ABSTRACT

Title of Thesis:	A DEEPER DIVE INTO THE WATER: A COMPARISON OF HYDROLOGIC FEATURES AS VARIABLES IN PRECONTACT SITE LOCATION PREDICTIVE MODELS FOR THE VIRGINIA PIEDMONT
	Jeffrey Wade Johnson, Master of Professional Studies, Cultural and Heritage Resource Management, 2024
Thesis Directed By:	Dr. Matthew M. Palus, Anthropology Department

The use of predictive modeling in Cultural Heritage Resource Management (CHRM) archaeology has become commonplace since its foundational principals were established in the 1980s, but criticisms of the practice persist, often centered around their lack of theory and dehumanization of the archaeological record. Proximity to water, typically expressed in the United States as distance to streamline data from the National Hydrography Dataset (NHD), is one of the most utilized variables when creating predictive models for Precontact period sites, but how does the variable "distance to streamline" compare to other hydrologic variables? In this thesis I seek to answer the question "how do distance to stream confluences and distance to wetlands compare to distance to streamline when attempting to predict Precontact site locations in the Virginia Piedmont?"

The publication *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling* (Altschul et al. 1988) is considered foundational to the practice of predictive modeling in archaeology; it is referenced frequently in modern theoretical works and throughout this thesis. The approaches to creating archaeological predictive models are typically divided into two camps: models that utilize an inductive, or correlative, approach and models that utilize a deductive, or theory driven, approach. Rather than establishing distance correlations between wetlands and stream confluences with previously recorded site data, I utilize a deductive approach where I establish the importance of those variables through archaeological theory pertaining to subsistence and settlement patterns and test their value with site data. Inductive associational models are very good at showing that archaeological site distribution is strongly patterned, but they often lack the explanatory framework that would be useful for management decisions based on their findings.

The Study Area the models are tested on is located within Orange County, Virginia near the town of Locust Grove, and encompasses about 686 acres. The Study Area contains two main streams, named Cormack Run and Mine Run, the confluence of those streams and other lower order streams, as well as wetlands located adjacent to the streams. Precontact occupations have likely occurred in this region for the past 12,000 years, if not longer.

The test results demonstrate that models created using deductively derived variables perform well enough to justify their use in CHRM contexts, but also include the added benefit of an explanatory framework. The guidelines for archaeological investigations in Virginia allow for the use of predictive models when conducting inventory surveys, meaning the archaeological predictive models (APM) created for this thesis could be utilized in a real-world context.

The primary focus of this thesis was to determine if using hydrologic features other than streams, specifically stream confluences and wetlands, to express the distance to water variable would improve the performance of an APM. I demonstrated that, yes, other hydrologic features may be better predictors of Precontact site locations in the Virginia Piedmont. Secondarily, I hoped to show that an APM created using a deductive approach would perform well enough to be considered appropriate for use in CHRM contexts. The high probability areas of all three of the APMs I created yielded Kg values high enough to be considered as having predictive utility. This demonstrates that the use of all three of the APMs I created could be considered appropriate to guide survey efforts in a CHRM context.

A DEEPER DIVE INTO THE WATER:
A COMPARISON OF HYDROLOGIC FEATURES AS VARIABLES IN PRECONTACT SITE LOCATION PREDICTIVE MODELS FOR THE VIRGINIA PIEDMONT

by

Jeffrey Wade Johnson

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Mater of Professional Studies
2024

Advisory Committee:
Dr. Matthew M. Palus, Chair
Dr. Kathryn Lafrenz Samuels
Dr. George Hambrecht

Acknowledgements

I would like to begin by thanking my advisory committee for their help, counsel, and contributions to the completion of my thesis. The chair of my committee, Dr. Matthew Palus, has guided me through every step of the process of completing this thesis. His insight, knowledge, and experience all contributed to the invaluable nature of his guidance, feedback, suggestions, and encouragement throughout the process of writing this thesis. Thank you to Dr. Kathryn Lafrenz Samuels for guiding this program, being my advisor during the early stages of my pursuit of this degree, and for your encouragement, guidance, and feedback that you gave to me. I would also like to thank Dr. George Hambrecht for taking the time to serve on my committee and providing constructive criticism of my work.

Next, I would like to express my appreciation of TRC for funding my education. The process of writing this thesis and obtaining my degree will undoubtedly contribute to my career at TRC and I am extremely grateful that I was able to do so without having to be concerned with the financial burden. Many thanks to Heather Millis, the Practice Leader of my office, for her support throughout the process of seeking my degree, providing feedback on my thesis, and assisting me in securing funding for my education. Special thanks to Matthew Paré, a Geographic Information System (GIS) specialist at TRC, for sharing his knowledge of APMs with me and assisting me with the GIS aspects of my research. A thank you should also go to Paul Webb, the Cultural Resources Market Leader at TRC, for suggesting I apply to this program at the University of Maryland (UMD) and encouraging me throughout my career and education.

I would also like to acknowledge every member of my cohort that proved their insights, feedback, and encouragement throughout the process of writing my thesis. Jordan Sly should also be recognized for being the Head Librarian for the Humanities and Social Sciences department at UMD and for providing guidance throughout the research process for my thesis.

Last, but certainly not the least, I would like to express my deepest gratitude to my wife, Chandra Wilson, for all the love and support she bestowed on me throughout this endeavor. I could not have undertaken this journey without her encouragement and willingness to take on a lion's share of my parental responsibilities while I was focusing on my studies. Her, and our son, Oliver, are my inspirations and motivations for bettering myself.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The use of predictive models for archaeological site location has been a common practice since the 1990s. The advances in Geographic Information System (GIS) capabilities over the past several decades allows for enormous amounts of data to be synthesized quickly, thus paving the way for an increased use of archaeological predictive models that attempt to predict the location of sites or artifacts within a defined area (Ebert 2004:320, Kohler and Parker 1986:400). Proximity to water is one of the most utilized variables when creating predictive models for Precontact period sites, but how does the variable "distance to streamline" compare to other hydrologic variables? I examine how two models, one using distance to wetlands and the other using distance to significant stream confluences, compare to a third model utilizing the traditional distance to streamline variable. Theories pertaining to Precontact Native American settlement patterns are the basis for selecting these three hydrologic features, and their predictive utility is tested using previously recorded site data from the Virginia Piedmont.

The inductive approach to model building is the most used methodology in CHRM contexts. The inductive approach to archaeological predictive modeling (APM) has largely been centered around correlations derived from previously collected quantitative data but is often criticized for not including archaeological theory and dehumanization of the archaeological record. A deductive approach to APMs uses archaeological theory to create the predictions, which provides an explanation for the findings that is often lacking from inductive models. A model based on archaeological theory could satisfy the needs of both CHRM professionals and future researchers. A Precontact period site location probability model can be based on archaeological theory, created from publicly available environmental data, and tested with previously recorded site information.

Distance to Water Variable

The importance of streams, stream confluences, and wetlands are all supported by Precontact settlement pattern theories. Wetland settings are not only sources of water but also support high levels of biodiversity. Settings near streams are considered attractive locations because the floodplains along them typically contain fertile well-drained soil, they are sources of water, and they are natural navigational routes. Stream confluences can be viewed as navigational points of interest, which is important to seasonal movement along the tributaries of major rivers. The formation of rapids occurs at stream confluences, which support diverse aquatic habitats.

The Study Area is drained by Mine Run and Cormack Run, as well as the lower order tributaries feeding into those streams. Mine Run and Cormack Run intersect with each other within the Study Area, and there are several other perennial stream confluences within the Study Area as well. There are also wetlands within the Study Areas, with the most extensive of them in the central portion of the Study Area along Cormack Run; smaller wetlands within the Study Area are in the western portion along Mine Run and in the southeastern portion along an unnamed tributary of Cormack Run.

