, provided units of analysis that were appropriate for predictive modeling purposes.

Secondarily, the procedures involved in investigating alternative models differed from those used for preliminary model development. In addition to descriptive data summarization, a statistical approach to predictive modeling - regression analysis - was conducted. When based on the data provided by the field survey project phase, both approaches generated potentially useful and reliable models.

Descriptive Model

A two-stage analytical procedure was used to generate a revised, descriptive model. The first stage involved variable selection and assessment of the degree of association between different variables and the distribution of archaeological material within New Hanover County. The variable or variables showing the strongest association with that distribution were then selected for inclusion in the model. The second stage involved assessing the nature or type of association between these selected variables and the archaeological resources.

Variable Selection

The association between each of the four environmental variables included in the preliminary model and the distribution of archaeological materials in New Hanover County was assessed by computing chi-square statistics. Sample sizes were reduced to correct for the effects of cluster sampling, and a significance level of 0.10 was selected. For each test the null hypothesis can be stated:

> H_0: There is no association between archaeological materials and probability zones defined on the basis of the given variable.

and the alternative hypotheses, which is confirmed if the null hypothesis can be rejected, is:

> H_1: An association does exist between archaeological materials and probability zones defined by the given variable.

The results indicate that of the four variables included in the original model, only soil type exhibits a statistically significant association with the presence of archaeological materials (Table 5.1). This variable was therefore selected as a basis for alternative predictive modeling purposes.

The soil types sampled during the survey can be placed into two general categories: those that yielded relatively large numbers of culturally non-sterile shovel tests and those that yielded very few or no such tests. The former soils can be considered high probability soils; the latter low probability soils. If all the soil types that occur within New Hanover had been sampled during the survey, this classification would provide a reliable predictive model. However, only approximately 50% of the soil types that actually occur within the county were sampled during the survey. In order to formulate a predictive model on the basis of soil type, it was thus necessary to generalize from sampled soil types to unsampled soil types.

Such generalization poses a special problem when soil type is the variable of interest. Soil type designations are made on the basis of numerous ancillary variables, such as fertility, permeability, drainage, etc. To generalize about site distributions from sampled to unsampled soil types, it was necessary to determine which of these various soil characteristics contributed to the desirability of certain soils as habitation locations and thus to the higher site densities exhibited by those soils. Predictions could then be made that any soil possessing those same desirable characteristics would exhibit similarly high site densities. A predictive model based on soil characteristics, rather than on soil types *per se*, was thus required.

Data for numerous soil characteristics were available from the United States Soil Conservation Service (Weaver 1977). Although much of this data (such as suitability for sanitary facilities) probably pertains exclusively to modern land-use practices, four characteristics for which data were available may have influenced prehistoric (as well as modern) land use practices. These included suitability for agricultural crops (corn), drainage, forest productivity, and wildlife potential. These four variables were therefore evaluated to determine which were the most highly associated with

Table 5.1 Chi-square tests of the associations between predictive variables and nonsterile shovel tests

Probability Zone	# Tests Obs. f_o	# Tests Obs. Cluster Correction f_{oc}	# Tests Exp. f_t	# Tests Exp. Cluster Correction f_{tc}	$\frac{(f_{oc}-f_{tc})^2}{f_{tc}}$
TYPE OF NEAREST WATER					
High (streams)	12	1.7	18.3	2.6	0.3
Medium (saltwater, ponds)	26	3.6	20.0	2.8	0.2
Low (lakes, swamps)	25	3.5	24.7	3.5	0.0
					X = 0.5
					df = 2.0
					not significant
DISTANCE TO NEAREST WATER					
Medium (0-30m)	30	4.2	33.8	4.7	0.1
High (30-118.6m)	25	3.5	24.3	3.4	0.0
Medium (118.6-207m)	8	1.1	4.8	0.7	0.2
					X = 0.3
					df = 2.0
					not significant
ELEVATION					
Low (0-1.2m)	0	0.0	0.0	0.0	0.0
Medium (1.2-4m)	23	3.2	14.7	2.1	0.5
High (4-7m)	32	4.5	25.3	3.5	0.3
Medium (7-10m)	6	0.8	13.8	1.9	0.6
Low (10-12.8m)	1	0.1	1.4	0.2	0.1
Non (12.8m)	1	0.1	7.4	1.0	0.8
					X = 2.4
					df - 5.0
					not significant
SOIL TYPE*					
High (LA,CR)	41	5.7	19.4	2.2	3.3
Medium (KR, WA, KE, LY, RM, LE)	21	2.9	37.2	5.2	1.0
Low (JO,MU SE,LS, WO,TO)	0	0.0	25.6	3.6	3.6
					X = 7.9
					df = 2.0
				significant when	α = .02

*Totals differ from those shown in Table 4.1 because disturbed (hence unidentifiable) soil profiles were encountered in 2 shovel tests. One of these shovel test contained cultural material, the other did not.

the occurrence of archaeological materials, and would thus constitute the best predictors of archaeological site distributions. Chi-square statistics were again employed for this purpose, and indicated that agricultural suitability and good drainage tended to be associated with archaeological materials (Table 5.2). The other two variables examined - forest productivity and wildlife potential - did not exhibit a significant association with the occurrence of these materials. Agricultural suitability and soil drainage character were thus selected as the best predictors of archaeological site distributions in New Hanover County.

Revised Descriptive Model

Developing a predictive model based on these two variables involved a further investigation of the association between each and the occurrence of archaeological material within New Hanover County. These associations are illustrated in Tables 5.3 and 5.4 for soil types sampled during the present survey. A comparison of these tables indicated that the soil drainage characteristic constituted a more powerful predictive variable than suitability for agricultural crops. All well-drained and excessively well-drained soils produced relatively high percentages of shovel tests containing archaeological material, whereas all poorly-drained and very poorly-drained soils produced relatively few such tests. In general, soil suitability for crops was also associated with the occurrence of archaeological material. However, some soils that are unsuitable for cultivation produced relatively high percentages of culturally nonsterile tests. Soil drainage character was thus more clearly associated with the distribution of archaeological resources, and was selected as the primary basis for a revised descriptive model. However, further refinement in this model was achieved by using both soil drainage character and agricultural suitability in a hierarchical fashion. When well-drained soils were subdivided into agriculturally suitable and unsuitable subsets, those soils that yielded the highest percentages of nonsterile tests were distinguished from those which yielded moderate percentages of nonsterile tests. The resulting classification included two high probability categories (very high and moderately high) and one low probability category. The association between these categories and the occurrence of archaeological material achieved a high level of statistical significance (chi-square = 11.30, df = 2, significant when $\alpha = 0.01$), thus providing an excellent basis for predicting the occurrence of prehistoric materials.

When used in combination, the drainage characteristic and the agricultural suitability of soils constitute a reliable basis for assigning relative probabilities of site occurrence to soil types that were not surveyed during the present study (Table 5.5). The result is a predictive model for New Hanover County which should prove to be a reliable cultural resource management tool. Because this model is based only on the distribution of soil types, it can be easily implemented. Maps for any portion of the county indicating which areas have the highest probability of containing archaeolgoical materials can be quickly and efficiently produced (Figure 5.1). Furthermore, the high probability areas defined by the model contain the majority of New Hanover County's archaeological resources. As defined

Table 5.2 Chi-square tests of the associations between soil characteristics and nonsterile shovel tests (cf. Weaver 1977)

Soil Characteristic	# Tests Obs. f_o	# Tests Obs. Cluster Correction f_{oc}	#Tests Exp. f_t	# Tests Exp. Cluster Correction f_{tc}	$\frac{(f_{oc} - f_{tc})^2}{f_{tc}}$
FERTILITY:					
Suitable	54	7.6	22.8	3.2	6.1
Unsuitable	8	1.1	37.8	5.3	3.3
					X = 9.4
					df = 1.0
				significant when	α =.01
DRAINAGE					
Well drained	60	8.4	32.7	4.6	3.1
Poorly drained	2	0.3	30.0	4.2	3.6
					X = 6.7
					df = 1.0
				significant when	α =.01
FOREST PRODUCTIVITY: (site index)					
High (90-110)	0	0.0	1.4	0.2	0.2
Medium (70-90)	56	7.8	50.4	7.1	0.1
Low (50-70)	6	0.8	8.9	1.3	0.2
					X = 0.5
					df = 2.0
					not significant
WILDLIFE POTENTIAL*:					
Good	12	1.7	10.1	1.4	0.1
Fair	34	14.8	24.2	3.4	0.6
Poor	8	1.1	14.3	2.0	0.4
Very Poor	6	0.8	12.4	1.7	0.5
					X = 1.6
					df = 3.0
					not significant

* Totals differ because data is not available for all soil types (cf. Weaver, 1977).

Table 5.3 Soil suitability for crops as a predictive variable

Individual Soil Types					Totals[b]		
Soil Type[a]	Cultivation of Maize	No. Tests Exc.	# Tests with Preh. Mat.	% Tests with Preh. Mat.	No. Tests Exc.	# Tests with Preh. Mat.	% Tests with Preh. Mat.
CR	suitable	35	7	20			
KE	suitable	12	5	42			
LA	suitable	90	34	38			
WA	suitable	15	8	53	152	54	36
LS	unsuitable	1	0	0			
LY	unsuitable	19	0	0			
SE	unsuitable	21	0	0			
JO	unsuitable	9	0	0			
KR	unsuitable	45	4	9			
LE	unsuitable	114	2	2			
MU	unsuitable	4	0	0	252	8	3
RM	unsuitable	14	2	14			
TO	unsuitable	19	0	0			
WO	unsuitable	6	0	0			

a. Soil type abbreviations are from Weaver (1977).

b. Totals differ from those reported in Table 4.1 because disturbed (hence unidentifiable) soil profiles were encountered in 2 shovel tests. These tests have been omitted from all analyses of the association between soil type and archaeological material.

Table 5.4 Soil drainage character as a predictive variable

	Individual Soil Types				Totals[b]		
Soil Type[a]	Drainage Character	No. Tests Exc.	# Tests with Preh. Mat.	% Tests with Preh. Mat.	No. Tests Exc.	# Tests with Preh. Mat.	% Tests with Preh. Mat.
CR	well drained	12	5	42			
KE	well drained	35	7	20			
KR	well drained	45	4	9	211	60	28
LA	well drained	90	34	38			
RM	well drained	14	2	14			
WA	well drained	15	8	53			
JO	poorly drained	9	0	0			
LE	poorly drained	114	2	2			
LS	poorly drained	1	0	0			
LY	poorly drained	19	0	0	193	2	1
MU	poorly drained	4	0	0			
SE	poorly drained	21	0	0			
TO	poorly drained	19	0	0			
WO	poorly drained	6	0	0			

a. Soil type abbreviations are from Weaver (1977).

b. Totals differ from those reported in Table 4.1 because disturbed (hence unidentifiable) soil profiles were encountered in 2 shovel tests. These tests have been omitted from all analyses of the association between soil type and archaeological material.

Table 5.5 A revised predictive model based on soil drainage and suitability for crops

Probability Zone	1. Soil Type	2. Drainage Character	Cultivation of Maize
	BE (BH)	well drained	suitable
	CR*	well drained	suitable
	KE*	well drained	suitable
	LA*	well drained	suitable
High 1	NO	well drained	suitable
	ON	well drained	suitable
	WA*	well drained	suitable
	WR	well drained	suitable
	KR (KU)*	well drained	unsuitable
High 2	NH	well drained	unsuitable
	RM*	well drained	unsuitable
	BA	poorly drained	unsuitable
	DO	poorly drained	unsuitable
	JO*	poorly drained	unsuitable
	LE (LO)*	poorly drained	unsuitable
	LS*	poorly drained	suitable
Low	LY*	poorly drained	suitable
	MU*	poorly drained	unsuitable
	PM	poorly drained	unsuitable
	PN	poorly drained	unsuitable
	RA	poorly drained	unsuitable
	SE (SH)*	poorly drained	suitable
	ST	poorly drained	suitable
	TM	poorly drained	unsuitable
	TO*	poorly drained	unsuitable
	WO*	poorly drained	unsuitable

* Sampled during present survey.

1. Soil type abbreviations are from Weaver (1977).
2. Soils classified as suitable but poorly drained were placed in the low probability category because they are only cultivatable through the use of modern agricultural practices.

Figure 5.1 The distribution of High 1 and High 2 probability soils in New Hanover County

in the revised model, approximately 26.0% of the county falls within the very high probability category. Present data indicates that 35.5% of all shovel tests excavated within these areas should produce archaeological materials; 4,431 hectares (10,950 acres) of this zone should thus be covered by archaeological sites. Similar calculations indicate that 12.7% of New Hanover County falls within the moderately high probability zone, and should contain an estimated 624 hectares (1,543 acres) of archaeological sites. Low probability areas, however, include 56.7% of the county, but should contain only 272 hectares (673 acres) of sites. Thus, fully 94.9% of the county's archaeological resources should be located within either the high or the moderately high probability zones, which together constitute only 38.7% of the county's total area. Soil classifications excluded from the model include urban areas, borrow pits and mine pits. Such areas account for 2,173 hectares (5,369 acres) or 4.5% of the county.

In contrast with the original model, the revised model should prove to be an effective planning tool. Since the high probability zones contain the majority of the archaeological resources within New Hanover County, identifying these zones within project areas should allow impacts on archaeological resources to be anticipated. Appropriate conservation measures can then be incorporated into the early phases of project planning.

It should not be assumed, however, that further archaeological investigation of low probability areas as defined in this model is unnecessary, or that identifying a project as lying entirely within this zone constitutes compliance with current preservation legislation. In its present form, the model has two important limitations. First, it is based on data derived from a small-scale survey. Before generalizing from this small sample to the remainder of New Hanover County, the model should be tested in the field. A second and more important limitation lies in the accuracy of currently available soils maps. The associations between soil characteristics and archaeological materials used to derive the model were determined from identification of soil types in the field. For unsurveyed areas, however, soil type can only be identified using soil maps. The accuracy of the model is thus limited by the accuracy of these maps. In New Hanover County, archaeological sites frequently occur on small, unmapped high probability soil units. While low probability areas as defined in the model may in actuality contain very few archaeological sites, low probability areas as mapped for unsurveyed regions may contain significant archaeological resources. Although such resources should be much less common in these low probability areas than in high probability areas, preservation legislation nevertheless requires that they be identified. Survey techniques must therefore be implemented in these areas to identify whatever sites may be present. Because of these limitations, the revised predictive model presented here should be used for limited planning and management purposes.

Regression Model

The revised predictive model developed in the previous section, although successful for New Hanover County, has inherent limitations which weaken its applicability to areas where the problem of archaeological prediction may be more complex. The limitations derive primarily from the methods used in model formulation: areas with especially high probabilities of site occurrence were identified largely through the inspection of descriptive data summaries.

In New Hanover County, local topographic variability has been a major factor in the choice of habitation location. Much of the county is flat. The water table is near the surface, and water sources are plentiful. Because of this, well-drained areas suitable for use as habitation sites and for the cultivation of crops are relatively restricted, whereas swampy and marshy areas are widespread. Prehistoric land-use practices have thus concentrated on these restricted, well-drained areas. Since soil characteristics are heavily influenced by drainage, soil classifications embody these variations in drainage character. As a result, a clear association between the distribution of certain soil types and archaeological materials is evident, and provides a simple and reliable method for predicting the occurrence of prehistoric materials in unsurveyed portions of the county.

The complexity of predicting the location of archaeological sites will increase in areas exhibiting greater environmental variability than New Hanover County. Water sources may be less widespread, soils may exhibit a wide range of drainage and fertility, and greater local topographic variability may exist. Under these conditions, choice of habitation locations may have been influenced in complex ways by several variables. Prior to in-depth settlement pattern analysis, the ways in which these different environmental variables determine site location may not be obvious from data inspection and simple descriptive statistics alone. Extension of the modeling procedures presented in the previous section to more complex situations may thus prove difficult.

In order to deal effectively with these more complex modeling situations, it is necessary to turn to more rigorous analytic procedures. Techniques such as regression analysis allow the predictive power of different variables to be evaluated from a mathematical perspective. The logic and computational methods of these techniques permit an objective definition of the variable combination which best models predictive relationships. Furthermore, statistical error estimates for the specific predictions that are generated provide an independent measure of model reliability.

For planning and review purposes, such models should prove powerful. Not only can areas of high probability be identified, but quantitative estimates of the amount of archaeological material that is likely to exist within a given area can be provided.

To illustrate the potential of more analytically powerful predictive approaches, a third and final predictive model was developed for archaeological resources in New Hanover County. The specific modeling approach selected for this purpose was regression analysis. The discussion presented below draws upon conventional regression concepts detailed in a variety of sources, such as Blalock (1972), McClave and Benson (1979), and Neter and Wasserman (1974).

Regression models are equations which express the relationship between a specified set of independent, predictor variables and a dependent, predicted variable. These equations exhibit the following general form:

$$E(Y) = B_0 + B_1X_1 + B_2X_2 \ldots + B_nX_n \qquad \text{(Eq. 5.1)}$$

where,

$E(Y)$ = the predicted (expected) value for Y, the de endent variable

$B_0, \ldots B_n$ = coefficients

$X_1, \ldots X_n$ = values for the independent variables

In archaeological terms, Y might be the amount of archaeological material to be expected at a given locality, or the percent of a unit of land surface covered by archaeological sites. The X values might represent measures for a specified set of environmental characteristics, such as distance from water, soil fertility, etc., for that same locality or unit of land surface. The B values are generated from the data by regression algorithms, and express the weight and the direction (hence, contribution) of each predictor variable to the prediction rule. If values for each X value are known for a given locality, the regression equation can be solved for Y, thus providing a prediction concerning the amount or extent of the archaeological resources to be expected at that locality. Confidence intervals around the predicted value of Y can also be computed, and provide an expected range of error for this value.

Generating a regression model involves a multi-stage analytical procedure. First, a unit of analysis appropriate to the predictive problem at hand must be defined. For archaeological predictive modeling purposes, this unit will generally consist of an arbitrarily defined segment of land surface, such as a square meter or a square kilometer. For the present analysis, an arbitrarily defined 100m x 200m grid unit corresponding to the size of the field test survey units was selected as the unit of analysis.

The second analytical stage involves variable selection; both independent and dependent variables must be defined. In the present example, the fertility and drainage character of the soils exhibited by a 100m x 200m unit were chosen as the independent variables, since these variables had previously been established as the best predictors of archaeological material in New Hanover County. Metric values for each variable were then computed for each 100m x 200m unit. Soil fertility was calculated by the equation

$$X_1 = \frac{\sum_{i=1}^{n} f_1}{n} \qquad \text{(Eq. 5.2)}$$

where,

X_1 = average soil fertility, in bushels of corn per acre (cf. Weaver 1977) for any given 100m x 200m unit
f_1 = the soil fertility at shovel test "i" in the 100m x 200m unit
n = the total number of shovel tests in the unit

This measure of soil fertility per field survey unit is effectively the average fertility of the individual shovel testing points within that unit.

Soil drainage character was quantified using

$$X_2 = \frac{w}{n} \times 100 \quad \text{(Eq. 5.3)}$$

where,

X_2 = the average drainage character of a 100m x 200m
w = the number of shovel testing points located within well-drained soils in the unit
n = the total number of shovel tests within the survey unit

Soil drainage for a given 100m x 200m unit is thus the percentage of well-drained soils within that unit, as determined at the individual shovel testing points. The scalar term (100) is incorporated to avoid potential rounding errors by computer data manipulation.

Values of the dependent variable in this analysis were determined as follows:

$$Y = \frac{a}{n} \times 100 \quad \text{(Eq. 5.4)}$$

where,

Y = the percentage of the surface of a 100m x 200m unit covered by archaeological materials
a = number of culturally nonsterile (i.e., artifact-producing shovel tests in that unit
n = total number of shovel tests excavated within that unit

The variable Y is thus also measured as a percentage. In this case, it represents the percent of culturally nonsterile shovel tests within any 100m x 200m unit of land surface. It should be noted that for each of these three measures (X_1, X_2, Y), any variation in the total number of shovel tests per unit is assumed to produce a constant measure.

In the third analytical stage, a representative sample must be provided of the items for which a prediction rule is desired. In this case, the 100m x 200m units that had been surveyed in order to test the original predictive model provided the required sample. For each such unit, values for X_1, X_2, and Y were calculated using equations 5.2, 5.3, and 5.4 (see Table 5.6).[2] The resulting data set was used as a basis for generating a regression

Table 5.6 X_1 (fertility), X_2 (drainage), and Y (amount of archaeological material) per survey unit

[a]Case (Survey Unit)		# Nonsterile Pitts/# Pits Excavated	# Pits/ Soil Type	[b]X_1 (Fertility)	[c]X_2 (Drainage)	[d]Y (Amount of Arch. Mat. Per Survey Unit)
1	(1)	2/30	19KR,11LE 3LY,16LE	0.0	63	7
2	(3)	2/25	4KR, 2WA	3.6	24	8
3	(8A)	10/27	6LE, 7LY 14LA	28.5	52	37
4	(11)	2/25	21KR, 4LE	0.0	84	8
5	(12)	6/23	18LA, 5LY	43.0	78	26
6	(16)	1/23	6RM,11SE 6JO, 2LE 2SE	0.0	26	4
7	(17)	0/27	6WO,17TO	0.0	0	0
8	(20)	4/27	26LA, 1LE	53.0	96	15
9	(29)	8/28	16LA,12KE	61.4	100	29
10	(31)	18/24	16LA, 1LS 7CR	67.3	96	75
11	(36)	1/28	28CR	105.0	100	4
12	(37)	0/3	3LE	0.0	0	0
13	(39)	1/18	10LE, 8RM 2JO, 7LE	0.0	44	6
14	(40)	7/23	1LY,13WA	25.4	56	30
15	(42)	0/15	8SE, 7LE	0.0	0	0
16	(46)	0/11	7LE, 4MU	0.0	0	0
17	(56)	0/26	25LE, 1JO	0.0	0	0
18	(75)	0/20	15LE, 3LY 2TO	0.0	0	0

a. One survey unit was located entirely in standing water (a large swamp). Since no shovel tests were excavated within that unit, it has been omitted from the regression analysis.

b. Average productivity (bushels of corn per acre) per survey unit. Calculated using Equation 5.2.

c. Calculated using Equation 5.3.

d. Calculated using Equation 5.4.

model predicting the extent of archaeological material to be expected within any 100m x 200m unit of land surface within New Hanover County.

The final stage of regression modeling establishes the actual relationship between the dependent variable and the independent, predictor variables. In other words, using the sample data values of X_1, X_2, and Y, a mathematical rule is established which best generalizes the relationship between these variables. With this rule, when new sets of values for X_1, and X_2 are obtained, the corresponding Y values can be calculated, hence predicted. Several concerns guide the creation of this predictive rule or regression model.

A first concern is the order or degree of the predictor variables in the regression equation. If any of the independent variables in a regression model exhibit a curvilinear rather than linear relationship with the predicted variable, the regression equation must contain a higher power term for that independent variable (e.g., a quadratic term such as X_2^2). Raw data frequency graphs are especially instructive in this context.

The independence of the predictor variables is another concern. If, as in the case of soil fertility and soil drainage, the two variables are meaningfully related (see Table 5.7), an interaction term must be provided in the regression equation (e.g., $X_1 X_2$). The absence of this interaction factor implies that soil fertility (X_1) determines (or effects) the extent of archaeological material independently of the affect of soil drainage (X_2).

However, when the predictor variables are highly correlated, redundant information may compromise the utility of the regression model. This effect is known as multicollinearity. In particular, highly correlated independent variables may yield an excellent predictive model - but only for the range of sample data used in its formulation. The predictive power of the regression model is directly affected.

These analytic concerns - the degree of independent variable interaction and the order of the polynomial function - can be resolved in part by statistical evaluation. The R^2 criterion and stepwise regression were used to analyze the New Hanover County data. Both of these techniques react to a set of potential independent variables, seeking that subset which best describes or predicts values of the dependent variable. Stepwise regression and the R^2 criterion are especially useful in modeling situations more complex than New Hanover County.

The R^2 criterion, or coefficient of determination, is a measure that describes the relative reduction in the sample (Y_i) variation ascribed to the use of the regression model. When R^2 is expressed as a ratio of sums of squares, the denominator, $\Sigma(Y_i-Y)^2$, is the sum of the squared deviations for the observed Y_i (hence, it is a measure of the total variation for the observed Y_i). The numerator in the ratio defining R^2 is the total variation for Y_i as determined by the regression model: $\Sigma(Y_i-\bar{Y})^2$. Values of R^2 range $0 \leq R^2 \leq 1$.

Table 5.7 Correlation coefficients (r)

	DRAINAGE (X_2)	EXTENT (Y)
FERTILITY (X_1)	.782 (.0001)	.538 (.02)
DRAINAGE (X_2)		.597 (.0009)

n = 18; significance levels given in parentheses

Table 5.8 Correlation coefficients (r) (corrected data)

	DRAINAGE (X_2)	EXTENT (Y)
FERTILITY (X_1)	.789 (.0002)	.83 (.0001)
DRAINAGE (X_2)		.686 (.002)

N = 17; significance levels given in parentheses

Stepwise regression, on the other hand, evaluates the set of independent variables by sequentially testing and eliminating single variables; statistical testing of the B coefficients establishes those variables which contribute to the model. Specifically,

H_0: $B_i = 0$ is tested against

H_1: $B_i \neq 0$ for each model, using the t-test

The set of predictor variables for the New Hanover County regression exercise had been effectively established during the first two predictive modeling strategies (see above). However, successful formulation of the mathematical expression relating these variables relied upon R^2 and stepwise regression. These procedures were applied to the sample data presented in Table 5.6, using options in the SAS computer library. Inspection of raw data plots suggested that the independent variable set should include for this regression analysis: X_2 (fertility), X_1^2, X_2 (drainage), X_2^2, and $X_1 X_2$.

Alternative regression models incorporating these variables were produced using the R^2 criterion. Three models yielded equivalent R^2 values, 0.64, which was the highest R^2 measure achieved.

The simplest of these equations was:

$$Y = B + B_1X_1 + B_2X_1^2 + B_3X_2 + B_4X_1X_2; \quad R^2 = 0.641 \qquad \text{(Eq. 5.5)}$$

This equation was selected for further analysis. The New Hanover County data (Table 5.6) defined the following B values for this model:

$$Y = 0.52 + 1.53X_1 - .01X_1^2 + .08X_2 - .005X_1X_2 \qquad \text{(Eq. 5.6)}$$

The relatively low R^2 value achieved by this model suggested that it might contain weaknesses. Further statistical and practical evaluations confirmed this hypothesis. T-tests of H_0: $B_i = 0$ versus H_1: $B_i > 0$ indicated that only B_1 and B_2 (for X_1 and X_1^2, respectively) were significant at $\alpha = 0.1$. For the remaining B values, the null hypothesis could not be rejected.

In addition to these statistical failings, the model generated predictions which were at variance with known site distributions. In particular, when drainage and fertility were assigned high values, the value for Y became low. This prediction conflicted with patterns revealed by lower level data analyses (cf. Tables 5.3 and 5.4). It was suspected that the problems with Equation 5.6 might be partially a result of the small size of the sample used as input to the regression procedure. Further comparisons of predicted Y values suggested that Equation 5.6 might be modeling random noise in this data, rather than underlying relationships among the predictor and predicted variables. An examination of the raw data offered at least one opportunity for tactical regrouping. Data case 11 (Table 5.6) did not conform to the more general patterns exhibited by the survey data. Some insight into this anomaly was gained from the knowledge that Survey Unit 36 (case 11) was located in a cultivated field. The archaeological site in this field (31NH616) had been collected for several years by the landowner. It can thus be argued that the amount of archaeological material recovered from this site was biased. For this reason, this data case was rejected as an analytically distorting outlier. The data set formed by exclusion of this case was then analyzed as before.

The effect of case 11 was immediately apparent from recalculated correlation coefficients (Table 5.8; cf. Table 5.7). The association between soil fertility and the extent of archaeological material in particular was much more evident. The following three data plots--fertility vs. drainage, fertility vs. extent, and drainage vs. extent (Figure 5.2)--illustrate these variable relationships.

Based on this revised data, a second regression model for archaeological resources in New Hanover County was developed. This model was selected by systematically evaluating equations that (1) had relatively high R^2 values, and that (2) were not responding primarily to data noise; i.e., that produced predictions consistent with known site distributions. The regression model best conforming to these criteria was:

$$Y = 2.38 + .616X_1 + .04X_2 \qquad \text{(Eq. 5.7)}$$

A graphic representation of this model is shown in Figure 5.3. The R^2 value for Equation 5.7 was 0.69; the global F-test was significant at α=0.001. The 90% confidence intervals around predicted values of the Y (the extent of archaeological material) are shown in Table 5.9.

This model should prove to be a powerful planning tool. If, for example, three alternative construction alignments are proposed for a highway to be built within some portion of New Hanover County, Equation 5.7 can be used to determine the relative impacts to archaeological resources of each alignment. The area to be affected can be subdivided into 100m x 200m units, and the average drainage character and soil fertility for each such unit can be determined from soils maps of the region. Equation 5.7 can then be used to estimate the areal extent of the archaeological materials to be expected within each of these same units. Through simple summation procedures, the probable impact to archaeological resources of each alignment can then be estimated and appropriate conservation procedures implemented.

It should be emphasized, however, that the regression model developed here possesses similar limitations to those of the descriptive model developed in the previous section. First, it is based on a small-scale survey. Second, its reliability when applied to unsurveyed areas is limited by the accuracy of currently available soils data. Therefore, the model should be subjected to field tests when used as a planning tool for specific project areas. In no instance should the use of regression models to predict the distribution of archaeological resources be used as clearance for areas that have not undergone surveys.

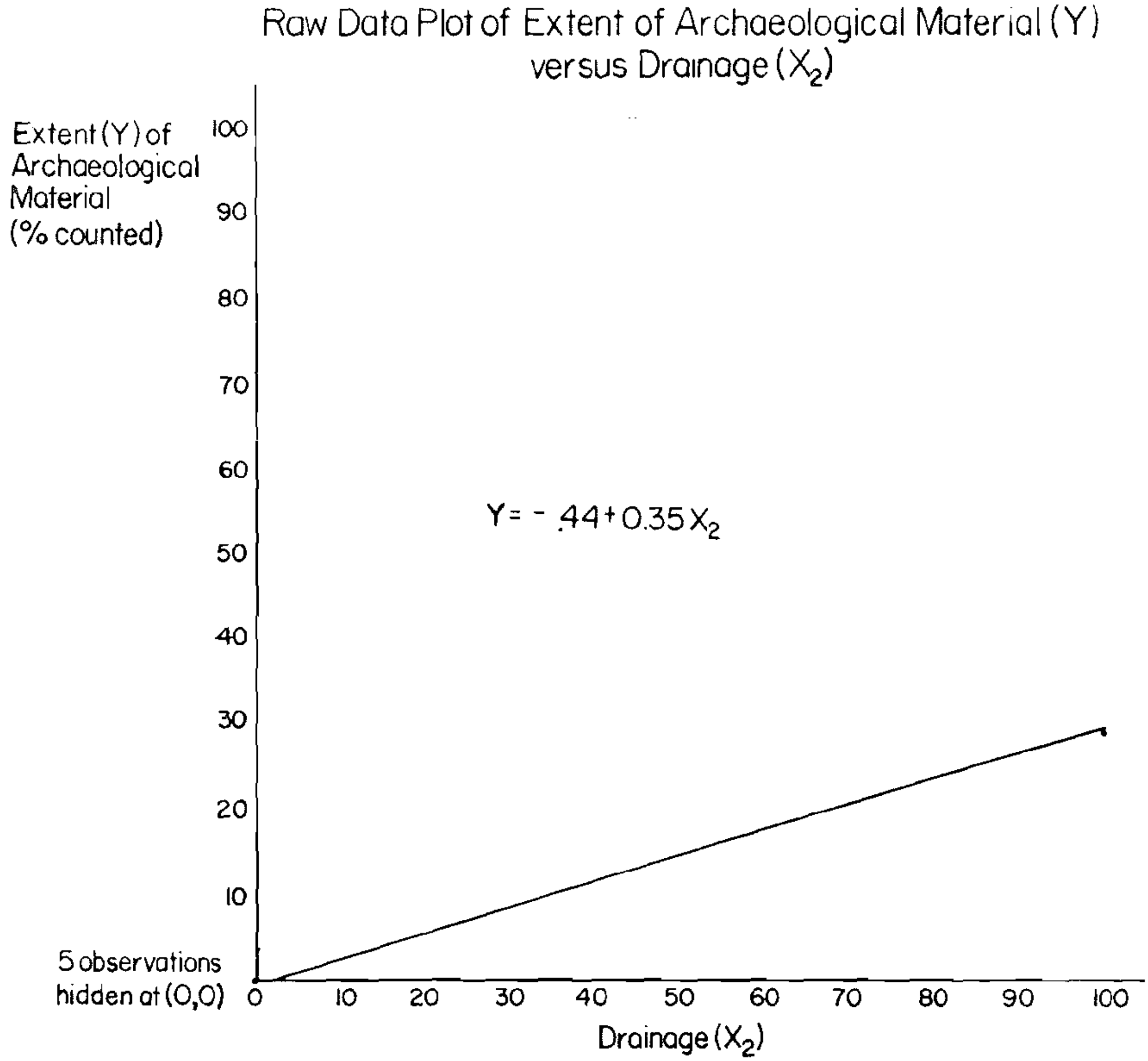
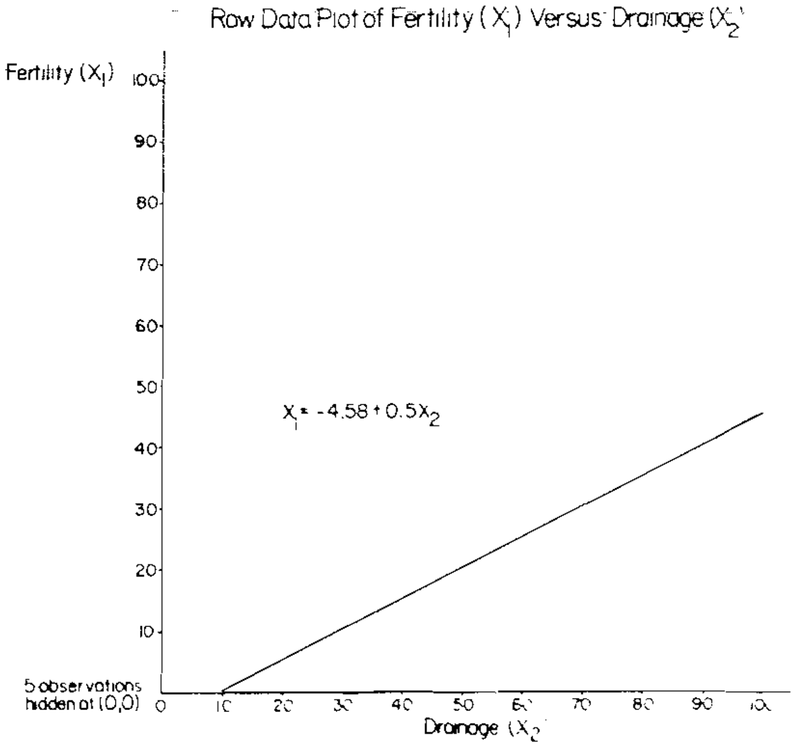


Figure 5.2 Raw data plots of X_1 (fertility) vs. Y (amount of archaeological material)



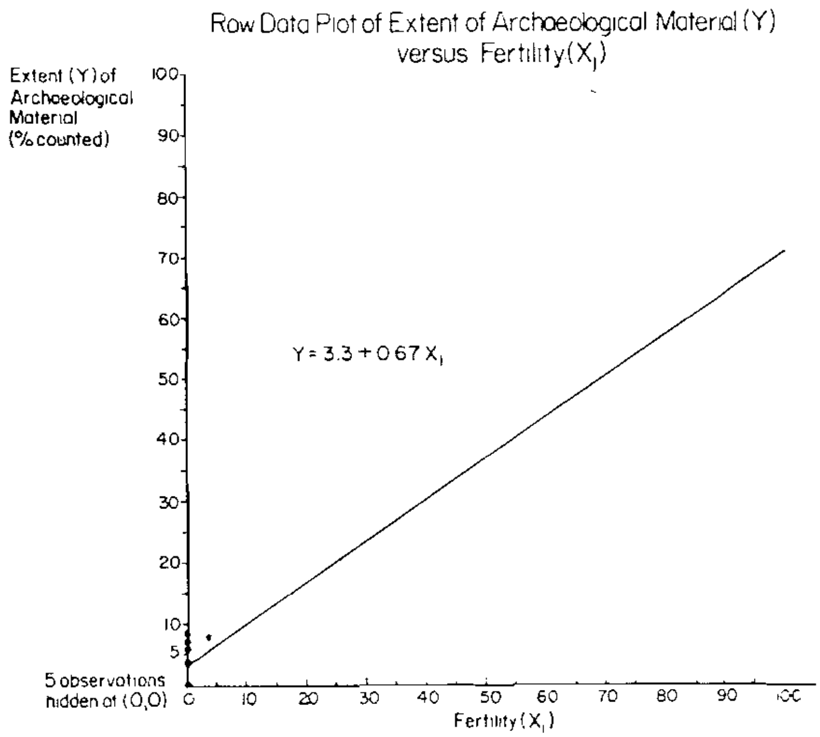


Figure 5.2 (continued) Raw data plots of X_2 (drainage) vs. Y; and X_1 (fertility) vs. X_2 (drainage)

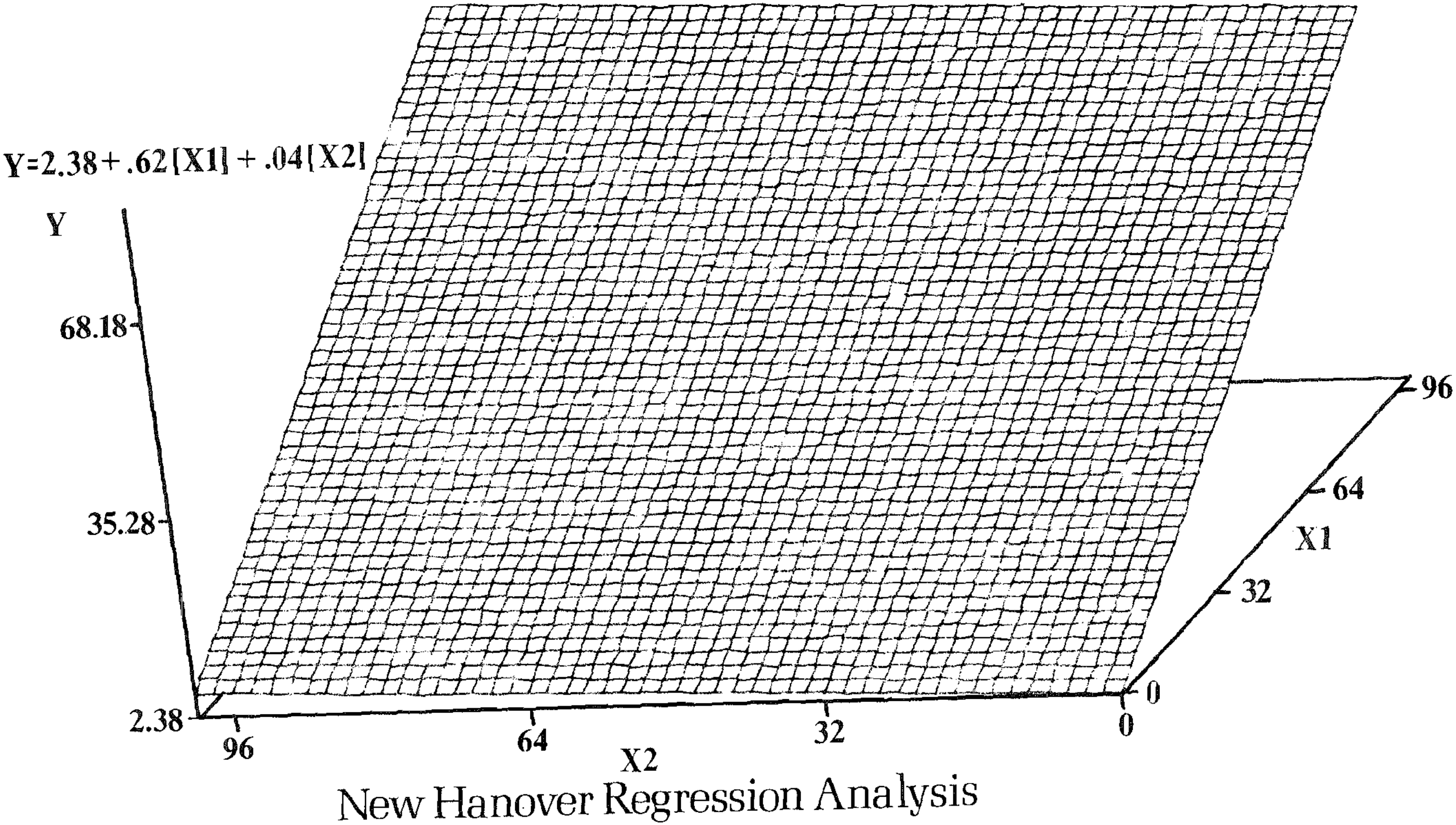


Figure 5.3 Response surface of final regression model

Table 5.9 90% Confidence intervals around predicted values of the extent of archaeological material (Y)

$$Y = 2.38 + .616X_1 + .04X_2$$

Case	Observed Value	Predicted Value	Lower 90% Confidence Interval	Upper 90% Confidence Interval
1	7.0	5.0	0.0	28.5
2	8.0	5.6	0.0	27.2
3	37.0	22.1	0.5	43.7
4	8.0	5.9	0.0	31.9
5	20.0	32.2	10.0	54.3
6	4.0	3.4	0.0	25.2
7	0.0	2.3	0.0	24.5
8	15.0	39.1	16.3	61.9
9	29.0	44.4	21.0	67.8
10	75.0	47.9	23.9	71.9
12	0.0	2.3	0.0	24.5
13	6.0	4.2	0.0	26.4
14	30.0	20.4	0.0	41.9
15	0.0	2.3	0.0	24.5
16	0.0	2.3	0.0	24.5
17	0.0	2.3	0.0	24.5
18	0.0	2.3	0.0	24.5

CHAPTER SIX

SUMMARY AND CONCLUSIONS

The New Hanover County Predictive Models Archaeological Research Project involved the development and evaluation of various approaches to the problem of predicting archaeological site distributions for areas with inadequate or incomplete site location data.