Predictive Modeling in Archaeology

The theoretical principals of APM were largely established in the 1980s and 1990s, but the first uses date back to the 1970s (e.g., Green 1973). The publication *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling* (Altschul et al. 1988) is considered foundational to the practice of predictive modeling in archaeology and is referenced frequently in modern theoretical works and throughout this thesis. As defined by Kohler, a predictive model for archaeological site location is “a simplified set of testable hypotheses, based either on behavioral assumptions or on empirical correlations, which, at a minimum, attempts to predict the loci of past human activities resulting in the deposition of artifacts or alteration of the landscape” (Kohler 1988:33). The proliferation of models to predict archaeological site location in the United States has been attributed to the passage of the National Historic Preservation Act (NHPA) of 1966 and its requirement that federal agencies must identify cultural resources to be managed or that may be impacted by their actions (Kohler 1988:34, Verhagen and Whitley 2012:53).

Predictive models typically fall into two methodological approaches: inductive and deductive. The inductive, or “data driven,” approach typically begins with a sample of known site locations, which are then correlated to a series of environmental independent variables to determine which variables are good predictors of site location, and then those variables are used to extrapolate the potential locations of sites in areas that have yet to be investigated (Nebbia 2023:223). The deductive, or “theory driven,” approach begins with the definition of features that influence human spatial behavior that are based on theoretical assumptions, which are then modeled for an area within a GIS platform using either primary or derived datasets (Nebbia 2023:223). Remarkably, the methodology of predictive modeling was developed before GIS software became affordable in the mid-1980s (Verhagen and Whitley 2012:51). The increase in computing power available and the decreased cost of computer technologies has influenced the development of predictive modeling in archaeology. The use of GIS has not only expanded the application of archaeological predictive models but also has become an integral part of archaeological data collection in academia and CHRM alike. While predictive modeling can be used in a research context to test models of settlement patterning related to specific human locational behaviors (Verhagen and Whitley 2012:50), it is primarily used as a cost-saving or cost-predicting tool in CHRM (Nebbia 2023:223). The use of APMs is commonplace in modern CHRM contexts, but criticisms of the practice persist, which are typically focused on their lack of theory and dehumanization of the archaeological record.

Summary of Chapters

Theoretical Background

The theoretical background chapter is presented as the theoretical framework of this thesis. It begins with outlining the theoretical basis of archaeological predictive modeling. Next, the two generally accepted approaches to creating APMs, inductive and deductive, are introduced and discussed. The APMs created for this thesis were done so utilizing a deductive approach, which is discussed in greater detail in the following section of this chapter. That discussion includes the reasoning behind why a deductive approach to model building was favored over an inductive approach.

After establishing the use of a deductive approach to creating the APMs, the utilization of Precontact settlement pattern theories as the basis for variable selection is discussed. This section includes an explanation of how settlement pattern theories can be integrated into APMs and how they relate to the environmental variables chosen to create the APMs presented in this thesis.

Historical Background and Context

This chapter begins by describing the investigation that occurred within the Study Area and identified the sites used to test the APMs. Following that is an overview of the environmental context of the Study Area used to test the APMs. This includes a description of the physiological setting of the Study Area, as well as the hydrology, soil types, flora, and fauna that are present within the Study Area. These are all relevant factors to this thesis because they are related to the environmental variables that were chosen to create the APMs. Those sections are followed by a summary of the understood paleoclimate of Virginia and how climatic shifts relate to the cultural adaptations that occurred throughout the generally accepted Precontact cultural sequence for the Virginia Piedmont.

Following that section is a description of the cultural contexts relevant to this thesis. The APMs created are intended to predict the location of Precontact Native American sites within the Virginia Piedmont. The cultural context discussion is divided into three sections based on the generally accepted cultural sequence within the Virginia Piedmont: Paleoindian, Archaic, and Woodland. As supported in the theoretical background chapter, the cultural context is centered on settlement pattern theories relevant to each of the three temporal periods. A deductive approach to APM building begins with archaeological theory, which is why an understanding of the cultural contexts is important; it establishes the reasoning behind the variables selected for the APMs I created.

Methods and Data Collection

The methods and data collection chapter begins with a description of the data sources used to create the APMs and the workflow within ArcGIS Pro 3 that was used to integrate that data into the APMs. All three of the APMs created contain the soil drainage classification and percent slope data. That data was converted into a raster and desirability scores were assigned to the different areas based on their theoretical attractiveness to Precontact Native Americans. The models differ from one another through the inclusion of one of the three hydrologic features being examined for predictive utility: distance to streamlines, distance to stream confluences, and distance to wetlands. That data was also converted into a raster and desirability scores were assigned. Next, graphic representations of the rasters that represent each of the five variables used to create the APMs are presented. That section is followed by a description of how the three different APMs were created in ArcGIS Pro 3 using the "Raster Calculator" tool.

Following the model creation sections, the author describes the test used to measure each APM's performance, which is known as Kvamme's gain (Kg). This section describes why Kvamme's gain was selected as an appropriate test of the APM's performance. It then goes on to describe the equation used and what the implications are of using Kvamme's gain to test APMs.

Results

This chapter begins with a brief review of how APMs are tested, including a general description of the results of other APM tests that were reviewed for this thesis. Next, the results of creating the APMs are presented. That section includes a description of how the scores for each area were calculated and it provides graphic representations of each of the three APMs created to answer the research question.

The next section of this chapter presents the results of testing each APM using Kvamme's gain. It explains how the archaeological site data were overlayed onto the models and how the values within Kvamme's gain equation were calculated. This is followed by a brief description of the test results for each of the APMs. Each subsection contains a table that displays the values used to calculate Kvamme's gain and the results of that equation. The high probability areas within the distance to stream confluence model resulted in the highest Kg value.

Analysis

The analysis chapter starts with an analysis of the results of each model that were presented in Chapter 5. This includes a summary of the results for each model and a brief description of what those results indicate. That section is followed by an analysis of Kvamme's gain. That analysis includes a restatement of Kvamme's gain equation, the implications of using Kvamme's gain to test the performance, or predictive utility, of APMs, as well as criticism of the practice. Following that is a section that analyzes the Study Area chosen to test the model, including a discussion of the advantages and disadvantages the Study Area presents in relation to answering the research question.

Conclusion

The final chapter of this thesis contains the conclusions that can be derived from this research. It begins with a restatement of the research question and the answer to that question based on the results of this research. Next, the research process of this thesis is outlined. Following that is a comparison of the APMs along with a discussion what the results of the analysis imply about the APMs and their predictive utilities. Future research directions pertaining to the APMs created are discussed next, which is followed by a presentation of possibly methodological adjustments that may have improved or altered the test results of the APMs created. This chapter ends with the author's final conclusions of their research.

Chapter 2: Theoretical Background

APMs typically focus on statistical correlations between environmental data and known archaeological site location data. This approach to predictive modeling is statistically sound, but it is often criticized for not including the role of human agency in the creation of the archaeological record, not including enough archaeological theory, and being limited by the amount of archaeological site data available for any given location. While predictive models that start with theory and end with confirming site location are created, they are more often created in academic settings rather than in CHRM settings. The focus on site identification and National Register of Historic Places (NRHP) eligibility assessment in CHRM means that a predictive model can be effective without relying on intricate archaeological theories, but without the inclusion of theory, the model is likely not useful beyond CHRM contexts.

The predictive models presented here are intended to predict the locations of archaeological sites dating to all three of the generally accepted major Precontact temporal periods that are recognized within the Virginia Piedmont: Paleoindian, Archaic, and Woodland. To predict site locations encompassing all three of these wide-ranging temporal periods, general theories about Precontact settlement patterns are more useful than theories specific to only one temporal period or sub-period. There are many ways to approach predictive modeling in archaeology and the variability in types and locations of CHRM projects means that almost every type of model has some utility (Kincaid 1988:556). Virginia does not utilize a state-wide APM but does allow for predictive models to be used when conducting phase I archaeological surveys, albeit with little guidance on how these models should be created or what they must entail (VDHR 2017:20, 88). Creating a predictive model for Precontact archaeological site locations within the Virginia Piedmont that is based on archaeological theory would be applicable to academic researchers and CHRM professionals alike, which is favored over an approach exclusively based on statistical correlation that lacks an explanatory framework (Verhagen 2008:285).

Theoretical Framework of Archaeological Predictive Modeling

As defined by Kohler, a predictive model for archaeological site location is "a simplified set of testable hypotheses, based either on behavioral assumptions or on empirical correlations, which, at a minimum, attempts to predict the loci of past human activities resulting in the deposition of artifacts or alteration of the landscape" (Kohler 1988:33). Correlative predictive models have been favored by CHRM archaeologists because they utilize existing data and only require a modest time investment to create (Church et al. 2000:136). Cultural resources that do not contain artifacts or evidence of landscape alteration do exist, like Traditional Cultural Properties (TCPs) for example, but they are better understood and identified through collaboration with descendant communities, not archaeological investigations. Predictive modeling relies on quantitative methods, which means it is best suited for quantifiable archaeological sites. There is very little, if any, literature pertaining to models intended to predict the location of cultural resources like TCPs.

The proliferation of models to predict archaeological site location in the United States has been attributed to the passage of the National Historic Preservation Act (NHPA) of 1966 and its requirement that federal agencies must identify cultural resources to be managed or that may be impacted by their actions (Kohler 1988:34, Verhagen and Whitley 2012:53). In other words, Section 106 of the NHPA established the need to identify archaeological resources in the CHRM context and therefore created a source of funding to develop predictive models in the hopes that they would be beneficial for cutting costs during the resource identification process. The cost saving potential of predictive models is typically realized in either recommending more intensive survey efforts on high probability areas and less intensive survey efforts on low probability areas, or in avoiding areas of high probability all together during the planning phase of an undertaking. Federal agencies are required to identify all cultural resources that may be impacted, and the undertakings being considered can involve enormous expanses of land to be investigated. Project budgets in CHRM are typically larger than academic research budgets, but contracts are typically awarded through a competitive bidding process, thus creating the need for cost reducing measures like predictive modeling (Kohler and Parker 1986:398).

Predictive models typically fall into two methodological approaches: inductive and deductive. The inductive, or “data driven,” approach typically begins with a sample of known site locations, which are then correlated to a series of environmental independent variables to determine which variables are good predictors of site location, and then those variables are used to extrapolate the potential locations of sites in areas that have yet to be investigated (Nebbia 2023:223). The deductive, or “theory driven,” approach begins with the definition of features that influence human spatial behavior and are based on theoretical assumptions, which are then modeled for an area within a GIS platform using either primary or derived datasets (Nebbia 2023:223). These two methods of creating predictive models start from seemingly different ends of the theory-building process, with inductive modeling relying on previously recorded data to make predictions about site location and deductive modeling beginning with the predictions and then working towards proving or correcting them with field data.

In the context of CHRM, practitioners are often not concerned about *why* a site is where it is, but rather prioritize site identification and NRHP eligibility assessment. It is important to remember that "research provides information about the basic operation of past human organizational systems; the discard of materials from these systems; the incorporation of archaeological materials into what is discovered and seen as the archaeological record; and the ways in which archaeologists discover, measure, and interpret this record" (Ebert and Kohler 1988:100). This is a key point because it demonstrates that, even if it is not the intention of CHRM professionals, the data collected and the models created can and will be used for archaeological research, and that research can in turn be used to improve future predictive models.

Astonishingly, before GIS software became affordable in the mid-1980s, the methodology of predictive modeling was already being developed (Verhagen and Whitley 2012:51). The increase in computing power available and the decreased cost of computer technologies has influenced the development of predictive modeling in archaeology. The use of GIS has not only expanded the application of archaeological predictive models but also has become an integral part of archaeological data collection in academia and CHRM alike. As stated by Wescott, GIS is "a powerful and efficient managerial tool for spatial data sets, allowing the land or resource manager the ability to access, analyze, and interpret large amounts of archaeological data in a fraction of the time previously required" (Wescott 2000:1).

The primary use of APMs is as a cost-saving or cost-predicting tool in CHRM contexts (Nebbia 2023:223), but it is also used in research contexts to test models of settlement patterning related to specific human locational behaviors (Verhagen and Whitley 2012:50). Criticism of using predictive models is often centered around the idea that no model can predict the location of all archaeological sites with 100% accuracy. The fear is that “low probability” areas will not be surveyed with the same rigor as “high probability” areas, thus creating the potential that a significant site within a “low probability” area could be missed during an inventory survey. While predictive models are often used in CHRM contexts to differentiate the levels of effort utilized in various areas based on their potential to contain archaeological sites, this is not their only application to CHRM. Probability models can be used in the early planning stages of a project to provide zones of higher and lower archaeological site potential that forewarn project planners about possible costs associated with different locations, which allows for the avoidance of areas with high potential for containing archaeological sites before the survey is conducted (Verhagen and Whitley 2012:54). An issue with implementing a predictive model for the purpose of targeted levels of effort is that it can create a self-fueled bias that would influence future model development (Wheatley 2004a). In other words, if site identification efforts are only focused in areas determined to be “high probability,” then sites will only be found in those areas, thus creating a sort of confirmation bias within the site data when testing the performance of the model. Predictive models are more appropriate for targeting survey efforts rather than being used to justify the elimination of entire areas from survey all together. Some level of survey effort should be employed on every part of an undertaking that could impact cultural resources. Ideally, when testing the validity of a model, the entire area should be surveyed using an equal level of effort across the entire area to more accurately determine where sites occur and where they do not.

Deductive Archaeological Predictive Models

Due to the high variability of the types and locations of CHRM projects, just about every type of model has some utility (Kincaid 1988:556). The inductive approach to model building is the most used methodology in CHRM contexts because of how easy it is to create and understand, but a model based on archaeological theory provides an explanation for the findings that is often lacking from inductive models. Some have gone as far as to say that since inductive models lack an explanatory framework, they can never be considered adequate models (Verhagen 2008:285, Wheatley 2004a and 2004b). Once variables have been selected with support from archaeological theory, they can be mapped within the area of interest as being either favorable or not favorable to site location. Areas with multiple overlying favorable conditions can be seen as the most likely to contain sites and those with no favorable conditions can be seen as the least likely to contain sites (Altschul 1988:66). Areas with overlapping favorable environmental conditions are considered the most likely to contain archaeological sites because it is theorized that people chose the locations of their activities based on the optimal combination of important resources (Altschul 1988:61).

Predictive modeling in archaeology has largely been centered around correlations derived from previously collected quantitative data and is often criticized for not including enough theory during the model building process. A model based on archaeological theory could satisfy the needs of both CHRM professionals and future researchers. The advancement of technology coinciding with the decrease in cost of GIS applications over time has contributed to their expanded use in archaeology, especially for the purpose of creating predictive models (Ebert 2004:320). GIS is an essential tool for predictive model building because of its ability to analyze vast quantities of data in a short amount of time (Wescott 2000:1). A Precontact period site location probability model can be based on archaeological theory, created from publicly available environmental data, and tested with previously recorded site information. The inclusion of regional archaeological theory suggests that the models presented would be applicable in other areas of the Virginia Piedmont rather than relying solely on an arbitrary set of survey data that may limit a model's usefulness in other locations. A new model can be tested by applying it to surveyed areas accompanied with site location data to determine if it performs well at predicting Precontact site locations.