During the initial project phase, a descriptive predictive model was developed for New Hanover County. This model was based on conventional archaeological site file data. It employed descriptive statistics and quantitative parameter definitions to identify areas within the county having high, medium, low, and nonprobabilities of site occurrence. In important respects the model developed conformed to current practice in the field of predictive modeling.

During the second project phase, this model was tested in the field and proved to have significant limitations. First, its predictions concerning the distribution of archaeological resources within the county contained inaccuracies; and second, the general congruence between the model's predictions and actual site distributions was not confirmed when subjected to statistical tests. Because of these difficulties, the model failed to clearly delineate those areas which contain the majority of New Hanover County's archaeological resources.

The problems that were encountered using the preliminary predictive model can be attributed to two sources, both of which are inherent to the type of data used in the model's formulation. First, these data contained unrecognized biases that led to misleading predictions concerning the distribution of sites. Second, the data were recorded in the conventional format (i.e., the site was the basic recording unit, and was perforce the unit of analysis as well). As a result, the unit of interest for predictive modeling purposes (a unit or units of land surface) was different from the unit of analysis (the archaeological site). This difficulty led to further predictive errors.

Because these difficulties were encountered during preliminary model formulation and testing, a third project phase involved the exploration of alternative approaches to the problem of predicting archaeological site distributions. The survey conducted for the purpose of testing the initial model provided a data base for these latter models. Since these data were collected using a probabilistic sampling design, they can be considered a representative sample of the county as a whole. Because data were recorded for all shovel tests and survey units regardless of whether archaeological materials were present or not, shovel tests or survey units rather than sites were the basic recording units; the unit of interest for predictive modeling purposes was thus consistent with the unit of analysis in the data set.

Two analytical approaches were used to formulate predictive generalizations concerning site distributions from these survey data. The first approach employed descriptive data summaries. A clear, statistically significant delineation of probability areas was achieved. Two qualitative variables - the suitability of soils for agriculture and soil drainage character - were found to be excellent predictors of this delineation, allowing the pattern for sampled soil types to be extrapolated to unsampled soil types. The resulting model can thus be utilized for all or part of New Hanover County using the data presently available on detailed soil maps. The model should be useful for planning and review purposes because it is based on currently available data, and because it is relatively simple to use. However, its accuracy is limited by two factors: the accuracy of county soil maps for the county, and the small scale of the survey from which it was derived. As a result, field surveys should be conducted to establish the reliability of the model when it is applied to specific project areas.

If confirmed by such tests, this latter model should prove to be an effective review and planning tool for New Hanover County. As a generalized type of model, however, it has limitations. Many areas for which predictive models are desired may not share New Hanover County's topography. Where water sources are less prevalent and elevational variability more distinct, soil type may not have been the prime determinant of site location; in such areas, accurate prediction may require the use of more complex models, incorporating several variables. Qualitative data analyses, of the kind used to generate the second model, become increasingly unsuitable under such circumstances. This is because the simultaneous effects of numerous variables may become cumbersome to analyze. More rigorous statistical modeling procedures are designed to deal with these complex predictive problems; and should be used in such cases.

A third and final predictive model using a statistical approach was developed. The specific approach employed for this purpose involved regression analysis; the same variables used in the previous model (soil fertility and drainage character) were selected as independent variables, and the extent of archaeological material to be expected within any given unit of land surface was selected as the dependent variable. In this case, an arbitrarily defined 100m X 200m tract was the unit of analysis. The equation,

$$Y = 2.38 + .616X_1 + .04X_2 \qquad \text{(Eq. 5.7)}$$

was derived. If values for X and X (fertility and drainage character, respectively) are known for any specific area, the equation will estimate the extent of the archaeological material (Y) to be expected within a 100m X 200m unit at that locality. Furthermore, a statistically derived range of probable error around the estimate can be calculated. With these capabilities, the model should prove to be a useful planning and review tool.

Several factors limit the rigor of the regression model, and should be emphasized. As with the previous descriptive model, the regression model is derived from a small-scale survey, and its accuracy is limited by the currently available data concerning the distribution of soil units within New Hanover County. It should thus be subjected to field testing when applied to specific project areas. Unlike the previous model, however, the regression model does not rely upon descriptive, qualitative data analyses. The procedures used in its formulation allow the incorporation of as many variables as considered desirable and should thus be appropriate for model formulation in other areas, regardless of the complexity of the predictive problem.

The results of the New Hanover County Project have important implications concerning the development and use of predictive models. Many of the models currently being developed share to some extent the limitations of the initial model developed here. It is therefore probable that many of these models contain misleading predictions concerning the distribution of archaeological resources. However, it is important that reliable predictive models be developed for areas that will undergo extensive construction activity. Only then will future economic development proceed in a manner consistent with the preservation of archaeological resources.

The results of the present project indicate that to contribute to this latter goal, predictive models should meet certain criteria:

1) Predictive models should be derived from probabilistic, representative samples of units of land surface, rather than from nonprobabilistic and potentially nonrepresentative samples of archaeological sites.

2) Predictive models derived using mathematical analytical procedures capable of handling the complexities of multivariate prediction may be required for many areas.

3) Statistically derived error estimates based on the variance of the sample used in the model's formulation should be provided, when appropriate.

Predictive models that conform to these criteria will initially be somewhat more expensive to develop than less rigorous models. However, they will be more efficient for planning purposes and should therefore facilitate the compliance process.

The results of the present study have further implications for the use of predictive models within the more general context of historic preservation compliance activities. Predictive models consist of generalizations about the distribution of sites in areas for which inadequate site location data are available. As such, they are inherently subject to error. Thus, the development of a predictive model _per se_ is not equivalent to an inventory of all significant historic and prehistoric properties in an impact area. This is true no matter how rigorous the procedures used in its formulation. Predictive models should instead be used in the planning process to minimize project impacts, thereby allowing additional time for the compliance process and aiding in the avoidance of project delays. In this way, predictive models can reduce compliance costs and minimize the likelihood that important archaeological properties will be lost in the course of ongoing economic development.

REFERENCES CITED

Bain, G.L.
1970 Geology and groundwater resources of New Hanover County, North Carolina. Department of Water and Air Resources, Raleigh, N.C., Groundwater Bulletin 17.

Bass, Q.
1977 Prehistoric settlement and subsistence patterns in the Great Smoky Mountains. Unpublished Masters thesis, Department of Anthropology, University of Tennessee, Knoxville.

Binford, L.R.
1964 Archeological and ethnohistoric investigation of cultural diversity and progressive development among aboriginal cultures of coastal Virginia and North Carolina. Ph.D. dissertation, University of Michigan, University Microfilms, Ann Arbor.

Blalock, H.M.,Jr.
1972 Social statistics (second ed.). McGraw-Hill, New York.

Broyles, B.J.
1971 Second preliminary report: the St. Albans site, Kanawha County, West Virginia. West Virginia Geological and Economic Survey Report of Archeological Investigations, No. 3.

Caldwell, J.R.
1958 Trend and tradition in the prehistory of the Eastern United States. American Anthropological Association, Memoir 88.

Carniero, R.
1970 A theory of the origin of the state. Science 169: 733-738.

Chapman, J.
1973 The Icehouse Bottom site, 40MR23. Department of Anthropology, University of Tennessee, Knoxville, Report of Investigations No. 13.
1975 The Rose Island site, Department of Anthropology, University of Tennessee, Knoxville, Report of Investigations No. 14.
1977 Archaic period research in the lower Little Tennessee River Valley. Department of Anthropology, University of Tennessee, Knoxville, Report of Investigations No. 18.

Clausen. C.J., A.D. Cohen, C. Emiliani, J.A. Holman and J.J. Stipp
1979 Little Salt Spring, Florida: a unique underwater site. Science 203:609-614.

Cochran, W.G.
1977 Sampling techniques (third ed.). J. Wiley and Sons, New York.

Coe, J.L.
1952 The cultural sequence of the Carolina piedmont. In Archeology of the Eastern United States, edited by J.B. Griffin, University of Chicago Press, Chicago.
1964 The formative cultures of the Carolina piedmont. Transactions of the American Philosophical Society, volume 54, part 5.

Coe, J.L. and O.F. McCormick, III.
1970 Archeological resources of the New Hope Reservoir area, North Carolina. Ms. on file, Research Laboratories of Anthropology, University of North Carolina, Chapel Hill.

Cooper, A.W.
1964 Smith Island and the Cape Fear Peninsula. North Carolina Academy of Science, Raleigh.

Cridlebaugh, P.
1977 An analysis of the Morrow Mountain component at the Icehouse Bottom site and a reassessment of the Morrow Mountain complex. Unpublished Masters thesis, Department of Anthropology, University of Tennessee, Knoxville.

Dickens, R.S., Jr.
1976 Cherokee Prehistory. University of Tennessee Press, Knoxville.

Ferguson, L.S. and R.J. Widmer
1976 Archeological examination of a transect through the Middle Savannah River Valley: the Bobby Jones expressway, Richmond County, Georgia. Institute of Archeology and Anthropology, University of South Carolina, Research Manuscript Series Number 89.

Fish, F.F.
1968, A catalog of the inland fishing waters in North Carolina. Division of Inland Fisheries, Raleigh, North Carolina.

Ford, R.I.
1974 Northeastern archeology: past and future directions. Annual Review of Anthropology 3:385.

Funderburg, J.B.
1955 The amphibians of New Hanover County, North Carolina. Journal of the Elisha Mitchel Scientific Society 71:19-28.

Gardner, W.M.
1977 Flint Run Paleoindian complex and its implication for Eastern North American prehistory. In Amerinds and their paleoenvironments in northeastern North America, edited by Newman and Salwen, Annals of the New York Academy of Sciences 288:257-263.

Griffin, J.B. (editor)
1952 Archeology of the Eastern United States. University of Chicago Press, Chicago.
1967 Eastern North American archeology: a summary. Science 156: 175-191.

Haag, W.G.
1958 The archeology of coastal North Carolina. Louisianna State University Coastal Studies Series Number 2.

Harris, R.J.
1975 A primer of multivariate statistics. Academic Press, New York.

Keel, B.C.
1976 Cherokee Archaeology University of Tennessee Press, Knoxville.

Larson, A.J.
1958 An agricultural geography of New Hanover and Pender counties, North Carolina. Unpublished Masters thesis, Departmenc of Geology and Geography, University of North Carolina, Chapel Hill.

Loftfield, T.C.
1976 "A briefe and true report. . ."; an archeological interpretation of the southern North Carolina coast. Unpublished Ph.D. dissertation, Department of Anthropology, University of North Carolina, Chapel Hill.

Mathis, M.A. (assembler)
1979 North Carolina statewide archeological survey: an introduction and application to three highway projects in Hertford, Wilkes, and Ashe counties. North Carolina Archeological Council, Publication 11.

McClave, J.T. and P.G. Benson
1979 Statistics for business and economics (revised ed.). Dellen, San Francisco.

Mendenhall, W., L. Ott and R.L. Scheaffer
1971 Elementary sampling. Wadsworth, Belmont, California.

Neter, J. and W. Wasserman
1974 Applied linear statistical models. Richard Irwin, Homewood, Illinois.

Newman, W.S. and B. Salwen (editors)
1977 Amerinds and their paleoenvironments in northeastern North America. Annals of the New York Academy of Sciences, Volume 288.

Nie, N.H., C.H. Hull, J.G. Jenkins, K. Steinbrenner and D.H. Bent
1975 Statistical package for the social sciences (SPSS) (second edition). McGraw-Hill, New York.

N.C. Department of Transportation
1976 Wilmington southeast outer loop from Shipyard Boulevard to U.S. 17 North: environmental impact statement. Ms. on file, N.C. Dept. of Transportation.

Pearson, T.G., C.S. Brimley and H.H. Brimley
1959 Birds of North Carolina. Bynum, Raleigh, North Carolina.

Perkinson, P.
1971 North Carolina fluted projectile points-survey report number one. Southern Indian Studies, 23:3-40.
1973 North Carolina fluted projectile points-survey report number two. Southern Indian Studies, 25:3-60.

Phelps, D.S.
1978 Archeological studies in the northern coastal plain of North Carolina. North Carolina Archeological Council, Publication 6.
1980 Archeology of N.C.'s coast and coastal plain: problems and hypotheses. Paper presented at "The Prehistory of North Carolina: an Archeological Symposium", March 28, 1980, Raleigh.

Potter, E.F., J.F. Parnell and R.P. Teulings
1980 Birds of the Carolinas. University of North Carolina Press, Chapel Hill.

Radford, A.E., H.E. Ahles and C.R. Bell
1964 Manual of the vascular flora of the Carolinas. University of North Carolina Press, Chapel Hill.

Richards, H.G.
1950 Geology of the coastal plain of North Carolina. Transactions of the American Philosophical Society, Volume 40, part 1.

Rights, D.L.
1957 The American Indian in North America. J.F. Blair, Winston-Salem, North Carolina.

Sharp, B.
1954 A new geography of North Carolina. Sharpe, Raleigh, N.C.

Shelford, V.E.
1963 The ecology of North America. University of Illinois Press, Urbana, Illinois.

Smith, G.P.
1965 An archeological survey of the New Hope Valley. Unpublished Masters thesis, Department of Anthropology, University of North Carolina, Chapel Hill.

Smith H.M.
1907 The fishes of North Carolina. E.M. Uzzell.

South, S.
1976 An archeological survey of southeastern coastal North Carolina. Ms. on file, Institute of Archeology and Anthropology, University of South Carolina.

Thomas, R., D.R. Griffith, C.L. Wise and R.E. Artusy, Jr.
1977 Environmental adaptation on Delaware's coastal plain. In Archeology of Eastern North America (Vol. 3).

U.S. Army Corps of Engineers
1976 Wilmington Harbor: feasibility report for channel deepening for navigation. Ms. on file, N.C. Archeology Branch.

Von Oesen, H. and Associates, Inc.
1976 Appendix T to greater Wilmington area 201 facilities: archeological and vegetative surveys. Ms. on file, Henry Van Oesen and Associates, Inc., Wilmington, N.C.

Weaver, A.
1977 Soil survey of New Hanover county, North Carolina. U.S. Department of Agriculture, Soil Conservation Service.

Wilde-Ramsing, M.
1978 A report on the New Hanover county archeological survey: a C.E.T.A. project, August 1977-July 1978. Ms. on file, N.C. Underwater Archeology Branch, Ft. Fisher, North Carolina.

Williams, S. and J. Stoltman
1965 An outline of southeastern United States prehistory with particular emphasis on the paleoindian era. In the quaternary of the United States, edited by H.E. Wright and D.D. Froy, Princeton University Press, Princeton.

Zim, H.S. and D.F. Hoffmeister
1955 Mammals: a guide to familar American species. Simon and Schuster, New York.

APPENDICES

APPENDIX A

This appendix lists important plant species in each of the ecological communities represented in New Hanover County. This appendix is not intended to be a complete listing of all the botanical components of each community. Sources for this information are Cooper (1964), Funderburg (1974), NC Department of Transportation (1976), Radford et al. (1964), Shelford (1963), Von Oesen (1976), and Wilde-Ramsing (1978). Both the common name and its Latin equivalent are provided for each species.

MARITIME FOREST

Live Oak	Quercus virginiana
Cabbage Palmetto	Sabal palmetto
Red bay	Persea borbonia
Carolina laurel cherry	Prunus caroliniana
American holly	Ilex opaca
Yaupon	Ilex vomitoria
Wax myrtle	Myrica cerifera
Wild olive	Osmanthus americana
Loblolly pine	Pinus taeda
Catbriar	Smilax rotundifolia
Southern bayberry (occasional)	Myrica heterophylla
French mulberry (occasional)	Morus rubra
Virginia creeper	Parthenocissus quinquefolia
Dogwood	Cornus florida

BEACH DUNE SCRUB

American beachgrass	Ammophila breviligulata
Bitter panic grass	Panicum amarum
Salt meadow cord grass	Spartina patens
Sea oat	Uniola paniculata
Broomsedge	Andropogon sp.
Sea myrtle	Baccharis halimifolia
Groundsel	Senecio vulgaris
Marsh elder	Iva imbricata
Yucca (occasional)	Yucca aloifolia
Southern bayberry (occasional)	Myrica heterophylla

POND VEGETATION

SUBMERGED PLANTS

Bladderwort	*Utricularia sp.*
Waterweed	*Elodea nuttalii*
Water nymph	*Najas sp.*
Egeria	*Egeria densa*
Pondweed	*Potamogeton sp.*
Widgeon grass	*Ruppia maritima*

FLOATING PLANTS

Water lily	*Nymphaea sp.*
Spatterdock	*Nuphar luteum*
Water shield	*Brasenia schreberi*
Duck weed	*Lemna sp.*
Alligator weed	*Alternanthera philoxeroides*

FRESHWATER MARSH

Giant reed grass	*Spartina cynosuroides*
Waterweed	*Elodea nuttalii*
Spatterdock	*Nuphar luteum*
Alligator weed	*Alternanthera philoxeroides*
Sedge	*Cyperus sp.*
Bullrush	*Scirpus sp.*
Eleocharis	*Eleocharis sp.*
Sawgrass	*Cledium jamaicense*
Wild rice	*Zizania sp.*

SALTWATER MARSH

REGULARLY FLOODED LOW MARSH

Smooth cord grass	*Spartina alterniflora*
Salt meadow cord grass (some)	*Spartina patens*
Sea ox-eye	*Borrichia frutescens*
Black needle rush	*Juncus roemerianus*
Cat tail	*Typha sp.*

SALTWATER MARSH CONTINUED

HIGHER MARSH (SOME FLOODING)

Salt measow cord grass	*Spartina patens*
Salt grass	*Distichlis spicata*
Suaeda	*Suaeda linearis*
Sedge	*Cyperus sp.*

RARELY FLOODED SHRUB ZONE

Sea myrtle	*Baccharus halimifolia*
Marsh elder	*Iva frutescens*
Wax myrtle	*Myrica cerifera*
Yaupon	*Ilex vomitoria*
False willow	*Baccharus anqustifolia*

SWAMP FOREST

Bald cypress	*Taxodium distichum*
Pond cypress	*Taxodium ascendens*
Black gum	*Nyssa sylvatica var. biflora*
Tupelo gum	*Nyssa aquatica*
Water Ash	*Fraxinus caroliniana*
Red maple	*Acer rubrum*
Water hickory	*Carya aquatica*
Swamp chestnut oak	*Quercus michauxii*
Sycamore	*Platanus occidentalis*

SLIGHTLY DRIER BOTTOMLAND HARDWOOD FOREST

Loblolly pine	*Pinus taeda*
Pond pine	*Pinus serotina*
Tulip poplar	*Liriodendron tulipifera*
Royal fern	*Osmunda regalis*
Cinnamon fern	*Osmunda cinnamomea*
Sweet gallberry	*Ilex coriacea*
Bamboo briar	*Smilax laurifolia*
Catbriar	*Smilax rotundifolia*
Swamp rose	*Rosa palustris*
Virginia willow	*Itea virginica*
Wax myrtle	*Myrica cerifera*
Water willow	*Decodon verticillatus*
Lizard's tail	*Saururus cernuus*
Pennywort	*Obolaria virginica*
Spanish moss	*Tillandsia usneoides*

PINE SAVANNAH

Longleaf pine	Pinus palustris
Loblolly pine	Pinus taeda
Sand myrtle	Leiophyllum buxifolium
Wax myrtle	Myrica cerifera
Wiregrass	Aristida stricta
Broomsedge	Andropogon sp.
Virginia chain fern	Woodwardia virginica
Goldenrod	Solidago sp.
Meadow beauty	Rhexia sp.
Yellow-eyed grass	Xyris sp.
Wild verbena	Verbena sp.
Pitcher plant	Sarracenia sp.
Pyxie moss	Pyxidanthera barbulata

POCOSIN

Pond pine	Pinus serotina
Loblolly bay	Gordonia lasianthus
Titi	Cyrilla racemiflora
Red bay	Persea borbonia
Bamboo briar	Smilax laurifolia
Sweet bay	Magnolia virginiana
Fetterbush	Lyonia lucida
Sweet gallberry	Ilex coriacea
Sweet pepperbush	Clethra alnifolia
Lambkill	Kalmia angustifolia
Cane	Arundinaria gigantea
Red maple	Acer rubrum
Dahoon	Ilex cassine

MIXED PINE-HARDWOOD FOREST

Loblolly pine	Pinus taeda
Live oak	Quercus virginiana
White oak	Quercus alba
Southern red oak	Quercus falcata
Willow oak	Quercus phellos
Water oak	Quercus nigra
Swamp chestnut oak	Quercus michauxii
Sweet gum	Liquidambar styraciflua
Mockernut hickory	Carya tomentosa
Dogwood	Cornus florida
Tulip poplar	Liriodendron tulipifera
Sweet pepperbush	Clethra alnifolia
Wild olive	Osmanthus americana
Catbriar	Smilax rotundifolia
Muscadine	Vitis rotundifolia

MIXED PINE-HARDWOOD FOREST CONTINUED

Bamboo briar	*Smilax laurifolia*
Wild ginger	*Hexastylis sp.*
Yellow jessamine	*Gelsemium sempervirens*
American holly	*Ilex opaca*
Titi	*Cyrilla racemiflora*
Southern bayberry	*Myrica heterophylla*
Yaupon	*Ilex vomitoria*
Sweet gallberry	*Ilex coriacea*

LONGLEAF AND LOBLOLLY PINE PLANTATIONS

Longleaf pine	*Pinus palustris*
Loblolly pine	*Pinus taeda*

UNDERSTORY, IF ANY

Turkey oak	*Quercus laevis*
Southern bayberry	*Myrica heterophylla*
American holly	*Ilex opaca*
Live oak	*Quercus virginiana*
Running oak	*Quercus pumila*

LONGLEAF PINE-TURKEY OAK FOREST

Longleaf pine	*Pinus palustris*
Turkey oak	*Quercus laevis*
Bluejack oak	*Quercus incana*
Live oak	*Quercus virginiana*
Wiregrass	*Aristida stricta*
Sweet pepperbush	*Clethra alnifolia*
Wild olive	*Osmanthus americana*
Catbriar	*Smilax rotundifolia*
Muscadine	*Vitis rotundifolia*
Bamboo briar	*Smilax laurifolia*
Wild ginger	*Hexastylis sp.*
Yellow jessamine	*Gelsemium sempervirens*

HARDWOOD FOREST

Water oak	*Quercus nigra*
Post oak	*Quercus stellata*
Dogwood	*Cornus florida*
White oak	*Quercus alba*
Red maple	*Acer rubrum*
Tulip poplar	*Liriodendron tulipifera*

HARDWOOD FOREST CONTINUED

Sweet gum	*Liquidambar styraciflua*
Hickory	*Carya sp.*
Black gum	*Nyssa sylvatica*
Longleaf pine	*Pinus palustris*
Loblolly pine	*Pinus taeda*
Sweet pepperbush	*Clethra alnifolia*
Wild olive	*Osmanthus americana*
Catbriar	*Smilax rotundifolia*
Muscadine	*Vitis rotundifolia*
Bamboo briar	*Smilax laurifolia*
Wild ginger	*Hexastylis sp.*

APPENDIX B

This appendix lists important animal species in each of the ecological communities represented in New Hanover County. This appendix is not intended to be a complete listing of all the faunal components of each community. The sources for this information are Pearson *et al.* (1959), Potter *et al.* (1980), Shelford (1963), Smith (1907), and *Zim* and Hoffmeister (1955). Both the common name and its Latin equivalent are provided for each species.

TIDAL MARSH

Raccoon	*Procyon lotor*
Mink	*Mustela vison*
Marsh rabbit	*Sylvilagus palustris*
Muskrat	*Ondatra zibethicus*
Marsh rice rat	*Oryzomys palustris*
Fiddler crab	*Uca rapax*
Other crabs	
Oysters	
Clams	
American alligator	*Alligator mississipiensis*
Eastern cottonmouth snake	*Agkistrodon piscivorus*
Yellow-bellied turtle	*Pseudemys scripta*
Carolina diamondback terrapin	*Mala clemmys centrata*
Northern diamondback terrapin	*Mala clemmys centrata concentrica*
Eastern mud turtle	*Kinosternon subrubrum*
Belted kingfisher	*Megaceryle alcyon*
Great blue heron	*Ardea herodias*
Marsh hawk	*Circus cyaneus*
Mallard	*Anas platyrhynchos*
Black duck	*Anas rubripes*
Herring gull	*Larus argentatus*
Ring-billed gull	*Larus delawarensis*
American bittern	*Botarus lentiginosus*
American oystercatcher	*Haematopus palliatus*
Caspian tern	*Sterna caspia*

PONDS

IMPORTANT FISH SPECIES

Bowfin	*Amia calva*
Chain pickerel	*Esox niger*
Redfin pickerel	*Esox americanus*
Lake chubsucker	*Erimyzon sucetta*
Golden shiner	*Notemigonus crysoleucas*
Yellow bullhead	*Ictalurus natalis*
Mosquito fish	*Gambusia affinis*
Flier	*Centrarchus macropterus*
Warmouth	*Lepomis gulosis*
Pumpkinseed	*Lepomis gibbosus*
Bluegill	*Lepomis macrochirus*
Largemouth bass	*Micropterus salmoides*
Yellow perch	*Perca flavescens*
Tadpole madtom	*Schilbeodes furiosus*
Eastern starhead topminnow	*Fundulus dispar lineolatus*
Sheepshead minnow	*Cyprinodon variegatus*
Blue spotted sunfish	*Enneacanthus gloriosus*

IMPORTANT REPTILES

Eastern cottonmouth	*Agkistrodon piscivorus*
Watersnake	*Natrix sp.*
Southern cricket frog	*Acris gryllus crepitans*
Southern leopard frog	*Rana pipiens*
Bullfrog	*Rana catesbiana*
Snapping turtle	*Chrysemys picta*
Yellow-bellied turtle	*Pseudemys scripta*
Red-bellied turtle	*Pseudemys rubriventris*
Greater siren red-spotted newt	*Siren lacertina*
Salamanders	

PERMANENT BIRD LIFE

Whistling swan	*Olar columbianus*
Ring-billed gull	*Larus delawarensis*
Caspian tern	*Sterna caspia*
American coot	*Fulica americana*
Belted kingfisher	*Megaceryle alcyon*
Pied-billed grebe	*Podilymbus podiceps*
Osprey	*Pandion haliaetus*
Common snipe	*Gallinago gallinago*
American woodcock	*Scolopax minor*

(See also migratory bird list)

RIVERS AND CREEKS

ESTUARIES SERVE AS NURSERIES FOR	
Crabs, oysters, clams, shrimp	
Anchovy	*Anchoa hepsetus*
Bluefish	*Pomatomus salatrix*
Menhaden	*Brevoortia tyrannus*
Mullet	*Mugil brasiliensis*
Silver perch	*Bairdiella chrysoura*
Blueback herring	*Pomolobus aestivalis*
Atlantic sturgeon	*Acipenser oxyrinchus*
Tarpon	*Tarpon atlanticus*
Alewife	*Alose pseudoharengus*
American shad	*Alosa sapidissima*
FRESHWATER FISH INCLUDE	
Largemouth bass	*Micropterus salmoides*
Chain pickerel	*Esox niger*
Golden shiner	*Notemigonus crysoleucas*
Yellow perch	*Perca flavescens*
Black crappie	*Pomoxis nigromaculatus*
Warmouth	*Lepomis gulosis*
Longnose gar	*Lepisosteus osseus*
Bowfin	*Amia calva*
Channel catfish	*Ictalurus punctatus*
White catfish	*Ictalurus catus*
Yellow bullhead	*Ictalurus natalis*
Gizzard shad	*Dorosoma cepedianum*
Redbreast sunfish	*Lepomis auritus*
Bluegill	*Lepomis macrochirus*
Pumpkinseed	*Lepomis gibbosus*
Carp	*Cyprinus carpio*
CURRENTLY SUPPORTS	
White-tailed deer	*Odocoileus virginianus*
Black bear	*Ursus americanus*
IN RECENT PAST SUPPORTED	
River otter	*Lutra canadensis*
Muskrat	*Ondatra zibethicus*
Mink	*Mustela vison*
Beaver	*Castor canadensis*
Nutria	*Myocastor coypus*
UP TO EARLY HISTORIC PERIOD SUPPORTED	
Wolf	*Canis lupus*
Cougar	*Felix concolor*

RIVERS AND CREEKS CONTINUED

BIRDS WHICH EXPLOIT OR LIVE ALONG RIVER	
Brown pelican	Pelecanus occidentalis
Double crested cormorant	Phalacrocorax auritus
Herring gull	Larus argentatus
Laughing gull	Larus atricilla
Ring-billed gull	Larus delawarensis
Various grebes	
Canadian goose	Branta canadensis
Black skimmer	Rynchops niger
See migratory bird list	

PINE AND MIXED HARDWOODS

Opossum	Didelphis marsupialis
Raccoon	Procyon lotor
Gray squirrel	Sciurus carolinensis
Southern flying squirrel	Glaucomys volans
Eastern cottontail rabbit	Sylvilagus floridanus
White-tailed deer	Odocoileus virginianus
Bobcat	Lynx rufus
Green anole	Anolis carolinensis
Black snake	Coluber constrictor
Cornsnake	Elaphe guttata
Eastern kingsnake	Lampropeltis getulus
Southern copperhead	Agkistrodon contortrix
Slimy salamander	Plethodon glutinosus
Eastern box turtle	Terrapene carolina
Ground skink	Leiolepisma laterale
Canebrake rattlesnake	Sistrurus miliarius
Southern toad	Bufo terrestris
Oak toad	Bufo quercicus
Pinewoods tree frog	Hyla femoralis
Sharp-shinned hawk	Accipiter striatus
Cooper's hawk	Accipiter cooperii
Great horned owl	Bubo virginianus
Screech owl	Otus asio
Carolina wren	Thryothorus ludovicianus
Ruby-crowned kinglet	Regulus calendula
Golden-crowned kinglet	Regulus satrapa
Bobwhite	Colinus virginianis
Warblers	Dendroica sp.
Various woodpeckers	

BEACH DUNE SCRUB

Eastern cottontail rabbit	Sylvilagus floridanus
Raccoon	Procyon lotor
Opossum	Didelphis marsupialis
Black snake	Coluber constrictor
Land turtles	
Sea turtles (to lay eggs)	

FREQUENT BIRDS

Laughing gull	Larus atricilla
Herring gull	Larus argentatus
Boat-tailed grackle	Quiscalus major
Red-winged blackbird	Agelaius phoeniceus
Mockingbird	Mimus polyglottos
Warblers	Dendroica sp.
Common tern	Sterna hirundo
Black skimmer	Rynchops niger

IMPORTANT SALTWATER FISH SPECIES

American shad	Alosa sapidissima
Alewife	Alose pseudoharengus
Striped bass	Roccus saxatilis
Flatheaded catfish	Tachysurus felis
Croaker	Micropogan undulatus
Pigfish	Orthopristis chrysopterus
Spot	Leiostomus xanthurus
Tarpon	Tarpon atlanticus
Whiting	Menticierhus saxatilis
Bluefish	Pomatomus salatrix
Red drum	Sciaenops ocellatus

MARITIME FOREST

Gray fox	Urocyon cinereo argenteus
Raccoon	Procyon lotor
Oppossum	Didelphis marsupialis
Gray squirrel	Sciurus carolinensis
Cotton mouse	Peromyscus gossypinus
Otter (rare)	Lutra canadensis
Mink (rare)	Mustela vison
Great crested flycatcher	Myiarchus crinitus
Parula warbler	Parula americana
Boat-tailed grackle	Quiscalus major
Clapper rail	Rallus longirostris
Common egret	Casmerodius albus
Wood ibis	Mycteria americana
Diamondback terrapin	Mala clemmys centrata
Atlantic loggerhead turtle	Caretta caretta

MARITIME FOREST CONTINUED

ONCE REPRESENTED BUT NO LONGER PRESENT	
Black bear	Ursus americanus
Wolf	Canis lupus
White-tailed deer	Odocoileus virginianus

PINE SAVANNAH

Opossum	Didelphis marsupialis
Raccoon	Procyon lotor
Bobcat	Lynx rufus
Gray squirrel	Sciurus carolinensis
Fox squirrel	Sciurus niger
White-tailed deer	Odocoileus virginianus
Striped skunk	Mephitis mephitis
Rat snake	Elaphe obsotela
Black snake	Coluber constrictor
Canebrake rattlesnake	Sistrurus miliarius
Various species of toads, lizards and frogs	
Bobwhite	Colinus virginianis
Mourning dove	Zenaida macroura
Various species of hawks, owls, woodpeckers and songbirds	

SWAMP FOREST

Opossum	Didelphis marsupialis
Raccoon	Procyon lotor
Mink	Mustela vison
River otter	Lutra canadensis
Short-tailed shrew	Blarina brevicauda
Gray squirrel	Sciurus carolinensis
Southern flying squirrel	Glaucomys volans
Marsh rabbit	Sylvilagus palustris
Bobcat	Lynx rufus
Black bear	Ursus americanus
Gray fox	Urocyon cinereo argenteus
White-tailed deer	Odocoileus virginianus
Eastern mud snake	Farancia abacura
Rat snake	Elaphe obsotela
Southern copperhead	Agkistrodon contortrix
Canebrake rattlesnake	Sistrurus miliarius
American alligator	Alligator mississipiensis
Yellow-bellied turtle	Pseudmys scripta
Numerous species of salamanders and frogs	
Wood duck	Aix sponsa
Red-shouldered hawk	Buteo lineatus
Blue jay	Cyanocitta cristata

SWAMP FOREST CONTINUED

Tufted titmouse	Parus bicolor
Foxsparrow (and other sparrows)	Passerella iliaca
Warblers	Dendroica sp.
Various woodpeckers and owls	
A stopping station for migratory birds	

HARDWOOD FOREST

Eastern gray squirrel	Sciurus carolinensis
Flying squirrel	Glaucomys volans
Cottontail rabbit	Sylvilagus floridanus
White-tailed deer	Odocoileus virginianus
Green snake	Opheodrys aestivus
Black snake	Coluber constrictor
Southern copperhead	Agkistrodon contortrix
Slimy salamander	Plethodon glutinosus
Box turtle	Terrapene carolina
Various species of hawks, owls, woodpeckers and songbirds	

LONGLEAF PINE-TURKEY OAK FOREST

Eastern cottontail rabbit	Sylvilagus floridanus
Gray squirrel	Sciurus niger
Gray fox	Urocyon cinereo argenteus
Striped skunk	Mephitis mephitis
White-tailed deer	Odocoileus virginianus
Crowned snake	Tantilla coronata
Eastern coachwhip	Masticophis flagellum
Black snake	Coluber constrictor
Rat snake	Elaphe obsotela
Corn snake	Elaphe guttata
Hognose snake	Heterodon platyrhinos
Eastern diamondback rattlesnake	Crotalus adamanteus
Various species of toads	

POCOSIN

A REFUGE HABITAT FOR

Marsh rabbit	Sylvilagus palustris
White-tailed deer	Odocoileus virginianus
Black bear	Ursus americanus
Mice of various species	

POCOSIN CONTINUED

FREQUENT BIRDS	
Carolina wren	Thryothorus ludovicianus
Catbird	Dumetella carolinensis
Hermit thrush	Catharus guttatus
White-eyed vireo	Vireo griseus
Rufous-sided towhee	Piplo erythrophthalmus
Mourning dove	Zenaida macroura
Common yellowthroat	Geothylypis trichas

MIGRATORY BIRD LIST
(Season and location)

Laughing gull	*Larus atricilla*	(Summer)
Herring gull	*Larus argentatus*	(Winter, rarely inland)
Common tern	*Sterna hirundo*	(Summer on coast)
Broad-winged hawk	*Buteo platypterus*	(Summer)
Snow goose	*Chen caerulescens*	(Winter, on coast)
Canadian goose	*Branta canadensis*	(Winter, on coast)
Common mallard	*Anas platyrhynchos*	(Winter, entire coast)
Black duck	*Anas rubripes*	(Winter)
American pintail	*Anas acuta*	(Winter, on coast)
Green-winged teal	*Anas crecca*	(Coast, late August-mid April)
Northern shoveler	*Anas clypeata*	(Winter)
American widgeon	*Anas american*	(Coast, Oct-May)
Wood duck	*Aix sponsa*	(Mostly fall & winter)
Great blue heron	*Ardea herodias*	(Summer)
Common egret	*Casmerodius albus*	(Summer)
Eastern hermit thrush	*Catharns guttatus*	(Winter)
Pied-billed grebe	*Podilymbus podiceps*	(Winter)
Brown pelican	*Pelecanus occidentalis*	(Mostly in summer)
Double-crested cormorant	*Phalacrocorax auritus*	(Winter)
Lesser scaup	*Aythya affinis*	(Winter)
Bufflehead	*Bucephala albeola*	(Winter, open water)
Whistling swan	*Olar columbianus*	(Winter)
Ring-billed gull	*Larus delawarensis*	(Winter)
Caspian tern	*Sterna caspia*	(Apr.-Nov.)
American coot	*Fulica americana*	(Winter)
Godwall	*Anas strepera*	(Winter, mainly fresh water)
Blue-winged teal	*Anas discors*	(Transient, spring & early fall)
Ring-necked duck	*Aythya collaris*	(Winter, prefers fresh water
Ruddy duck	*Oxyura jamaicensis*	(Winter)
Hooded merganser	*Lophodytes cucullatus*	(Winter)
Canvasback	*Aythya valisineria*	(Winter)

APPENDIX C

GEOREFERENCING PREDICTIVE MODELS

Predictive modeling has been an important concept in cultural resource management for nearly a decade. However, the tying of predictive models to the surface of the earth, or georeferencing, has received little consideration. Georeferencing is critical to utilizing predictive models within the framework of planning and environmental review. This appendix will discuss various methods of georeferencing predictive models, the methods utilized in this study and will suggest methods for future studies.

Types of Georeferencing

Three types of predictive modeling have been previously described, namely, descriptive, behavioral and statistical. Descriptive and behavioral models utilize environmental variables as the unit of analysis or reference to the environment. These variables would include soil type, slope, elevation and type of nearest water. Some of these variables are available on maps while others are calculated from map data (e.g., slope).

Descriptive and behavioral predictive models that rely on single variables may be manually georeferenced. However, the possibility of overlooking or misidentifying areas defined by a model increases with the complexity of the map source. For example, the use of general soils (associations) maps for developing manual predictive models would have little chance of error, but the use of detailed soils (series) maps would be subject to greater classification error due to map complexity. Similarly, when models become dependent upon several environmental variables, developing a manual georeferenced model becomes complex. Map references may be at differenct scales or variables must be calculated from other variables (e.g., slope from elevation). These complexities rule out manual georeferencing in all but the simplest cases.

Statistical predictive models use land area as the unit of analysis. Usually the units of analysis are square or rectangular land units of uniform size. These units or grid cells manifest environmental conditions such as slope, elevation and distance from a water source. These variables usually represent the numeric average for the cell. Soil type assignment, for example, is usually based on the type that covers the greatest portion of a given cell unless a key soil type (as defined by the model) exists in the cell. In that situation the cell may be coded with the critical soil type.

To collect information on grid cells, a grid overlay is placed on a map source and values are extracted for each cell. If there is a large number of cells, manual extraction of this information becomes tedious and accuracy is difficult to maintain. The use of automated methods is critical in georeferencing all but the most simple statistical predictive models.

Whether manual or automated, the accuracy of georeferencing directly reflects the data/map sources used. As a result, georeferencing and the modeling process will not be totally accurate. Several such problems are identified in the text of this report such as two small pockets of soil types not being recorded on the soils maps and the inaccuracy in recording water resources. These criteria should be noted as qualifiers to any predictive modeling scheme.