Inductive APMs have shown a strong correlation between proximity to water sources and Precontact site locations, which is typically expressed by "distance to stream." All five of the models reviewed for this research include some type of distance to water as a variable, either as distance to stream or distance to a combination of hydrologic features (Ahlman et al. 2005, Harris et al. 2015, Klein et al. 2015, Madry et al. 2006, and NCDOT 2017). Those models offer a variety of statistical calculations to demonstrate the correlation between Precontact site locations and proximity to water, but they do not provide an explanation as to why that is the case. Two of the five APMs, one created by the Pennsylvania Department of Transportation (Harris et al. 2015) and one created by the North Carolina Department of Transportation (Madry et al. 2006), utilize water sources other than streams as a predictive variable. The distance to water variables selected for those models were determined through statistical tests and used in conjunction with one another. I was unable to find a model that compares the predictive power of various types of hydrologic features beyond initial statistical correlations with known site locations during the selection of variables in the model building process, which is what inspired me to conduct this research. I created three separate APMs, with each one utilizing a different hydrologic feature, so that their performance could be tested against one another to demonstrate which hydrologic feature is the best predictor of Precontact archaeological sites. I based my selection of hydrologic features on Precontact settlement pattern theories in hopes of shedding light on why those features were attractive rather than simply demonstrating which features correlated best with Precontact sites.

Precontact Settlement Patterns

An important body of theory that has been used since the early days of creating predictive models is that of settlement patterns. Precontact settlement patterns are typically based on resource procurement, where reconstructions of important resource distributions, availability, and productivity are created using current environmental conditions as a baseline from which to work backwards (Parsons 1972:146). It seems that creating and testing predictive models based on proposed settlement patterns in areas that have not been archaeologically investigated would yield valuable information on the validity of proposed settlement patterns, as well as help refine hypotheses of proposed settlement patterns in the future. Including variables from settlement patterns into predictive models for site location can add an element of human behavioral patterns that may be lacking from models created primarily from statistical correlation. Probability models are typically created to predict site location, not necessarily cultural affiliation, so they should be effective across cultural boundaries, especially when interaction between groups creates fluid cultural exchange.

Two of the most common variables used in predicting the locations of Precontact archaeological sites are distance to water and slope (Verhagen and Whitley 2012:51, Nebbia 2023:224). Slope and soil drainage class are included in all three models presented and proximity to hydrologic features is the variable being examined in this thesis. Level areas on both uplands and floodplains are weighted favorably in all three models through the percent slope variable. Precontact archaeological sites are more frequently located on well-drained soils than poorly drained soils (Mandel and Bettis 2001:182). By including soil drainage class as a variable, areas of well-drained soil will be weighted higher than areas with poorly drained soils. Areas inundated with water, such as wetlands and marshes, are often exploited for their resources (Egghart at al. 2020:58, 109, 114), but submerged settings are not desirable locations for campsites or permanent settlements. The distance to wetland variable is intended to test if wetlands are a strong indicator of Precontact settlement. Wetland settings are not only sources of water but also support high levels of biodiversity. The distance to stream variable being tested gives more weight to floodplains than level uplands due to floodplains being adjacent to streams, which is important because floodplains are favored environments throughout all temporal periods while uplands became less favorable as subsistence patterns became more focused on horticulture and eventually agriculture (Egghart et al. 2020:14). Settlement patterns proposed for the Paleoindian, Archaic, and Woodland Periods focus on seasonal movement along the tributaries of major rivers (Anderson and Hanson 1988:8, Stewart 1995:186), thus demonstrating the importance of stream confluences as navigational points of interest as well as preferred locations for short- and long-term settlements (Egghart et al. 2020:56, 114, 147). Another theory as to why Precontact sites are found near stream confluences is the formation of rapids at those points, which support diverse aquatic habitats that were exploited for their resources (Seramur et al. 2007:44). Distance to stream confluence is included as a variable to test if these locations were favored over areas simply located along a stream.

Chapter 3: Historical Background and Context

The Study Area is located near the unincorporated town of Locust Grove in Orange County, Virginia (Figure 1). The Study Area encompasses approximately 686 acres, including areas of forest, cleared agricultural or pastoral land, residential properties, streams, and wetlands. The named streams flowing through the Study Area are Mine Run and Cormack Run, and there are several unnamed tributaries also within the Study Area that flow into those named streams.

Investigation of the Study Area

A cultural resources survey of the Study Area was conducted by TRC from April–May 2018 (Millis and Hayden 2018). The goal of that survey was to inventory archaeological and architectural resources over 50-years-old within an approximately 686-acre area in the northern Piedmont of Virginia and assess those resources using the criteria for the NRHP (Millis and Hayden 2018:35). The investigation of the Study Area involved systematic surface inspection, shovel testing, and metal detecting (Millis and Hayden 2018:35–37). All the open land within the Study Area was either overgrown fallow agricultural fields or pasture, meaning it had not been plowed recently and yielded little to no ground surface visibility, thus inhibiting the likelihood that Precontact period artifacts would be observed on the ground surface. The metal detecting was only conducted in areas with potential for containing Civil War era deposits (Millis and Hayden 2018:37). For those reasons, the surface inspection and metal detecting aspects of the investigation are not discussed further. A Virginia Department of Environmental Quality-approved Cultural Resources Sensitivity and Survey Plan (CRSSP) was used to guide the systematic shovel testing, which entailed shovel testing at 15-m intervals on 100% of the high probability areas, 25% of the medium probability areas, and 10% of the low probability areas (Millis and Hayden 2018:36). The CRSSP is essentially an APM in that it was based on both historic documents and environmental conditions and was created to predict an area's level of potential for containing both Precontact and postcontact sites prior to the actual investigation (Millis and Hayden 2018:30). Factors used to create the CRSSP include environmental attributes, such as proximity to streams, soil drainage classification, elevation, and percent slope, as well as historic documents, such as aerial photography and maps (Millis and Hayden 2018:30–33). In total, 6,673 shovel tests were excavated within the Study Area and 29 archaeological resources were identified; of those resources, 23 contained Precontact period components (Millis and Hayden 2018:41, 145).

Environmental Context

The Study Area is in the Outer Piedmont sub-province of the Piedmont physiographic province of Virginia (Roberts and Bailey 2000). The Outer Piedmont sub-province is a broad upland with low to moderate slopes (Roberts and Bailey 2000). The Study Area is situated in the eastern portion of Orange County with elevations ranging between 310–410 ft AMSL. The physiography of Orange County is well dissected and contains mostly rolling hills (Carter et al. 1971:164), and the Study Area reflects that physiography.

The Study Area is drained by Mine Run and Cormack Run, as well as the lower order tributaries feeding into those streams. Cormack Run, a tributary of Mine Run, flows from east to west roughly bisecting the Study Area. Mine Run flows south to north along the western edge of the Study Area. Mine Run flows generally north until it meets the Rapidan River approximately

5.5 miles north of the Study Area. The streams within the Study Area are represented in the APMs by flowline shapefiles obtained from the National Hydrography Dataset (NHD).

The most extensive wetlands within the Study Area are in the central portion along Cormack Run; smaller wetlands within the Study Area are in the western portion along Mine Run and in the southeastern portion along an unnamed tributary of Cormack Run. As defined by the United States Environmental Protection Agency (USEPA), wetlands are “areas where water covers the soil, or is present either at or near the surface of the soil all year or for varying periods of time during the year, including during the growing season” (USEPA 2024). Wetlands within the Study Area are considered riparian, meaning they are inland non-tidal wetlands that are commonly found on stream floodplains, isolated depressions, along the margins of lakes and ponds, or other low-lying areas (USEPA 2024). The wetlands within the Study Area are represented in the APMs by wetland shapefiles obtained from the United States Department of Agriculture (USDA).

The most prominent commercial soapstone (steatite) quarry in Virginia was in Rhoadesville, which is located a little less than five miles west-southwest of the Study Area (Webb and Sweet 1992:34). Despite no archaeological sites being recorded in the vicinity of that quarry (VDHR 2024), soapstone was likely procured from small quarries near present-day Rhoadesville. Precontact Native Americans quarried soapstone to create vessels, evident in the numerous soapstone vessel fragments found at archaeological sites (Webb and Sweet 1992:34). Soapstone vessels were primarily utilized prior to the proliferation of ceramic technology during the Woodland period, but soapstone continued to be utilized as a ceramic tempering agent throughout the Woodland period (Egghart et al. 2020:79, 84–85, 188–189).

There are 22 different soil units mapped within the Study Area (USDA NRCS 2024; Figure 2). Nathalie sandy loam is by far the most common soil type found in the Study Area, encompassing forty percent of the total area (USDA NRCS 2024). Nathalie sandy loam is well-drained, found on hillslopes, and formed from residuum weathered from granite and gneiss parent material (USDA NRCS 2024). Codorus silt loam, found within the floodplain of Cormack Run, is the second most common (13.3%) soil unit mapped within the Study Area (USDA NRCS 2024). Codorus silt loam is frequently flooded, somewhat poorly drained, and formed from alluvium (USDA NRCS 2022). Littlejoe loam and Clifford fine sandy loam are mapped within 7.9 percent and 7.6 percent of the Study Area, respectively (USDA NRCS 2022). Both Littlejoe loam and Clifford fine sandy loam are similar to Nathalie sandy loam in that they are well- drained, found on hillslopes, and formed from weathered residuum (USDA NRCS 2022). All other soil units (Table 1) individually encompass less than seven percent (or 48 acres) of the Study Area. This information was used to create the soil drainage classification variable of the APMs.

The Study Area is within the Northern Inner Piedmont Ecoregion (45e) as defined by the U.S. Environmental Protection Agency (Woods et al. 2003). A long history of land clearing, agriculture, logging, and other anthropogenic disturbance has severely altered the present-day vegetation of the Piedmont in Virginia, including the Study Area (VDCR DNH 2016:37). Those anthropogenic disturbances also likely impacted the integrity of archaeological sites within the Study Area. The natural flora of this area is mapped as Oak-Hickory-Pine Forest with dominant species including hickory (*Carya spp*.), shortleaf pine (*Pinus echinata*), loblolly pine (*Pinus taeda*), white oak (*Quercus alba*), and post oak (*Quercus stellata*) (Kuchler 1964, Woods and Omernik 1999:4).

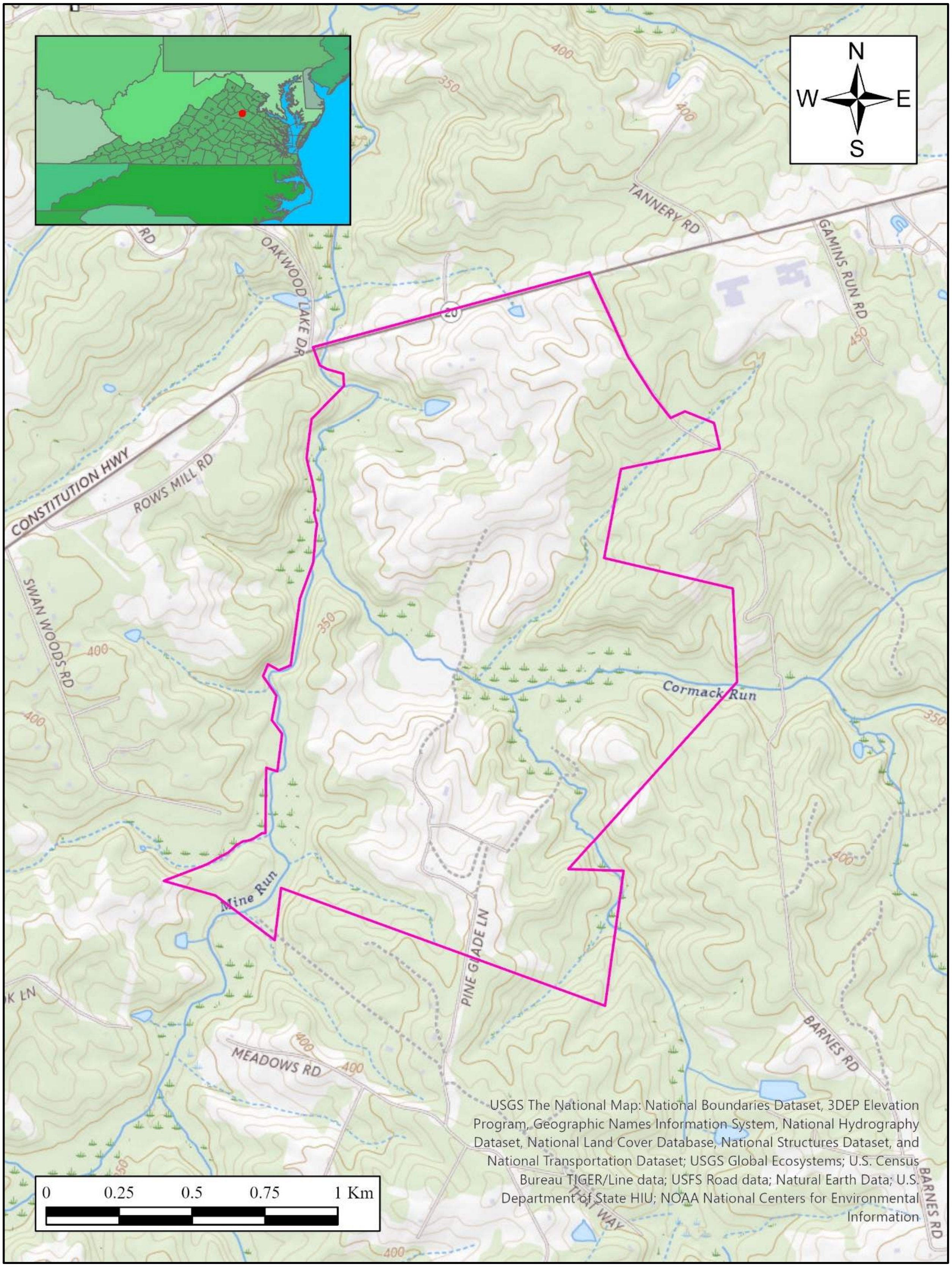

Figure 1. Project Area shown on USGS National Map (original work of the author).

Figure 2. Soil units mapped within the Project Area (original work of the author).

Table 1. USDA-NRCS Soil Units mapped within the Study Area

Map Unit Symbol	Map Unit Name	Acres in Study Area	Percent of Study Area
ApB	Nathalie Sandy Loam, 2–7% slopes	2.2	0.3%
ApB2	Nathalie Sandy Loam, 2–7% slopes, moderately eroded	78.9	11.5%
ApC2	Nathalie Sandy Loam, 7–15% slopes, moderately eroded	193.7	28.2%
BrC	Bremo silt loam, 4–15% slopes	6.0	0.9%
CeB2	Clifford fine sandy loam, 2–7% slopes, moderately eroded	33.5	4.9%
CeC2	Clifford fine sandy loam, 7–15% slopes, moderately eroded	18.4	2.7%
CmB2	Clifford loam, 2–7% slopes, moderately eroded	18.7	2.7%
CmC2	Clifford loam, 7–15% slopes, moderately eroded	7.5	1.1%
CsC3	Clifford clay loam, 4–15% slopes, severely eroded	3.0	0.4%
Cw	Codorus silt loam, 0–2% slopes, frequently flooded	91.0	13.3%
CxB	Belvoir loam, 2–7% slopes	38.2	5.6%
EsB2	Elsinboro loam, 2–7% slopes, moderately eroded	1.2	0.2%
LgB	Grassland silt loam, 2–7% slopes	5.8	0.8%
LoC	Philomont sandy loam, 5–15% slopes	17.4	2.5%
LoD	Philomont sandy loam, 15–25% slopes	2.9	0.4%
MrD	Bugley very channery silt loam, 15–25% slopes	6.4	0.9%
Mx	Hatboro-Codorus complex, 0–3% slopes, frequently flooded	21.4	3.1%
NaC2	Buffstat loam, 7–15% slopes, moderately eroded	0.1	>0.1%
OgA	Orange silt loam, concretionary variant, 0–2% slopes	>0.1	>0.1%
RnC	Rock land, acidic, sloping	25.9	3.8%
RnD	Rock land, acidic, moderately steep	5.0	0.7%
SeB	Meadowville fine sandy loam, 2–7% slopes	12.4	1.8%
SrC	Meadowville silt loam, 2–10% slopes	1.4	0.2%
TaB2	Littlejoe loam, 2–7% slopes, moderately eroded	28.6	4.2%
TaC2	Littlejoe silty clay loam, 7–15% slopes, moderately eroded	25.3	3.7%
TtC3	Littlejoe silty clay loam, 7–15% slopes, severely eroded	12.3	1.8%
TuB2	Germanna loam, 2–7% slopes, moderately eroded	5.9	0.9%
VaB	Casville fine sandy loam, 2–7% slopes	2.5	0.4%
W	Water	1.1	0.2%
We	Hatboro silt loam, 0–2% slopes, frequently flooded	14.7	2.1%
WoB	Delila silt loam, 2–7% slopes	5.6	0.8%