Data Capture

Several types of geographic data are stored in computers. Data can be stored as points using a single x and y coordinate pair. Data can be stored as lines using a series of x and y coordinate pairs with discrete beginning and ending points. Polygons represent a special case of a line that closes on itself where the beginning coordinate pair is identical to the ending coordinate pair. Finally, there is the special case of the polygon which contains four points, defining a regular grid cell.

Agencies which record and map geographic data are beginning to store their data as described above. The advantages of this storage include the ability to reproduce map data to any scale. However, the accuracy of the data reflects the resolution or scale of the base data. The data can be easily updated without reentering or redrafting the entire data base or map. Digital geographic information provides data that can be easily analyzed or integrated with similarly stored data.

Elevational information was the only extant geographic data base stored at LRIS that was used in this project. This data was stored in grid cell format. All other data bases used by this study were digitized or stored into the system as part of the project.

Capturing geographic data through digitizing is a costly and time-consuming process. For example, each 9 3/4 by 15 inch detailed soils sheet took approximately 40 hours to store error free. Thus, the use of extant geographic data bases is most cost effective.

Isolating the Model Geographically

Two methods may be used to georeference predictive models when more than one variable is used. When the raw geographic data is stored as polygons, the intersection of one area with another can be extracted through calculation. For example, all areas which have Craven fine sandy loam and elevations between 4 and 7 meters could be extracted. Note that grid cells are a special form of polygon and can be processed with irregular polygons. The use of polygon overlay methods produces the most accurate geographic referencing of a model. However, these methods require a large number of calculations which in turn require a large amount of computer time.

The alternative method is to use grid cells. This may require the automated conversion of polygons to grid cells. Once in this format, grid cells which contain the desired environmental characters can be identified. This process is extremely simple and efficient. However, this method lacks the accuracy of the polygon overlay process. Accuracy will vary with the size of the grid cell. The larger the grid cell the less accurate the georeferencing. The smaller the grid cell the more accurate the georeferencing. Smaller grid cells will require more processing time to convert polygons to grid cells and to extract the model. The decision to use polygon overlays versus grid cells for georeferencing depends upon available geographic data bases and the resolution needed.

Georeferencing the New Hanover Predictive Models

The state's Land Resources Information Service (LRIS) stored and processed the geographic information used in this study. LRIS uses Comarc's Geographic Information System software and hardware to collect and manage geographic data bases. This system was installed late in 1977. LRIS and the Division of Archives and History reached a cooperative agreement late in 1977 to test the applicability of the Comarc system's ability to georeference archeological predictive models using polygon overlays. In spring of 1978 work began to digitize detailed soils, roads, water sources, surveyed areas, and site locations into the system. In December and January 1978-1979 the initial predictive model was partially georeferenced. In April and May 1979 the revised descriptive model was georeferenced.

The georeferencing of archeological predictive models at LRIS was the largest and most complex project initiated at LRIS at the time. In many ways this cooperative project served as an initial test of the analytical portions of the system software. Throughout the process problems with the software were identified. During the georeferencing of the initial model, several computer programs were identified that did not work and other programs were determined not to work at the level of efficiency desired. Several of the overlay runs took over ten hours and the calculation of the distance to nearest water was not possible. The final overlay of distance to nearest water and type of nearest water was completed manually. However, this overlay was facilitated by a LRIS-drawn map of the county's water resources. Since that time the programs have been corrected and are now more efficient. The refined descriptive predictive model was totally georeferenced by LRIS (see figures 5.1 and 5.2).

Aside from the problems with the software, a good deal was learned about the problems inherent in georeferencing archeological predictive models. This includes the time needed to store the geographic data necessary for georeferencing the predictive model. The computer time needed for the polygon overlay process was many times what was anticipated. If the initial model had been completely georeferenced by LRIS, the computer time to calculate the overlays would have been in excess of four days (96 hours) of computer time. It should be noted that this estimate reflects the time needed on a small mini computer which processes data at a much slower rate than a large computer.

Suggested Methods

The cost of storing environmental data used for georeferencing predictive models of archeological site occurrence is dropping. New digitizing methods and software have reduced digitizing time by 75 percent.

This time will also be further reduced when automatic digitizing hardware becomes reliable and affordable.

The sharing of digital geographic information through services like LRIS will provide a reliable and inexpensive library of information for the georeferencing of predictive models. Efforts in predictive modeling should center on available environmental data bases. These include LANDSAT, elevational information, general soils associations, and land use. Additional data bases will become available in the future.

Many of these data bases are stored in grid cell format with cells roughly 220 feet to a side. Given this resolution and the efficiency of grid cell processing, it is suggested that future research investigate the utility of grid cell processing with 200 to 250 foot grid cell sizes. The major advantage in the use of small grid cell methods, when compared to polygon overlay methods, is the ease and speed of georeferencing revised predictive models with little loss in resolution.

APPENDIX D

This appendix contains maps of the 19 units sampled during the project. As mentioned in the text, the units are not numbered consecutively because of the manner in which the final sample was chosen. A map depicting approximate topographic contours has been overlaid on the map of shovel test locations for each unit. A key to the symbols used on the maps appears below.

■ Sterile shovel test

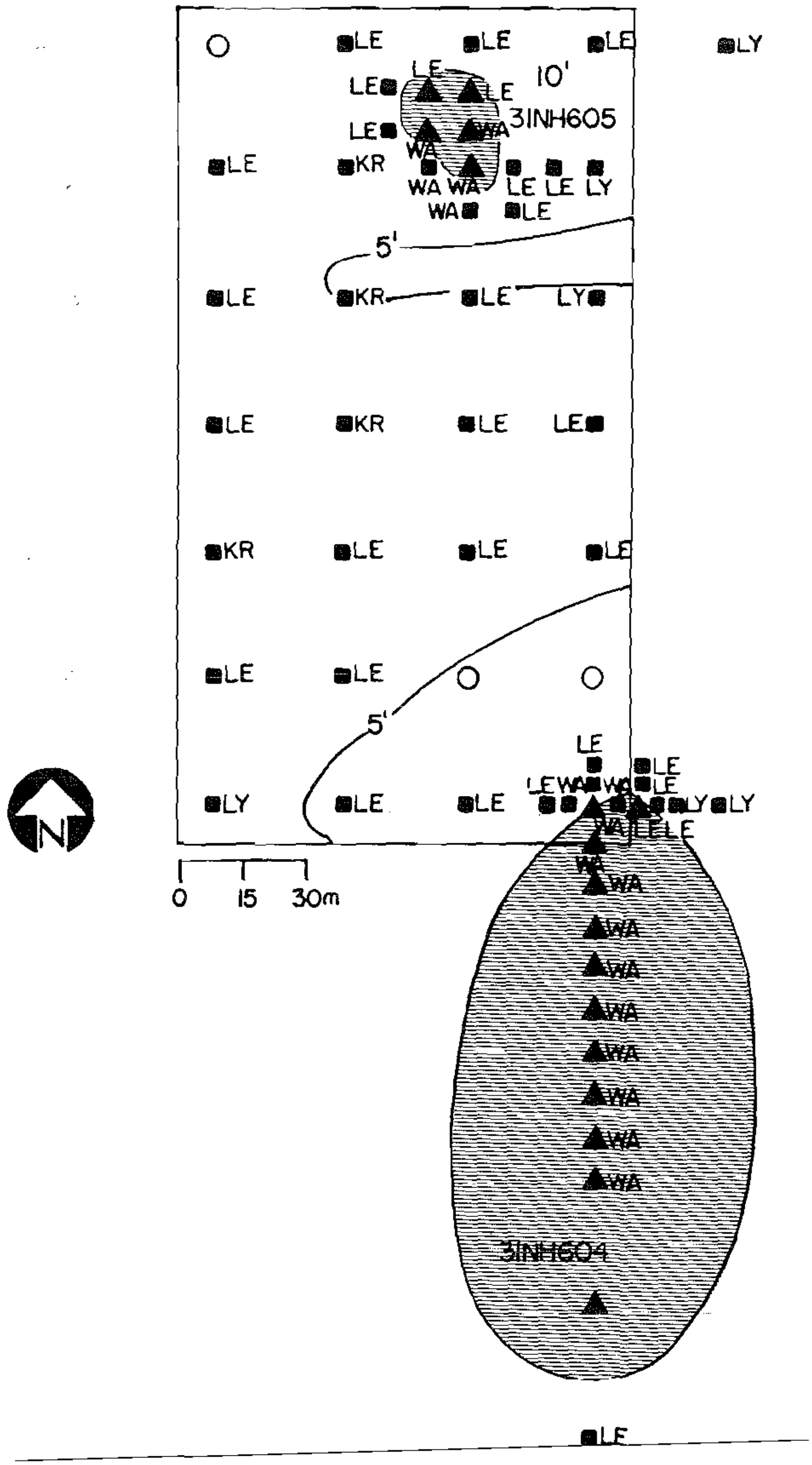
△ Nonsterile shovel test

○ Inaccessible shovel test

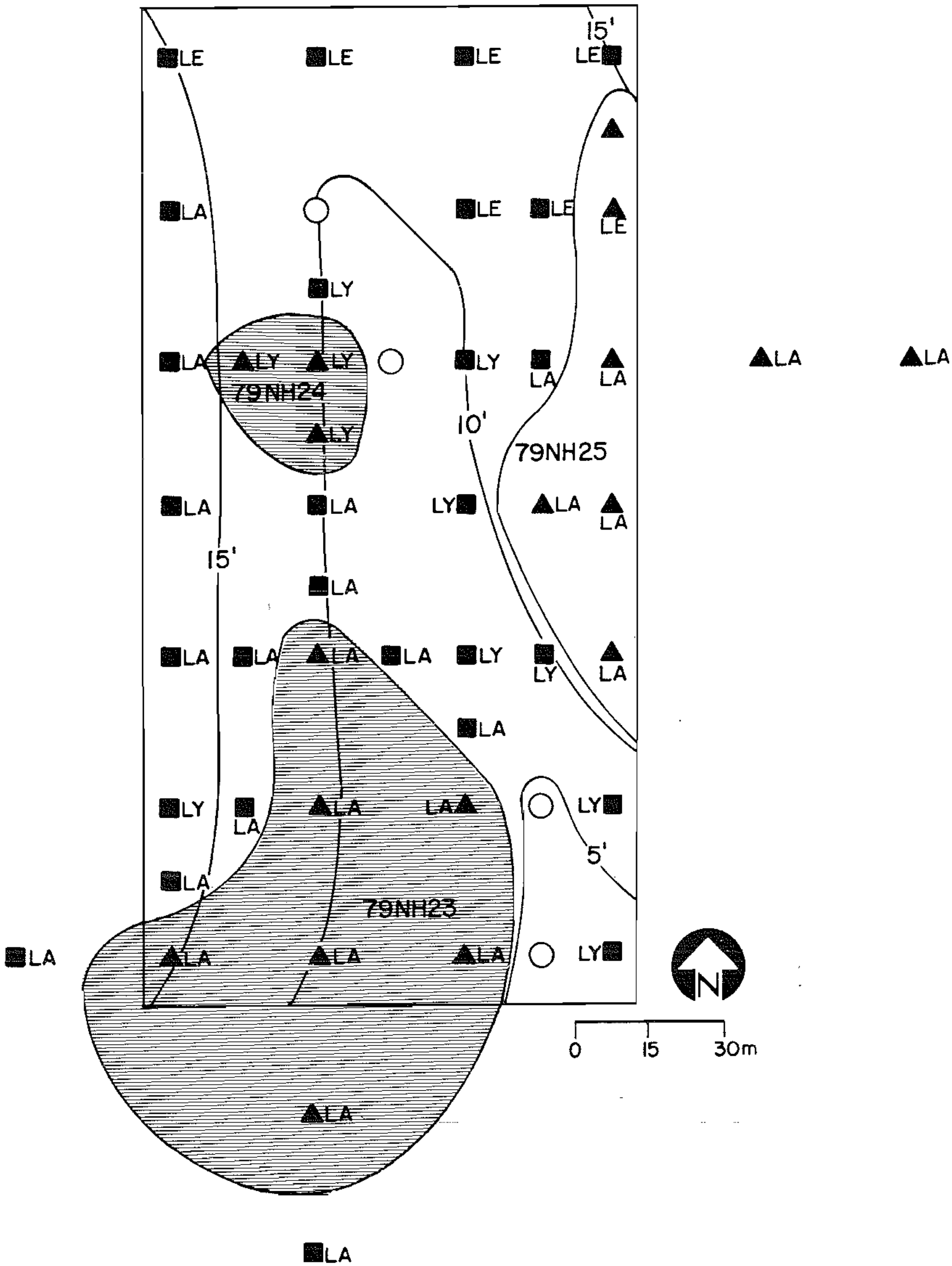
Archaeological site

LE Soil series classification

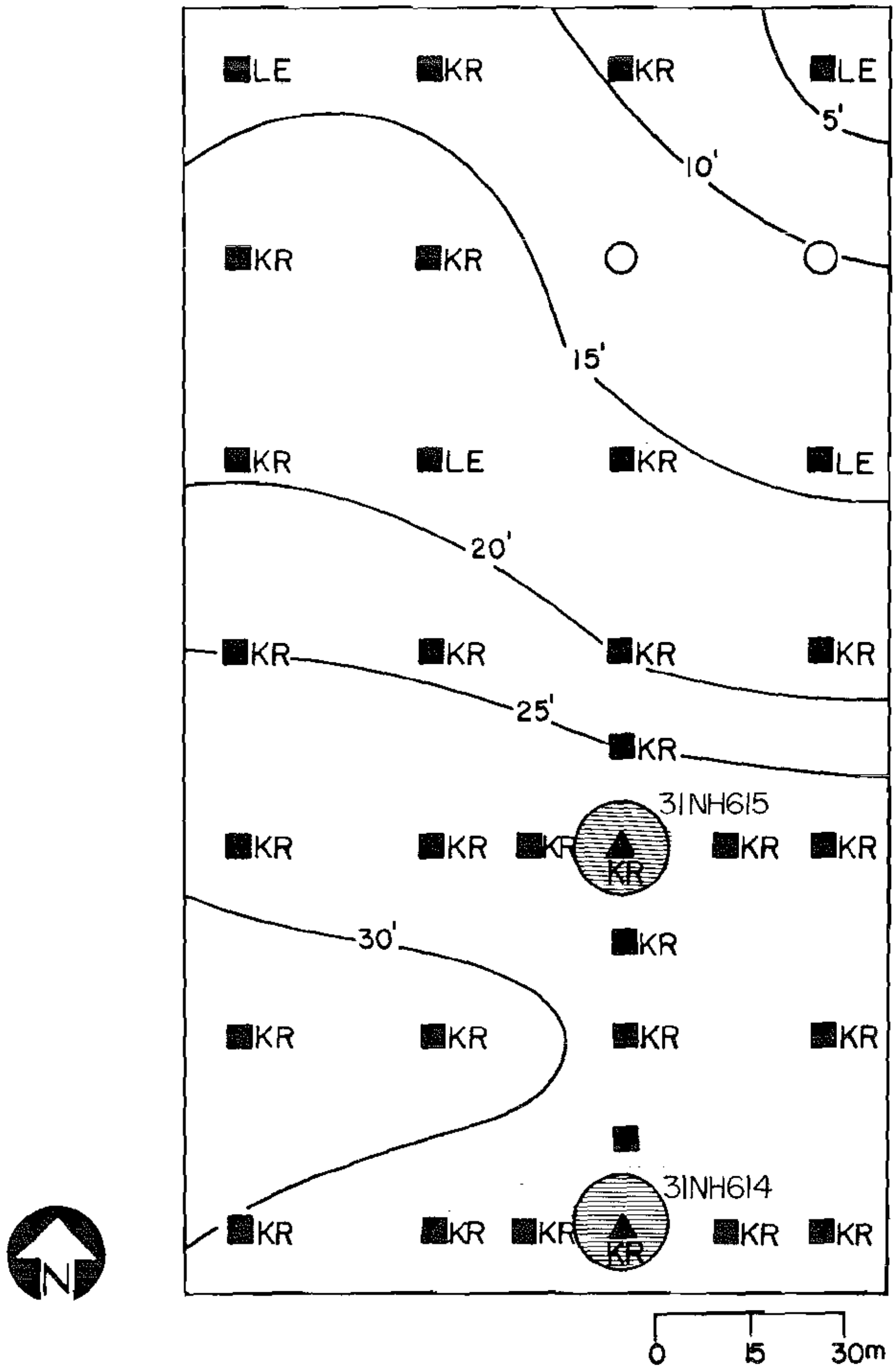
Unit 3 (medium probability)



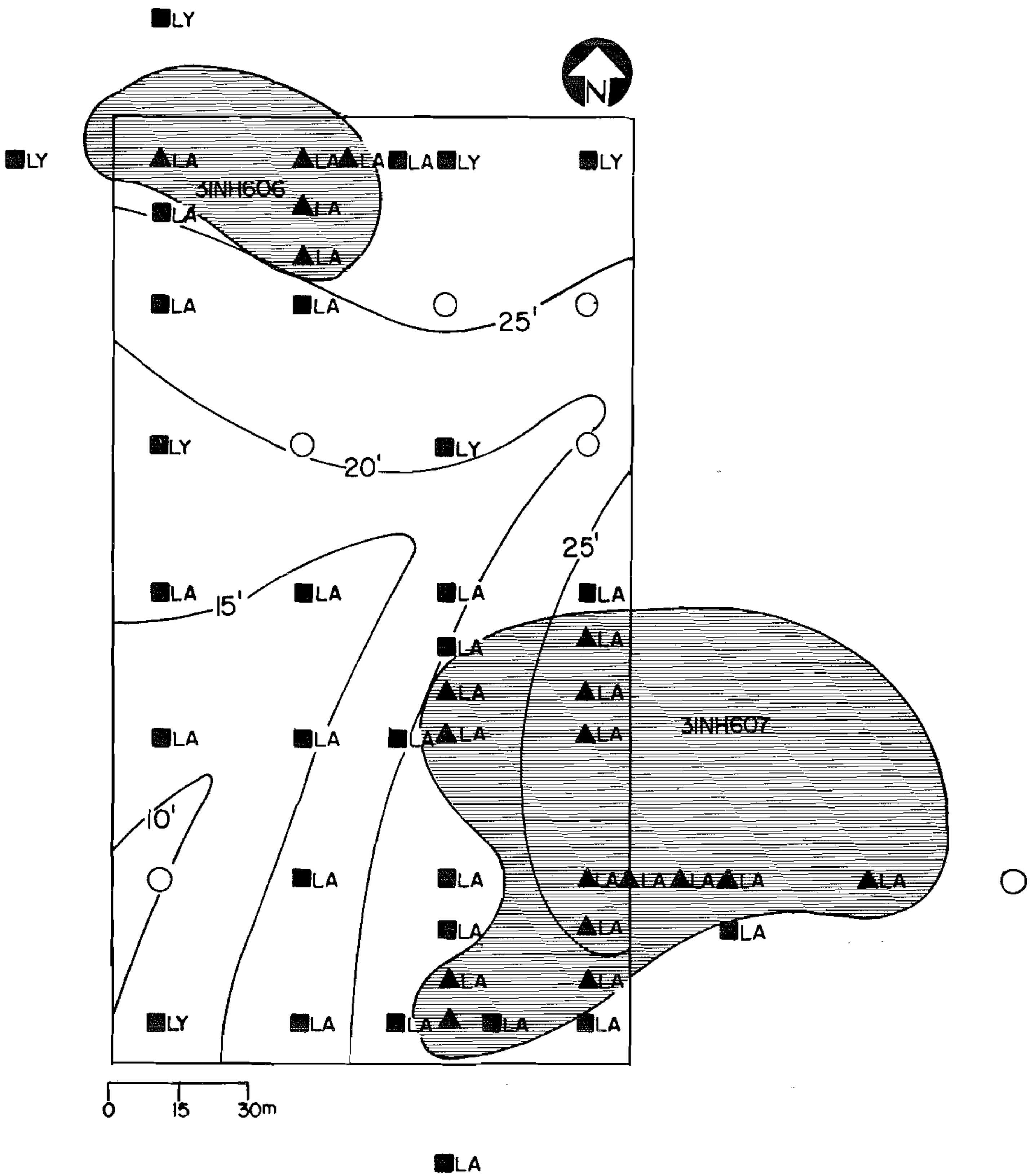
Unit 8A (high probability)