Common mammals found in the forests of the Virginia Piedmont include white-tailed deer (*Odocoileus virginianus*), Virginia opossums (*Didelphis viriniana*), eastern cottontail rabbits (*Sylvilagus floridanus*), eastern gray squirrels (*Sciurus carolinensis*), eastern raccoons (*Procyon lotor*), red (*Vulpes vulpes*) and gray (*Urocyon cinereoargenteus*) foxes, and at least 17 species of bats (*Chiroptera)* (Godfrey 1997:182–184). A wide array of avian, reptilian, and amphibian species are abundant in the forests of the Piedmont (Godfrey 1997). Leaf litter and animal remains that make their way into the Piedmont's rivers and streams provide a constant food source for abundant animal life, such as amphibians, fish, invertebrates, and gastropods (Godfrey 1997:243–244). These are all examples of the extant endemic species that would have been utilized by Native Americans prior to European contact. Streams and wetlands are also water sources for flora and fauna, meaning that areas near hydrologic features are more likely to contain an abundance and diversity of species that were utilized as resources by Precontact Native Americans than areas farther away from those features. Areas with concentrations of resources are more likely to contain archaeological sites resulting from resource procurement activities.

Paleoclimate

The paleoclimate of Virginia experienced significant change over the millennia since humans first began to occupy the area. The glacial maximum of the last glacial cycle occurred around 21,400 (calendar years before present (cal. BP) or 18,000 extended radiocarbon calibration before present (rcbp), which was followed by a 2,000-year post-glacial warming event (Anderson 2001:152). During the full glacial extent, boreal conifers like spruce and jack-pine dominated the forests of the northern part of the Eastern United States, which eventually began to give way to forests mostly comprised of oak and hickory after the onset of rapid deglaciation (Anderson 2001:152). This period of glacial retreat exhibited slowly rising sea levels with colder intervals possibly triggering minor reversals (Fiedel 2000). Diverse species of fauna were present in the Eastern United States during the Late Pleistocene, including extinct megafauna such as mammoth, mastodon, horse, giant sloth, saber-tooth tiger, and camel, as well as extant fauna like white-tailed deer, racoon, and rabbit (Anderson 2001:153). Over 30 genera of large animals, including all members of *Equidae, Camelidae,* and *Proboscidea,* were extinct by the end of the Pleistocene (Martin 1984:361–363). It seems likely that humans played some role in the mass extinction of the megafauna at the end of the Pleistocene (Martin and Klein, eds., 1984).

The earliest known evidence of human occupation in Virginia is from the Cactus Hill site (44SX202) on the Nottaway River in Sussex County, which has produced C14 dates of 15,070 +/- 70 and 14,180 +/- 80 BP (Egghart et al. 2020:31). The number of Early Paleoindian sites that have been discovered is very low, therefore most scholars must rely on broad-reaching regional perspectives rather than focused subregional data. The Middle Paleoindian or Clovis era is considered the period when human populations rapidly expanded across the Eastern United States and began to permanently settle in some areas (Anderson 2001:154). The sudden onset of a period of extreme cold spanning across the globe, called the Younger Dryas, marked the end of the Middle Paleoindian Period, and continued through much of the subsequent Late Paleoindian Period (Anderson 2001:155). The onset of Holocene climate and biota marked the end of the Clovis culture, which was followed by regional and subregional traditions becoming established, population growth, and a change in technologies to adjust to the new climate (Anderson 2001:155).

The transition from the Paleoindian Period to the Archaic Period coincides with the end of the Younger Dryas and the onset of the Middle Holocene Hypsithermal warming episode (Anderson et al. 1996:14–15). Rather than a higher average global temperature like we experience today, the Middle Holocene seems to have had hotter summers but colder winters in the northern hemisphere (Ganopolski et al. 1998). The Middle Holocene appears to have been a time of interrelated environmental stress and population pressure resulting from the cold winters and widespread aridity (Anderson 2001:160). Cultural trends initiated in the Middle Archaic continued to progress in the Late Archaic (5,700–3,2000 cal. BP/5,000–3,000 rcbp) during the period referred to as the Initial Late Holocene (Anderson 2001:160). It was during this period that essentially modern climate, sea level, and vegetation took hold (Anderson 2001:161).

Paleoclimate data from the Late Holocene is more abundant, which allows for more precise reconstructions of climatic events during this period. The Middle Woodland Period corresponds with the Roman Optimum, a period where environmental conditions appear to have been favorable for agriculture in Eastern North America, but it has yet to be determined if these climate conditions were an influential factor in that transition (Anderson 2001:164). Fairly dramatic cooling events occurred throughout the Late Holocene, which seem to roughly align with the Late Archaic/Early Woodland and Middle/Late Woodland transitions, but it is not clear if those events were causal or contributory to the change of those lifeways (Anderson 2001:164–165). Warming and cooling events continued for the remainder of the Precontact period, such as the Medieval Warm Period and the Little Ice Age, which continued to shape technological adaptations, subsistence strategies, and the cultural evolution of the Tribes inhabiting the land currently known as Virginia.

The modern fluvial regime of alternating cycles of cutting and filling of streambeds has been dominant for about the past 9000 years in Virgina (Haynes 2000), or since the later portion of the Early Archaic period, signifying that the modern riverine systems within the Study Area have been present for a majority of the time that there is evidence of human occupation in the Virginia Piedmont. That does not appear to be the case for the wetlands in the Virginia Piedmont. The USEPA estimates that more than 220 million acres of wetlands existed in the lower 48 states during the 17th century and today over half of those wetlands have been drained and converted to other uses (USEPA 2001). Wetland loss was largely attributed to the expansion of agricultural land in the past, but currently it is urban development that is causing the most loss of wetlands in the U.S. (USEPA 2001). Another example of why the extensiveness of wetlands in North America has declined since European contact is the near eradication of beavers. Beginning in the 1700s and continuing until about 1900, beaver populations in North America experienced a massive decline due to over trapping because of the demand for beaver pelt hats in Europe (Pollock et al. 2015:24). Beaver dams form surface water pools, which result in the expansion of riparian wetlands and wetlands along streams (Pollock et al. 2015:5–6). If there are no beavers to build dams that slow surface water flow, the stream channels become incised, which results in rapid draining of surface water that would typically form wetlands if drained slowly (Pollock et al. 2015:10–13). Fortunately, beaver populations are on the rise again, but there are still far fewer wetlands in the U.S. than are believed to have existed prior to European contact.

Cultural Context

The generally accepted Precontact cultural sequence in Virginia begins with the Paleoindian Period, which is followed by the Archaic Period, and ends during the Woodland Period with the arrival of Europeans on this continent. It is not exactly known when humans began to populate present-day Virginia. Clovis assemblages across the Eastern United States date to around 13,450 cal. BP/11,500 rcbp, but there is evidence from numerous sites that suggests the initial human occupation of the area could have begun many thousands of years before that (Anderson 2001:153–154). There is mounting evidence that humans arrived in the Americas well before the currently defined beginning of the Paleoindian period, but the number of sites containing that evidence is very low and the environmental conditions at that time were vastly different than today's, thus making it exceedingly difficult to predict the locations of those sites. As stated above, the Cactus Hill site has produced the earliest known evidence of human occupation in Virginia dating to approximately 15,000 years ago. The cultural context presented here is focused on subsistence and settlement patterns because theories about those patterns are the most relevant when creating models that predict Precontact site location.

Paleoindian Period

In Virginia, the Paleoindian Period began about 16,000 years before present (BP), or earlier, and continued until about 10,000 BP (Boyd 2020:31). The Paleoindian settlement pattern model for Virginia developed by Gardner (1977:258–259) includes five functionally separate site types: quarries, quarry reduction sites, quarry-related base camps, periodically visited hunting sites, and sporadically visited hunting sites. Quarries and quarry reduction sites tend to be in upland environments or near bluffs, quarry-related base camps and periodically visited hunting sites are typically within or near floodplain environments, and sporadically visited hunting sites have been found in upland and lowland areas (Turner 1994:72–74). The nearest Paleoindian site to this Project Area that has been investigated substantially is the Brook Run Site (44CU0122), a jasper quarry that includes both Paleoindian and Early Archaic components and is recorded approximately 10 miles north of the Study Area in Culpeper County (Monaghan et. al. 2004, Voigt 2001, Voigt et. al. 2004). Quarry sites are typically found in upland environments, but the Brook Run Site is located adjacent to a stream of the same name.

Archaic Period

The Archaic Period in Virginia began around 10,000 BP and lasted until about 3100 BP (Egghart et al. 2020:50–100). The Archaic Period is typically divided into three temporal subperiods: Early (10,000–8500 BP), Middle (8500–4500 BP), and Late (4500–3100 BP). Settlement-subsistence patterns and level of social complexity of Early and Middle Archaic peoples that inhabited the Middle Atlantic region are typically described similarly by archaeologists because environmental changes that occurred during those periods are similar when compared to those that occurred during the preceding Paleoindian or succeeding Late Archaic Periods (Parker 1990:100). Hypothesized subsistence models for the Early and Middle Archaic in Virginia involve seasonal movements of groups between different types of camps focused on the procurement of resources (Custer 1980, Garner 1974, Hoffman and Foss 1980). These camps generally fall into three categories: macro-band base camps located on floodplains and occupied during the Spring and Summer; micro-band camps located in the uplands and occupied during the Fall and Winter; and special procurement camps located wherever specific resources are plentiful during various times of the year (Parker 1990:101–102). Current models deemphasize fission-fusion cycles and instead propose a system involving small groups that occupy camps focused on a diverse set of seasonally available resources with frequent movements between camps spread over a relatively limited home territory and regular contact with neighboring groups (Egghart 2020:67). Both antiquated and modern models for Early to Middle Archaic settlement patterns in Virginia include mobility along river systems and the importance of wetland environments due to the high biodiversity they support.

Despite the environmental changes that occurred in Virginia during the Late Archaic Period, site distribution patters and locations are almost identical when compared to those of the Middle Archaic Period (Klein and Klatka 1991:155). The data Klein and Klatka presented suggests that a clear shift toward favoring riverine environments did not occur during the Late Archaic Period, but the authors acknowledged that this could be the result of survey bias (Klein and Klatka 1991:157). Two more recent studies also found that Late Archaic sites are fairly evenly distributed between both riverine and upland environments. Nash reports that Late Archaic settlements in Madison County, Virginia were documented on the full spectrum of topographic and ecological settings (Nash 2009). Barber found that Late Archaic component sites are not only the most documented but also are located on the widest variety of landforms and elevations (Barber 2010). All these studies postulate that the increase in the number of sites identified with Late Archaic components when compared to earlier components is an indication of population growth during that period, but again it is possible that this is the result of survey bias or site visibility rather than a significant increase in population.

Woodland Period

The Woodland Period is also commonly divided into three temporal subperiods: Early (3100–2500 BP), Middle (2500–1100 BP), and Late (1100 BP–European Contact). The Woodland Period is conventionally defined by the emergence of ceramic vessels in the archaeological record and a shift toward more sedentary subsistence strategies coinciding with the development of horticulture (Hodges 1991:225–232). Studies focusing on the transition from the Late Archaic Period to the Early Woodland Period suggest that populations continued to increase, the mobility of groups decreased, and riverine settings were utilized more than upland settings (Nash 2009, Klein and Klatka 1991). Nash found that stream confluences were most intensively utilized in Virginia's Piedmont during the Early Woodland Period and, Klein and Klatka observed that there were significantly more Early Woodland sites recorded on floodplains compared to upland settings in the Piedmont of Virginia. Egghart and Manson examined artifacts from a compilation of surface sites recorded by Robert Ogle along the Nottoway River Fall Zone in Sussex County, Virginia and showed that sites in this area with Early Woodland components tended to be located near swamps and wetlands (Egghart and Manson 2016).

The transition from the Early to Middle Woodland Period in Virginia's Northern Piedmont and a clear demarcation of the Middle Woodland from its preceding and succeeding subperiods have been difficult to establish because of the low number of Middle Woodland component sites that have been identified, especially when compared to the Coastal Plain where a majority of Virginia's Middle Woodland component sites have been recorded (Nash 2020:123–124). The generally recognized indicators of the Middle Woodland Period in the Piedmont of Virginia include relatively sedentary residence, a subsistence strategy focused on terrestrial resources and supplemented by an increased utilization of resources from riverine and estuarine environments, more complexity of social organization than previous periods, far-reaching trade networks, and population mobility (Stewart 1992:4). One study of Precontact sites in the Upper Rappahannock Valley found that the high order stream outer floodplain is the most common setting for sites dating to the later portion of the Middle Woodland Period (sometimes referred to as the Middle Woodland II), which is also the most common setting of Late Woodland sites in the Northern Piedmont, signifying a change in subsistence strategies taking place during the Middle Woodland Period (Nash 2009).

It was during the Late Woodland Period that the subsistence strategies of the inhabitants of the Northern Piedmont of Virginia shifted to permanent village life with a reliance on full-scale maize horticulture (Means 2020:163). The larger villages of the Late Woodland Period tend to be located along larger rivers (Kavanaugh 1983:49–52). It has been suggested that the fertile soils present on higher order stream floodplains is what drew Woodland people to those settings (Holland 1978:34–39). While there was a reliance on the cultivation of maize, hunting and gathering continued to be important aspects of the subsistence strategies of Late Woodland people in the Piedmont. Faunal analysis from five village sites in the Virginia Piedmont shows that terrestrial animals were the primary focus and aquatic species were utilized to supplement their diets (Barfield and Barber 1992:228). The end of the Late Woodland Period is typically recognized as the point of contact with Europeans sometime between 1600 and 1650, thus marking the end of the Precontact period in Virginia.

In general, sedentism increased throughout the Woodland Period in contrast to the more nomadic subsistence patterns of the Paleoindian and Archaic Periods. Archaeological sites predating the Woodland Period tend to be located on both upland and floodplain settings but become more concentrated in floodplain environments during the Woodland Period due to an increased reliance on horticulture. As subsistence strategies shifted toward horticulture, overall population and individual group sizes increased. Because riverine environments were utilized for resource procurement and as travel corridors during every temporal period, and most intensely utilized when populations were estimated to be the highest, it is reasonable to conclude that the number of sites would be the highest near streams and rivers. This explains why "distance to water" tends to be the strongest independent variable when creating probability models for Precontact period site location.

Conclusion

The review of the environmental context, paleoclimate, and cultural history of Orange County and the Virgina Piedmont presented in this chapter is intended to demonstrate why particular features within the Study Area influenced the location of Precontact sites. A comprehension of the relationship between the environmental and cultural factors that influenced Precontact Native American activities is of paramount importance when attempting to understand the "why" of site locations rather than simply the "where." This thesis attempts to answer the question "which hydrologic feature is the best predictor of Precontact site location?" while also shedding light on why that is the case. The following chapter discusses the data sources and methods used to create the three APMs as well as the test that was used to determine their performance in predicting Precontact site locations.