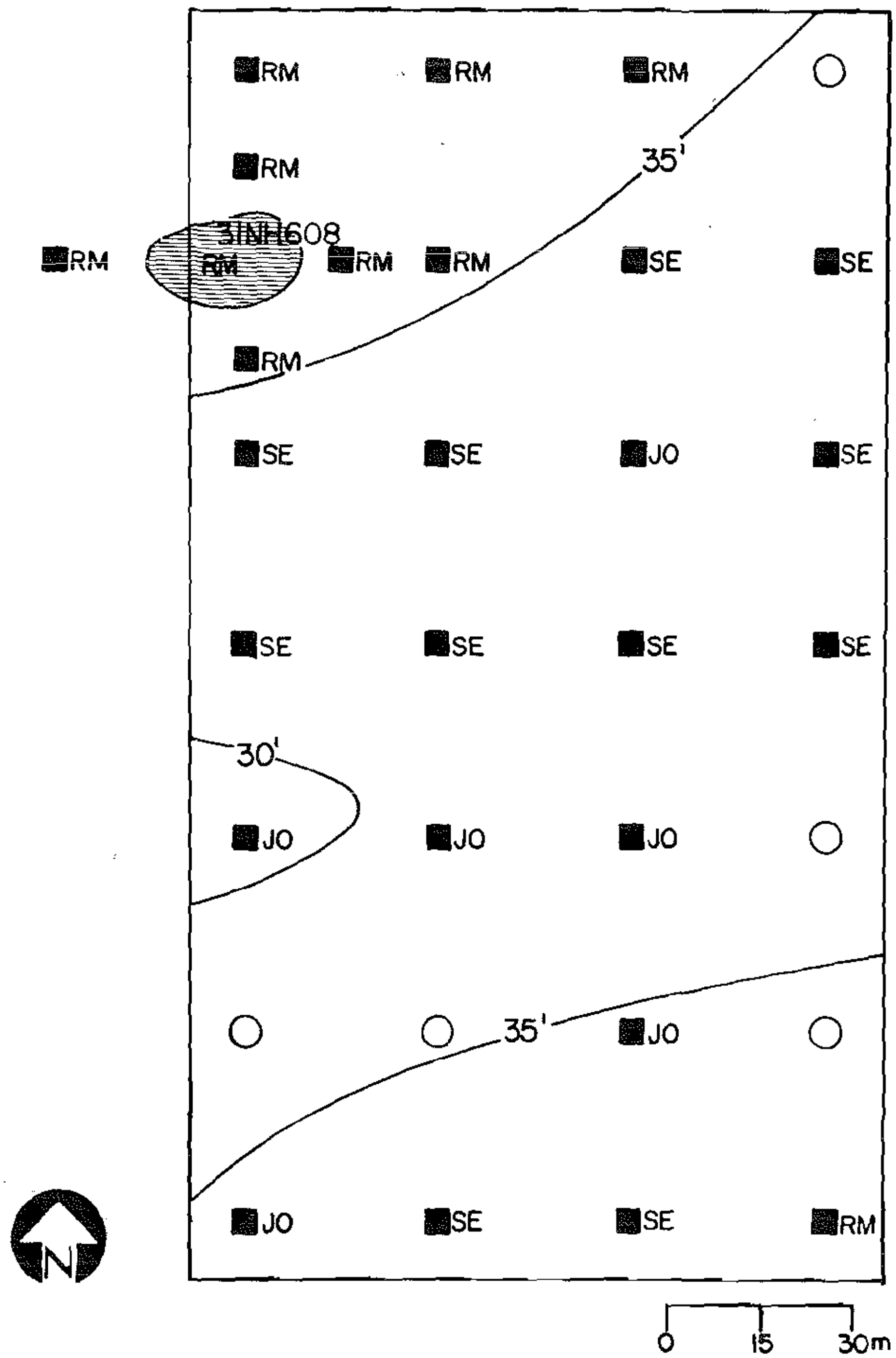
Unit II (medium probability)



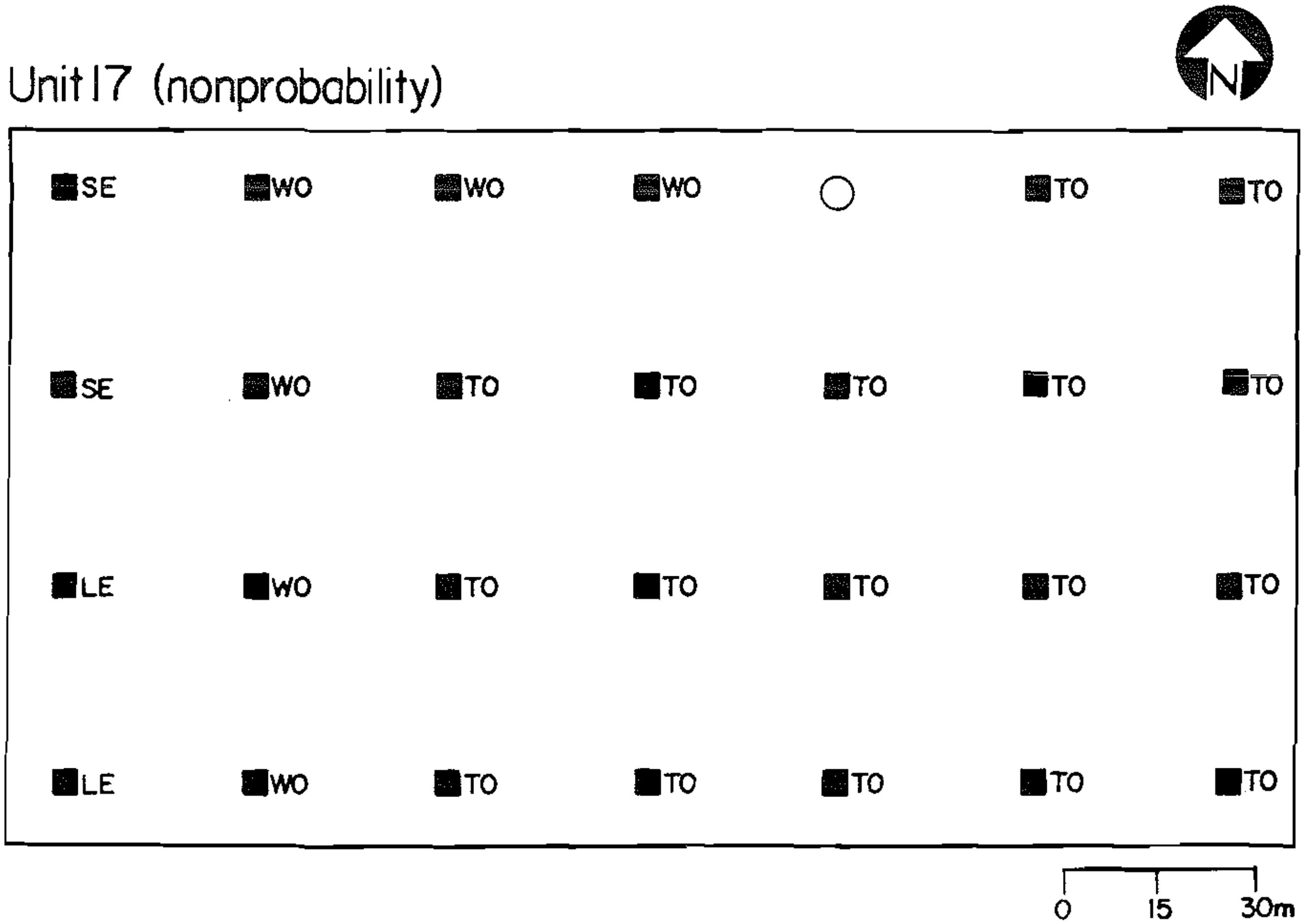
Unit 12 (high probability)



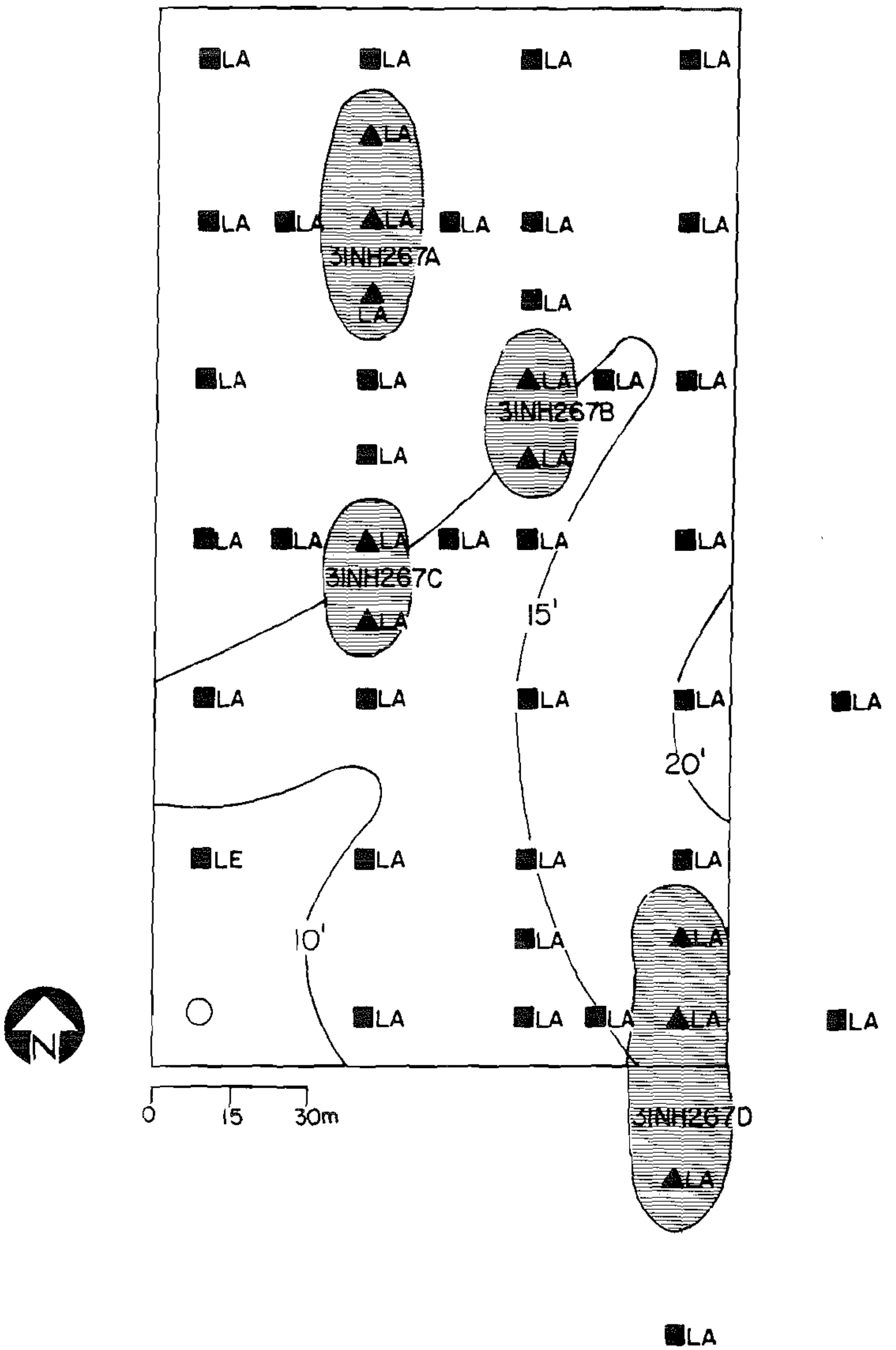
Unit 16 (low probability)



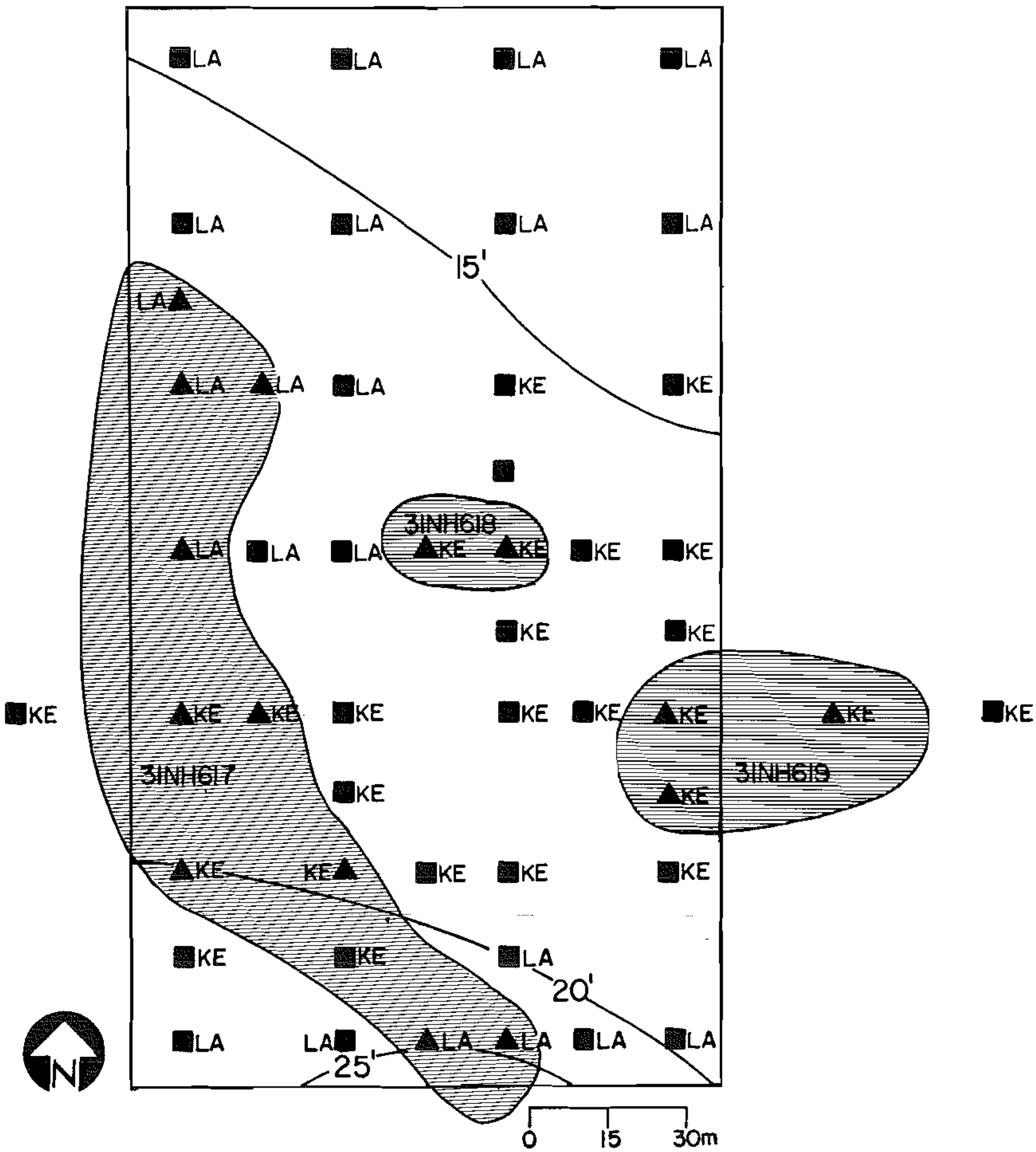
Unit 17 (nonprobability)



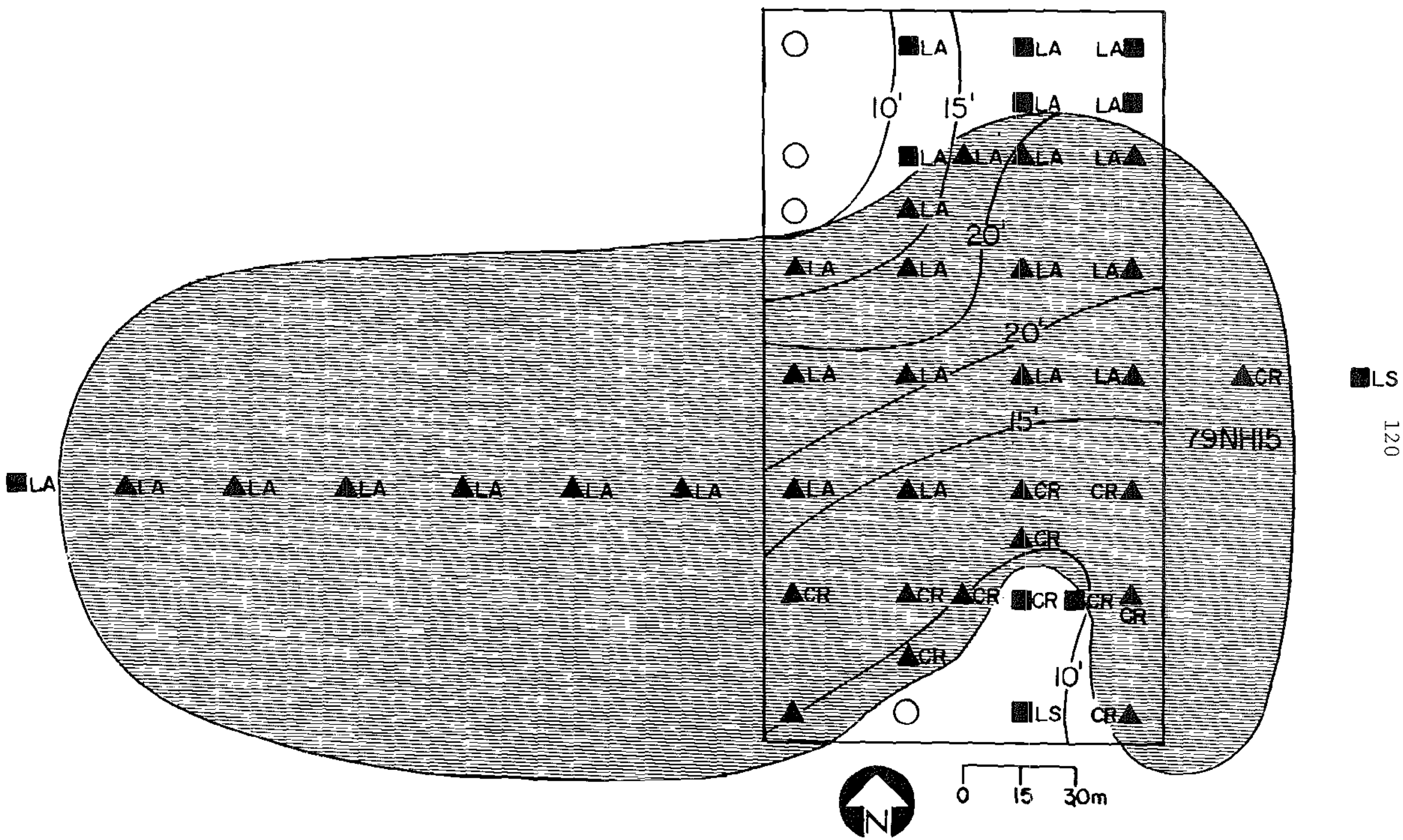
Unit 20 (high probability)



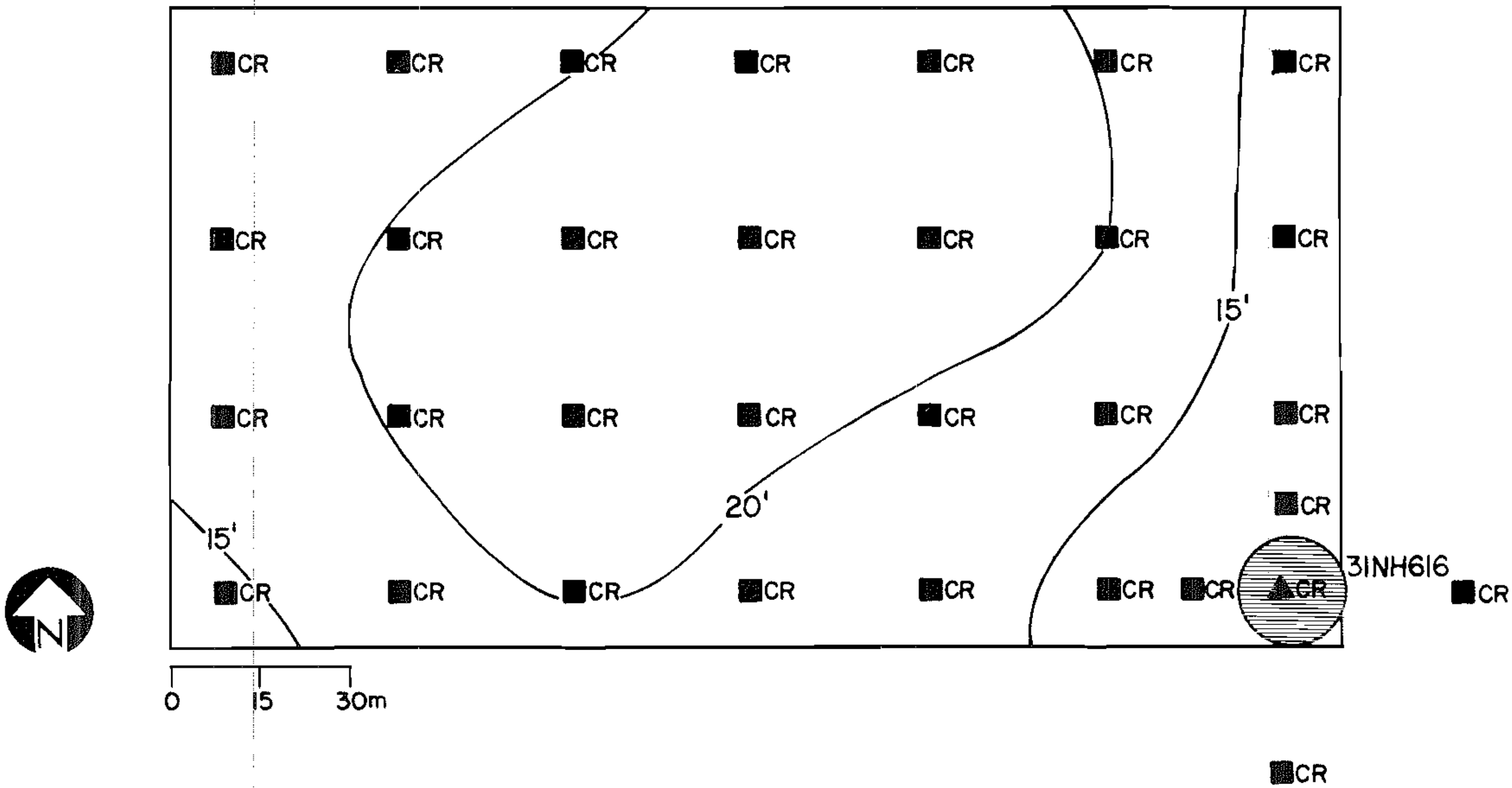
Unit 29 (low probability)



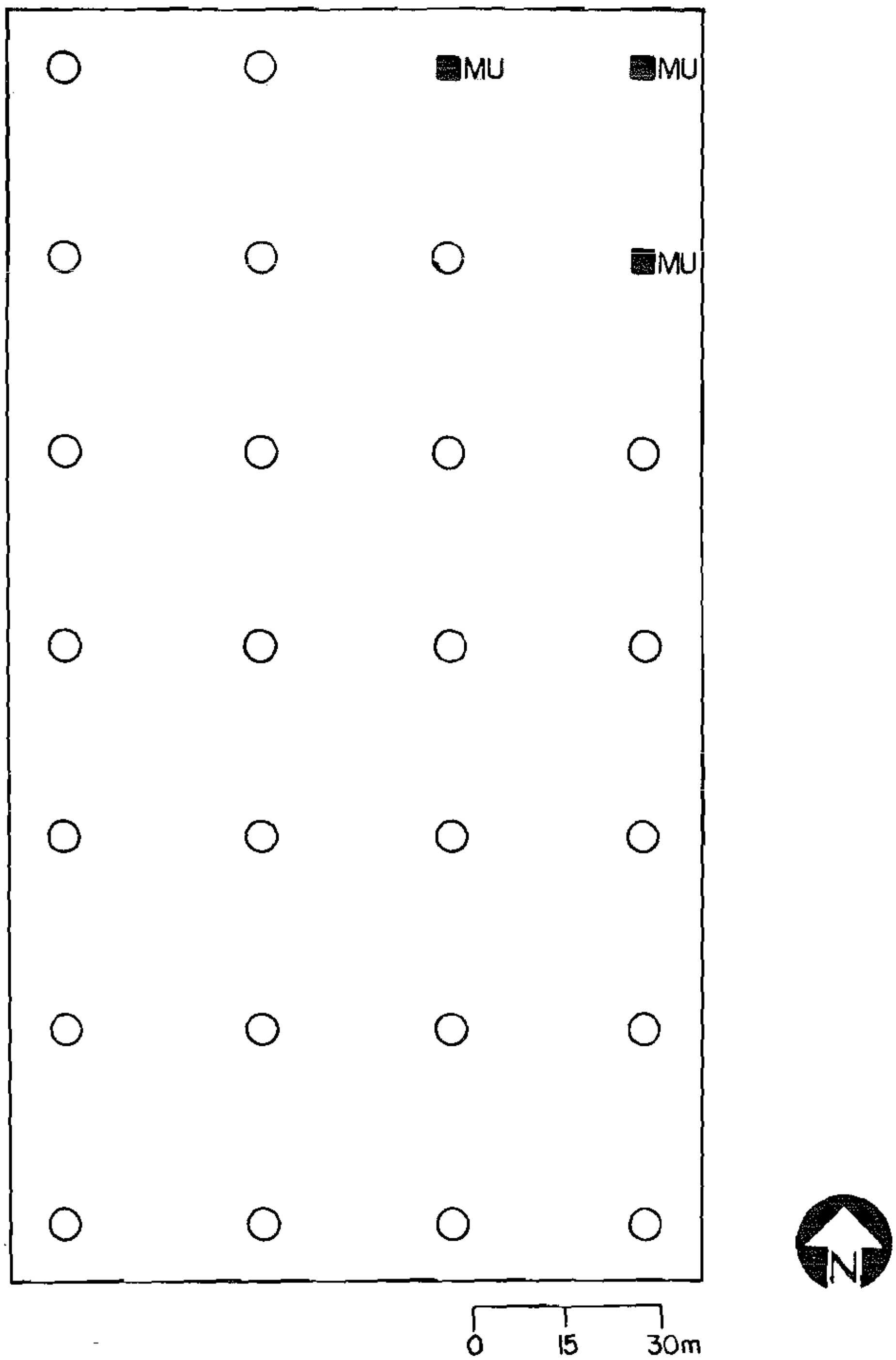
Unit 31 (high probability)



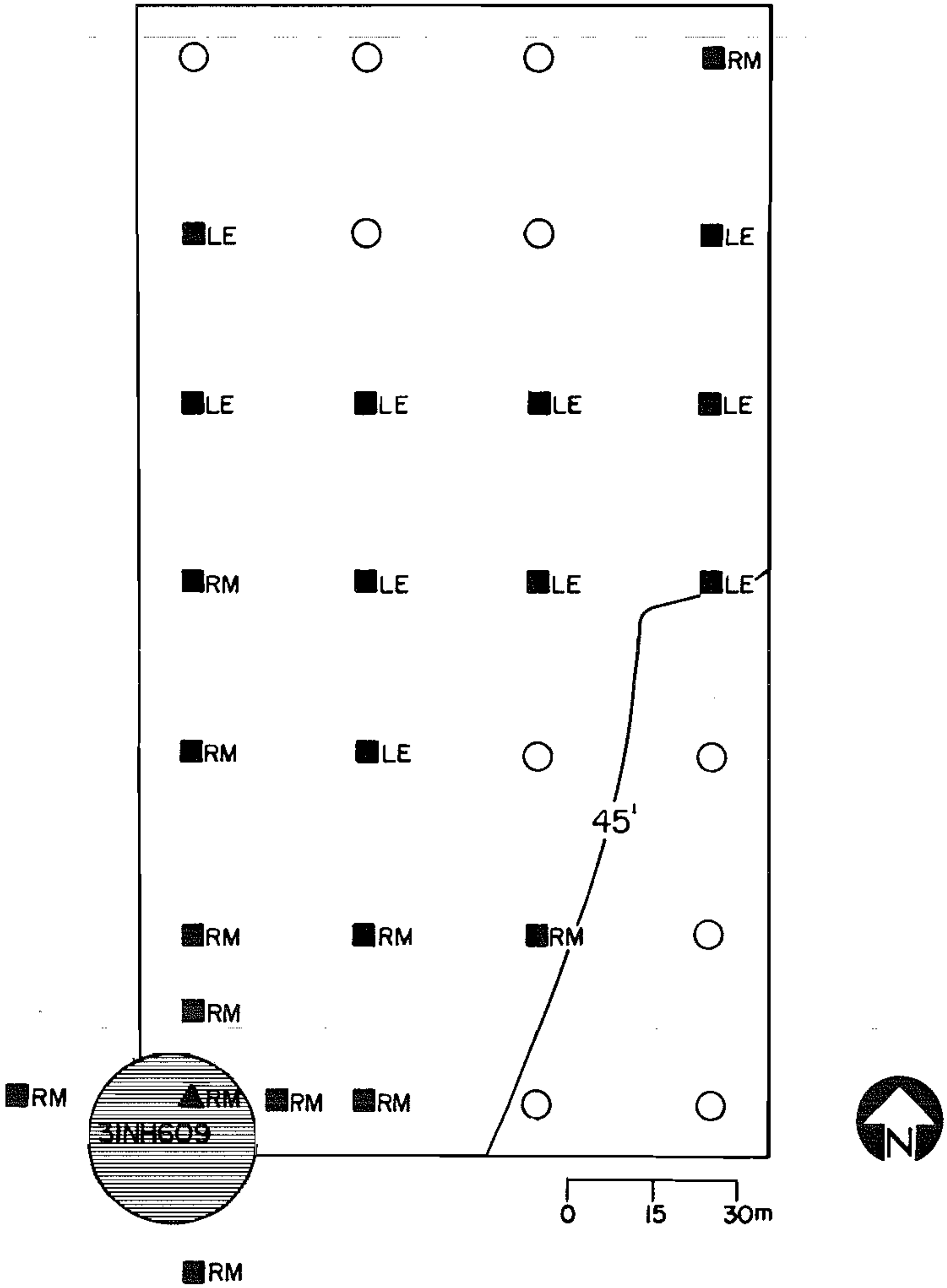
Unit 36 (high probability)



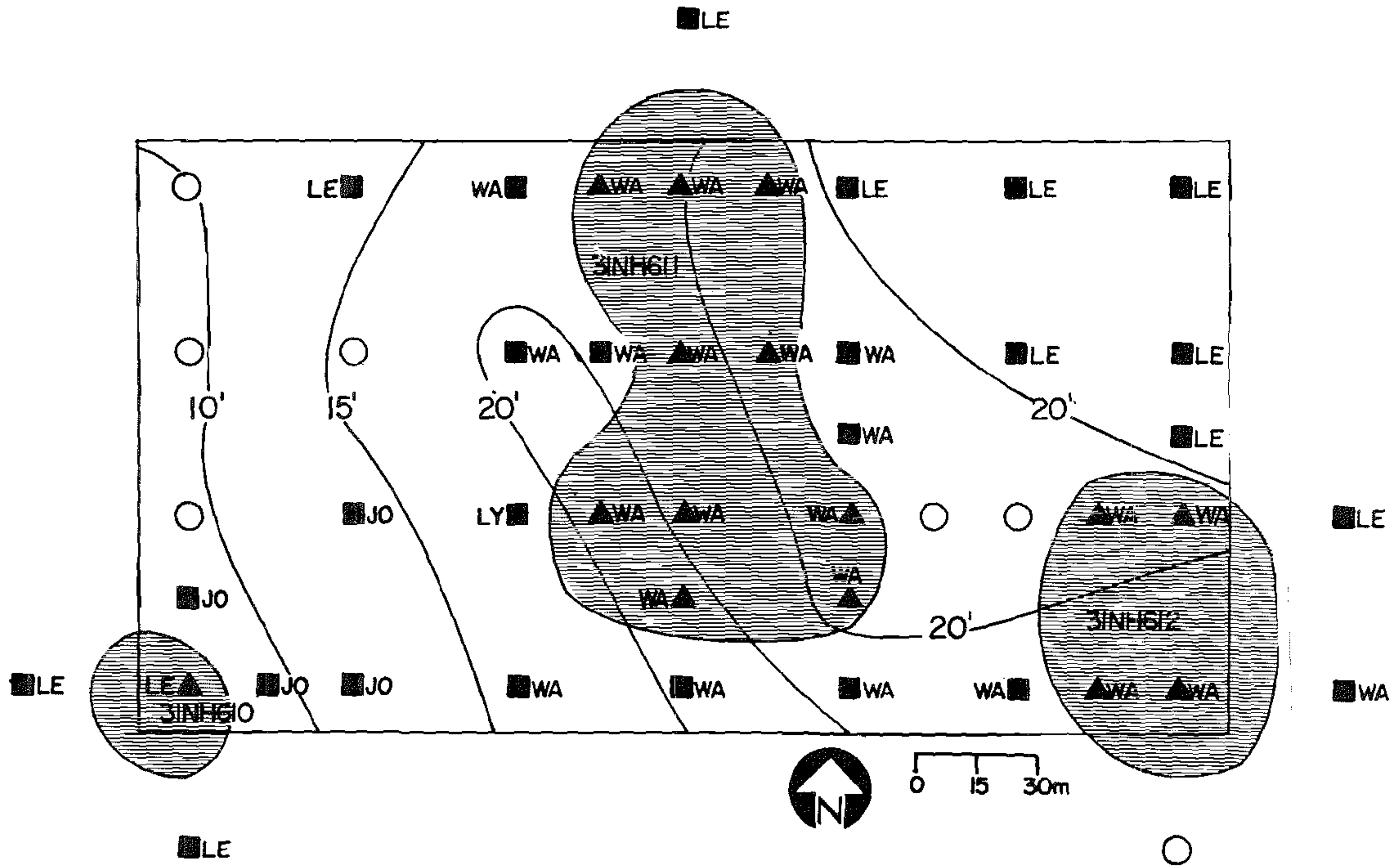
Unit 37 (nonprobability)



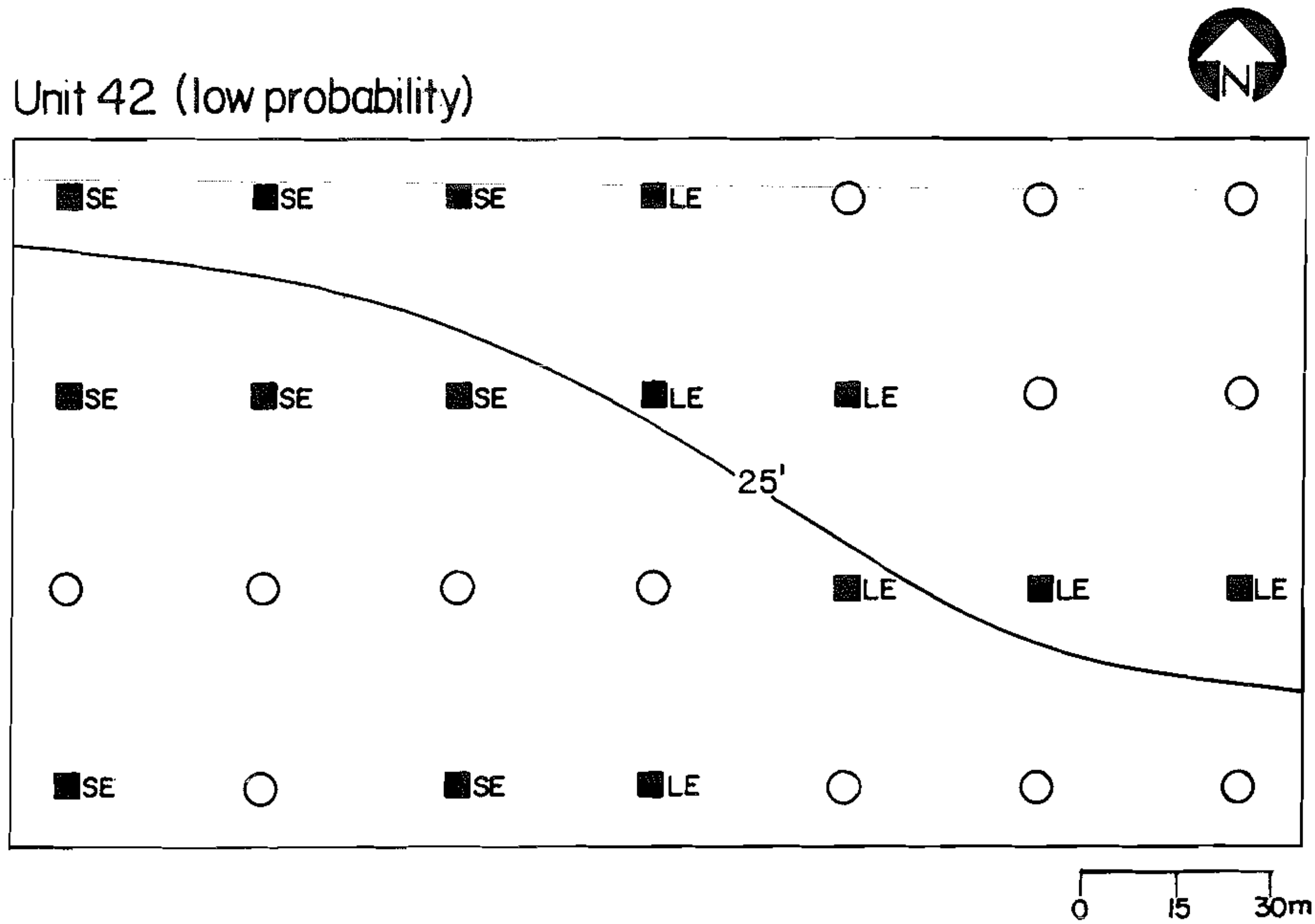
Unit 39 (medium probability)



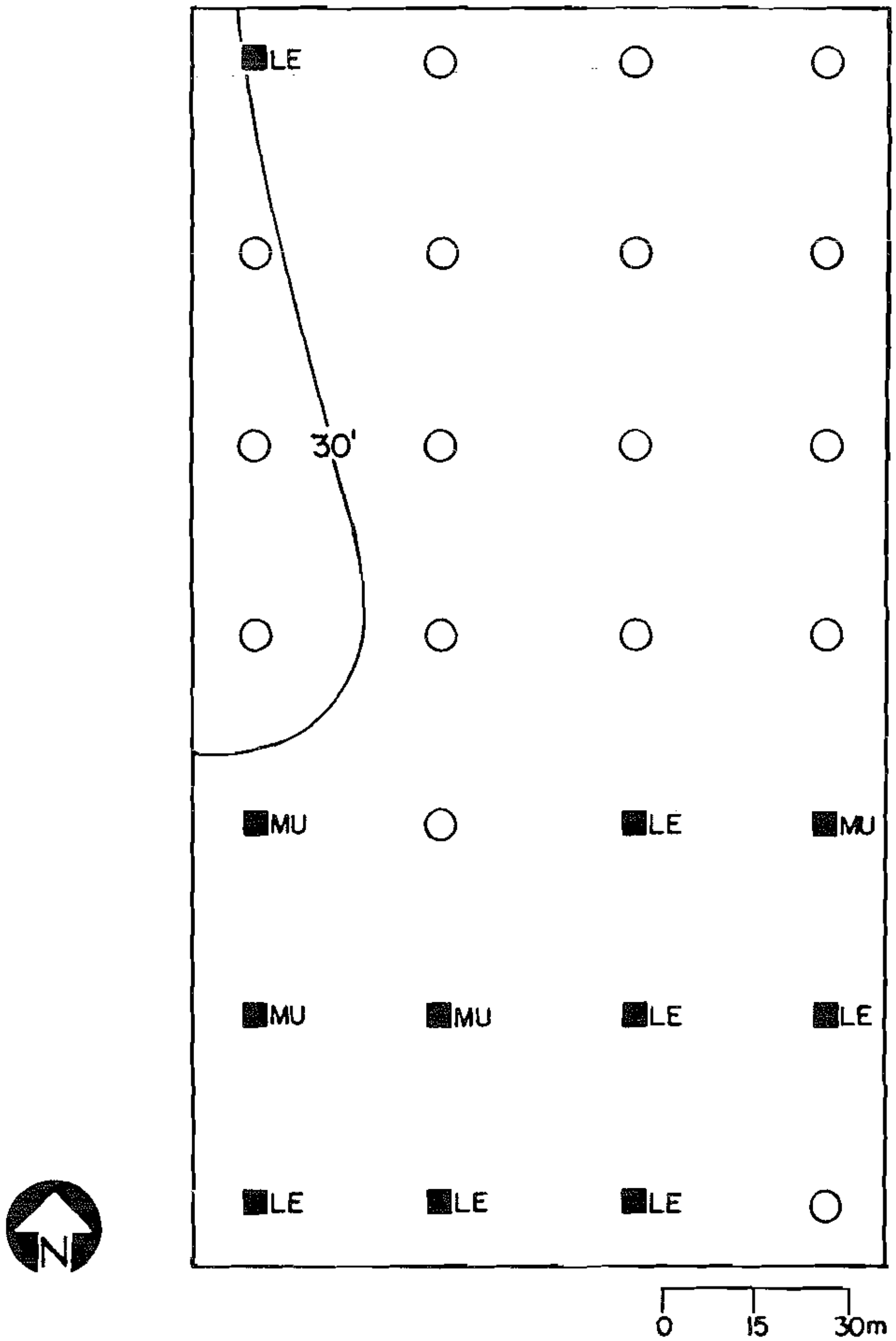
Unit 40 (medium probability)



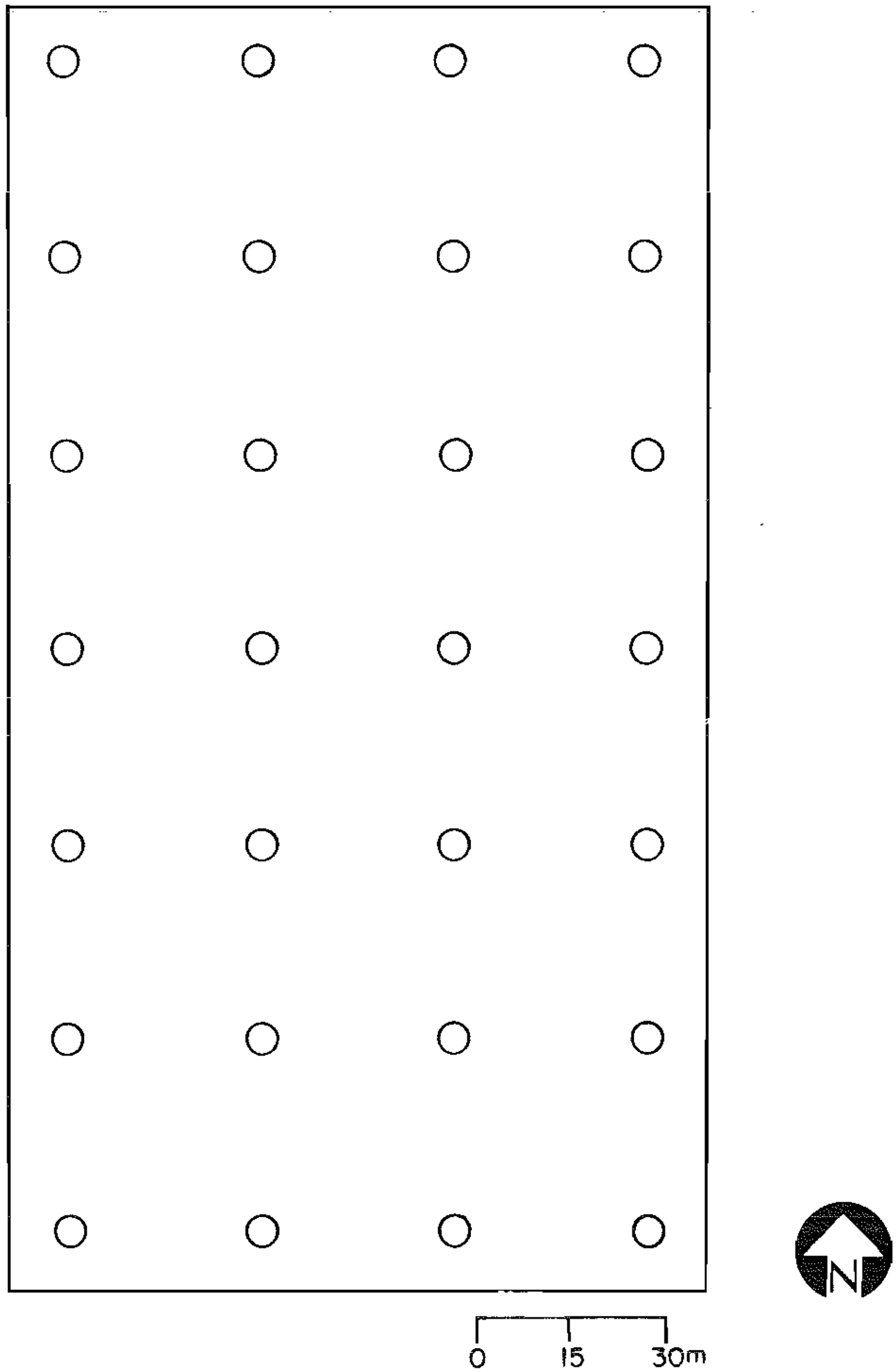
Unit 42 (low probability)



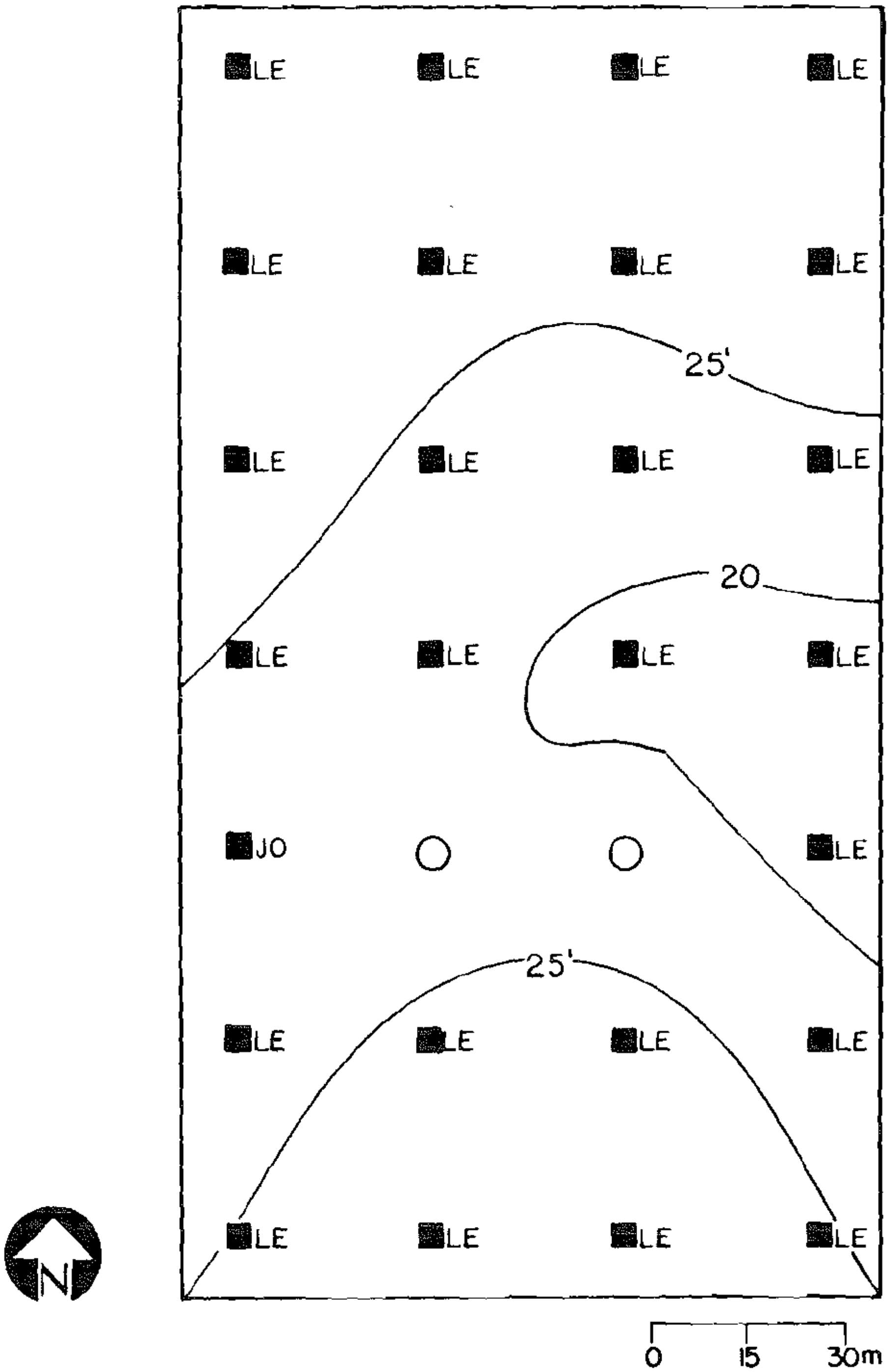
Unit 46 (nonprobability)



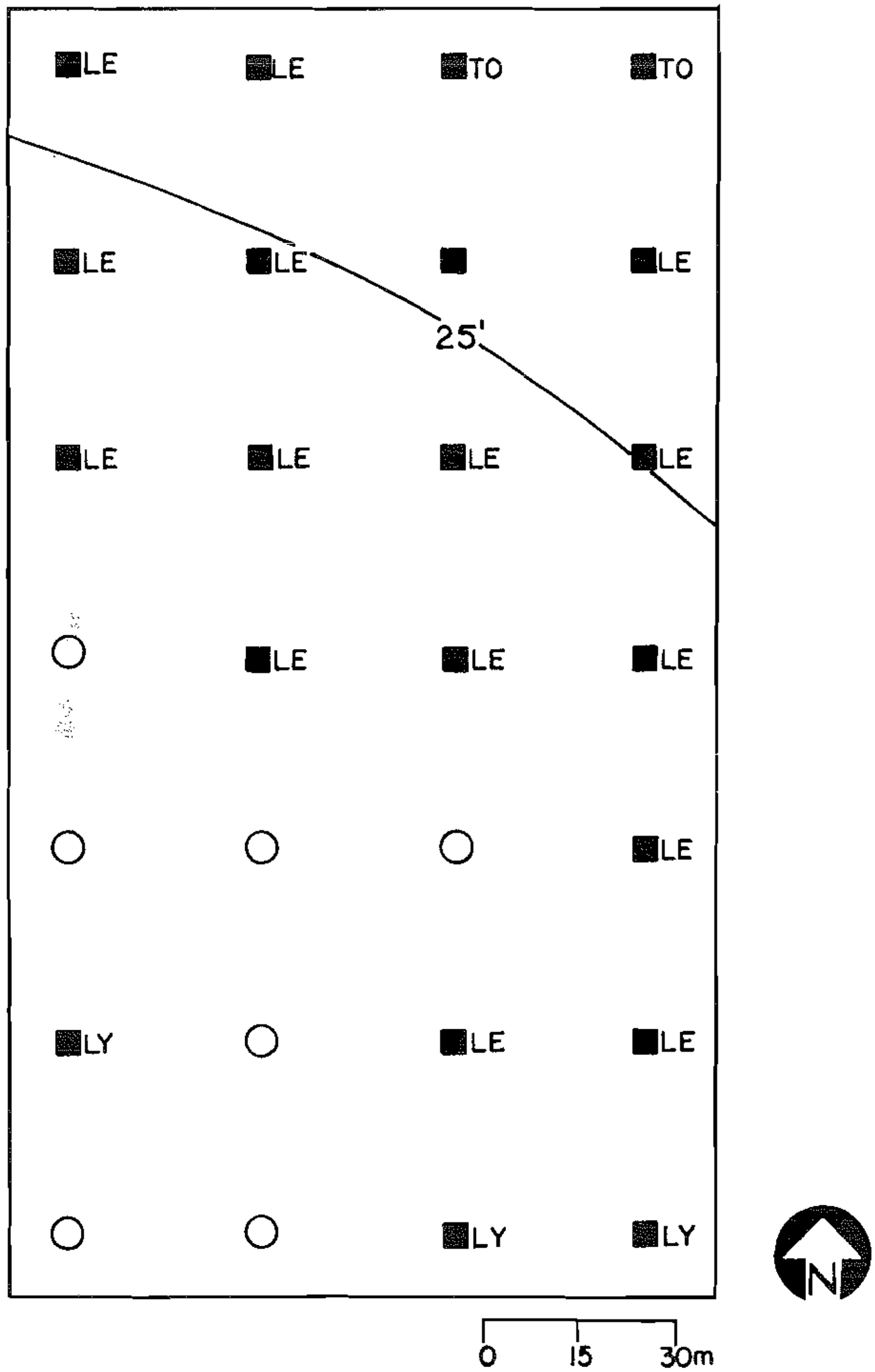
Unit 47 (nonprobability)



Unit 56 (medium probability)



Unit 75 (nonprobability)



APPENDIX E

This appendix contains brief descriptions of all sites located during the New Hanover County Predictive Models Project (NHCPMP). The sites are listed by the permanent numbers assigned to them by the Research Laboratory of Anthropology, University of North Carolina, Chapel Hill. The NHCPMP temporary numbers appear in parentheses after the permanent number. Some of the sites described within this appendix were not included in the mathematical analyses for various reasons. 31NH602(79NH-2) and 31NH603 (79NH-3) contain only historic artifacts and were therefore excluded from the sample. 31NH261(79NH-10) and 31NH325(79NH-26) were both located outside the boundaries of the sample units. The site assigned the temporary number 79NH-8 was outside the survey area and has not been assigned a permanent number.

31NH261 (79NH-10)

This site contains a Woodland period component, represented by both lithic and ceramic materials; and a possible Mississippian period component, indicated by the presence of daub. The site is located in a plowed field on an upland flat. Further test investigations for the full evaluation of the site are recommended. This site was located outside the boundaries of the sample unit and had been noted by a previous survey.

31NH267 (79NH-19)

This site contains a Woodland period component as well as some Historic period material. It had been recorded by a previous survey. The site is located in a plowed field on a low rise. Due to the definition of site used in this study (see text), this site was counted as four sites, namely, 19A, 19B, 19C and 19D. No further testing is considered necessary for the evlauation of this site.

31NH601 (79NH-1)

This Woodland period site is on a knoll in a scrub pine clearing 180 meters from a swamp. Site size is in the range between 601 and 5000 square meters. Only prehistoric ceramics were found. It is recommended that test investigations be conducted to fully evaluate the site.

31NH602** (79NH-2)

This historic site consists of some possible Civil War period earthworks. One cannonball fragment was found nearby. No prehistoric component exists at this site. No further investigation of the site is necessary.

31NH603** (79NH-3)

This historic site consists of an isolated occurrence of three pieces of historic ceramics. The area has been disturbed by bulldozing and no further investigation is warranted.

31NH604 (79NH-4)

This Woodland period site is in a plowed field on a terrace remnant which is in the floodplain of a stream located 290 meters away. Only prehistoric ceramics were found in an area between 5001 to 10,000 square meters. This site may represent a Woodland period village. It is recommended that test investigations be conducted to fully evaluate the site.

31NH605 (79NH-5)

This site, which may represent a small Woodland period village or hamlet, is on a hill in the lawns of a residential development. The core area of the site is between 101 and 600 square meters. The site is defined by the presence of lithic material of an indeterminate nature, prehistoric ceramics and preserved features. Test investigations should be conducted to fully evaluate the site.

31NH606 (79NH-6)

This site, which may represent a Woodland period hamlet, is on a forested hill. Its size ranges between 601 and 5000 square meters. Prehistoric ceramics define the site. It is recommended that test investigations be conducted to fully evaluate the site.

31NH607 (79NH-7)

This prehistoric site may consist of several components. It is characterized by Woodland period ceramics and lithic material of an indeterminate nature. Located on a forested hilltop, the site ranges between 601 and 5000 square meters in extent. It is considered unlikely that any undisturbed deposits remain. Further test investigations should be conducted to fully evaluate the site.

31NH608 (79NH-9)

This prehistoric site consists of two isolated lithic flakes. They were found on a forested hilltop 50 meters from a stream. No further investigation is needed at this site.

31NH609 (79NH-11)

This prehistoric site consists of one isolated flake. It is in a level area of a tree farm. No further investigation is considered necessary.

31NH610 (79NH-12)

This Woodland period habitation site is located on a forested upland flat. The core area is one to ten square meters in extent. The site is represented by prehistoric ceramics and a small isolated feature which may be in a dateable context. It is recommended that test investigations be conducted to fully evaluate the site.

31NH611 (79NH-13)

This Woodland period habitation site is located on a forested hilltop. It is characterized by lithic material of an indeterminate nature, prehistoric ceramics and a possible feature. The site area is between 601 and 5000 square meters. It is recommended that test investigations be conducted to fully evaluate the site.

31NH612 (79NH-14)

This site has both a prehistoric and an historic component. The prehistoric component consists of Woodland period ceramics. The historic component consists of a brass or bronze nail which is probably part of a slave burial, indicating that this site may be part of an historic slave cemetary. The site is located on a forested upland flat and extends between 601 and 5000 square meters. Test investigations are recommended to fully evaluate the site.

31NH613 (79NH-15)

This site, probably a multi-component Woodland period village, is located on a prominant rise surrounded by swamps and tidal marsh. This large (25,001 to 50,000 square meters) site is represented by a dense concentration of prehistoric ceramics and some indeterminate lithic meterial. It is recommended that test investigations be conducted to fully evaluate the site.

31NH614 (79NH-16)

This Woodland period site consists of one prehistoric potsherd. This isolated artifact was found on a forested upland flat. No further investigation is necessary at this site.

31NH615 (79NH-17)

This prehistoric site consists of one flake. This isolated piece of lithic material was found on a forested upland flat. No further investagation is necessary at this site.

31NH616 (79NH-18)

This prehistoric site is represented by Woodland period ceramics and some lithic material, including a possible Morrow Mountain projectile point. The artifact scatter was sparse and widely scattered in a plowed field on an upland flat. The site size is greater than 50,000 square meters. No further investigation is considered necessary at this site.

31NH617 (79NH-20)

This site consists of a moderately dense concentration of prehistoric ceramics, possibly representing a small Woodland village. It extends between 5001 and 10,000 square meters and is located on a forested upland flat. Test investigations should be conducted to fully evaluate the site.

31NH618 (79NH-21)

The small (101 to 600 square meters) concentration of Woodland period ceramics at this site may represent a village or hamlet. It is located on a wooded upland flat. Test investigations should be conducted to fully evaluate the site.

31NH619 (79NH-22)

This prehistoric site consists of Woodland period ceramics concentrated in an area between 601 and 5000 square meters. It is located on a forested upland flat and probably represents a village or hamlet site. Test investigations should be conducted to fully evaluate the site.

31NH620 (79NH-23)

This prehistoric site is characterized by a dense concentration of Woodland period ceramics and indeterminate lithic material. The areal extent of the site is between 601 and 5000 square meters. The site is located on a forested low rise adjacent to a stream. The relatively high concentration of lithics may represent an Archaic component, and the ceramics may represent a Woodland period village site. Test investigations should be conducted to fully evaluate the site.

31NH621 (79NH-24)

This prehistoric site consists of a sparse scatter of Woodland period ceramic material, covering an area between 601 and 5000 square meters. It is located on a low rise in an abandoned vineyard. No further work is necessary at this site.

31NH622 (79NH-25)

This site has both prehistoric and historic components. A dense concentration of Woodland period ceramics represents the prehistoric component. An abandoned house, and various historic materials scattered around the house, comprise the historic component. The site is between 5001 and 10,000 square meters and is located on a broad, low rise. Test investigations should be conducted to fully evaluate the site.

31NH623 (79NH-26)

This site consists of nondiagnostic lithic material and Woodland period ceramics. It is located on a low rise in a plowed field. It is recommended that test investigations be conducted to fully evaluate the site.

APPENDIX F

CERAMIC ARTIFACTS RECOVERED PER SITE

NC SITE NO.	SHERD TEMP.			SAND TEMP.				SHELL TEMP.		
	FI	CM	O	FI	CM	P	O	CM	P	O
31NH601		1	7	11	1		37	2		10
31NH604	3	3	7	1	3		13			2
31NH605	9	2	6				1			
31NH606						3	1		5	
31NH607	10		1		6		16			
31NH608										
31NH609										
31NH610					8					
31NH611	4			1		1	5			
31NH612							11		2	
31NH613	5		44	1	8		84		2	1
31NH614	(ONE SHERD LOST IN THE FIELD)									
31NH615										
31NH616			12				4			
31NH267A			1				2			
31NH267B			2							
31NH267C							4			
31NH267D			3				1			
31NH617		2	4	1	3		7		1	
31NH618	1	1					3			
31NH619	1	1	1							
31NH620	14	1	7				5			
31NH621	1		1	2			2			
31NH622	4	1	5	7			17			

FI = Fabric-impressed
CM = Cord-marked
O = Other
P = Plain

1. No net-impressed sherds of any type were collected.
2. No Tom's Creek sherds were collected.
3. No fabric-impressed, shell tempered sherds were collected.

APPENDIX G

LITHIC ARTIFACTS RECOVERED PER SITE

SITE NO.	FLAKES	CORES/PREFORMS	BIFACES
31NH601	1		
31NH604			
31NH605	1		
31NH606			
31NH607	8	1	
31NH608	2		
31NH609	2		
31NH610			
31NH611	3	1	
31NH612			
31NH613	1		
31NH614			
31NH615	1		
31NH616	5	2	1
31NH267A			
31NH267B			
31NH267C			
31NH267D			
31NH617			
31NH618			
31NH619			
31NH620	7		
31NH621			
31NH622			