Chapter 4: Methods and Data Collection

Predictive models for Precontact site location often correlate current environmental data with previously recorded site location data to form their projections. As stated above, a review of past predictive models suggests that distance to water is one of the most used variables when attempting to predict Precontact archaeological site locations (Ahlman et al. 2005, Harris et al. 2015, Klein et al. 2015, Madry et al. 2006, NCDOT 2017). Creators of predictive models typically use large-scale statistical analysis to determine which data sets show the strongest correlations with Precontact site locations. While models created using statistical correlations are mathematically sound, they are often criticized for oversimplifying the archaeological record and their dependance on archaeological site location data that can be considered inaccurate or unreliable. Site location data is considered unreliable because survey coverage is fragmented across large areas and because survey methods have been highly variable over time, especially considering the numerous site locations recorded by avocational archaeologists. The models referenced above were all created for large-scale, high budget, multi-year projects, but most CHRM projects do not provide for the time, money, or workforce to create similar models. The primary question this thesis attempts to answer is which of the three hydrologic features selected (streams, stream confluences, and wetlands) is the most accurate and precise predictor of Precontact site location in the Virginia Piedmont. Secondarily, I hope to demonstrate that models based on archaeological theory, rather than purely relying on statistical correlations, have value when creating Precontact site location predictive modes, especially in smaller scale CHRM contexts.

The methodology presented in this chapter details what environmental data was used for these models and where that data was obtained from, how that data was used to create the three models, and how the models were tested. The model testing presented at the end of this chapter is intended to compare the predictive power of hydrologic variables, specifically whether distance to wetlands and stream confluences are stronger variables than distance to streams for predicting Precontact site location.

Data Sources

Soil Drainage Classification

The soil data are in the form of polygon shapefiles obtained from the United Stated Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS). The data were downloaded from the Web Soil Survey (WSS), which was produced by the National Cooperative Soil Survey (USDA NRCS 2024). The soil unit data include seven different soil drainage classifications, including excessively drained, somewhat excessively drained, well drained, moderately well drained, somewhat poorly drained, poorly drained, and very poorly drained (USDA NRCS 2024). None of the soils within the Study Area are described as excessively drained, but all the other six classifications are represented. Two other soil drainage classifications are also represented within the Study Area: “exposed bed rock” and “water.”

The soil data polygons were imported into ArcGIS Pro 3 and the “Clip” tool was used to isolate the soil data polygons within the Study Area. A “soil drainage rating” column was then added to the soil data attribute table. Each of the classes of soil drainage were assigned a rating based on their theoretical desirability to Precontact peoples, with a rating of three being the most desirable and a rating of one being the least. Precontact archaeological sites are more frequently located on well drained soils than poorly drained soils (Mandel and Bettis 2001:182). Based on that theory, well drained soils were assigned a rating of three, somewhat excessively drained and moderately well drained soils were assigned a rating of two, and all other classifications (somewhat poorly drained, poorly drained, and very poorly drained) were assigned a rating of one. The areas classified as exposed bedrock and water were assigned a rating of one as well because excavating shovel tests on those areas is not possible. These areas should not be completely excluded from the model because archaeological sites, such as rock shelters, artifacts exposed by erosion, petroglyphs, or cupules, could be present.

The soil data polygons were then converted to a raster format using the "Feature to Raster" tool with the soil drainage score field selected to assign the values of the output raster. The default cell size was used for this raster. The resulting raster (Figure 3) was used in all three models to calculate the overall desirability score using the "Raster Calculator" tool.

Percent Slope

The percent slope was derived from Digital Elevation Models (DEMs) acquired from the United States Geographical Society (USGS). The one-meter resolution DEM tile used was produced through the USGS National Geospatial Program (NGP) 3D Elevation Program (3DEP) utilizing high resolution light detection and ranging (LiDAR) source data of one-meter or higher resolution (USGS 2020). The Study Area falls completely within the USGS_one_meter_x28y415_VA Sandy_2014.tif tile published in 2020 and downloaded from the USGS ScienceBase Catalog (USGS 2020).

The DEM file was imported into ArcGIS Pro 3 and the "Clip" tool was used to isolate the portion of the tile encompassing the Study Area. The "Slope" tool was used to create a "percent slope" raster, with the clipped DEM as the input surface raster, the output measurement set to percent rise, and the cell size set to 15 m. A larger cell size was preferred when creating this raster because the default cell size results in a very detailed but fragmented raster. The "Slope" tool calculates the average slope (in percent rise for this case) within each cell. A cell size of 15 × 15 m was used because the VA DHR archaeological survey standards specify that shovel tests should be excavated at 15-m intervals.

Using the "Symbology" window, the resulting raster was manually classified into three categories based on the calculated present slope values for each cell: 0–5% slope, 5–10% slope, and greater than 10% slope. Next, the "Reclassify" tool was used to convert the slope categories into values, with 0–5% slopes assigned a value of three, 5–10% slopes assigned a value of two, and greater than 10% slopes assigned a value of one. More level areas were valued higher than more sloped areas because level areas have been valued throughout human history for both campsites and long-term settlements alike. Steep slopes are often impacted by erosion, which means artifacts deposited on steep slopes may erode down slope away from their original location of deposition. The resulting raster (Figure 4) was used in all three models to calculate the overall desirability score using the "Raster Calculator" tool.

Distance to Stream

The flowlines data for Orange County, Virginia used to calculate the distance to stream variable were obtained from the National Hydrography Dataset (NHD). The NHD was created and is managed by the USEPA. The polyline shapefile for stream flowlines was downloaded from the USGS NHD and imported into ArcGIS Pro 3 (USGS NHD 2022). To capture flowlines adjacent to but outside the Study Area, a polygon representing a 300-m buffer surrounding the Study Area was created using the “Buffer” tool. The “Clip” tool was then used to isolate the flowline polygons within the 300-m buffer of the Study Area.

The “Multiple Ring Buffer” tool was used to create buffer polygons at 100-m increments around the flowline polylines. A shapefile with three “distance to stream” buffer polygons was created, including 0–100 m, 100–200 m, and greater than 200 m (capped at 400 m to limit size while still encompassing the entire Study Area), and the maximum distance of each of those ranges is recorded in the “buffer distance” attribute field the tool creates. The ring buffers were then converted to a raster using the “Feature to Raster” tool, with the “buffer distance” field used to assign values to the output raster and the cell size set to 15 m. Next, the “Reclassify” tool was used to convert the distance to stream measurements into values, with 0–100 m assigned a value of three, 100–200 m assigned a value of two, and greater than 200 m assigned a value of one. The “distance to water” variable is one of the most commonly variables used in archaeological site predictive modeling (Ahlman et al. 2005, Harris et al. 2015, Klein et al. 2015, Madry et al. 2006, NCDOT 2017). Streams and the areas along them were important throughout the Archaic and Woodland periods in Virginia because they are sources of water, are areas of high biodiversity valuable for resource procurement, are natural travel corridors, and have nutrient-rich well-drained soils ideal for natural or agricultural plant growth.

The resulting raster (Figure 5) was used in conjunction with the percent slope raster and soil drainage raster to calculate the overall desirability score for all areas within the Study Area using the “Raster Calculator” tool. The final “Distance to Stream” model was categorized into three areas, with the “high” probability areas (red) receiving a total desirability score of 8–9, the “medium” probability areas (yellow) receiving a total desirability score of 6–7, and the “low” probability areas (green) receiving a total desirability score of 3–6.

Distance to Stream Confluence

There are no publicly available datasets for stream confluences. Also, I did not want to include the confluence of every ephemeral drainage with a first order stream, but instead focus on the confluences of perennial streams that could be considered water sources year-round. To capture said confluences, a point shapefile was created on ArcGIS Pro 3 and points were manually placed at the confluences of second order or higher stream confluences within the Study Area utilizing the USDA NHD flowlines shapefile. This is a practice that is applicable to a smaller-scale Study Area such as this one and likely impractical if applied to a large-scale predictive model. Three such confluences are present within the Study Area: one along Cormack Run, one along Mine Run, and one at the confluence of those two streams. To capture confluences adjacent to but outside the Study Area, a polygon representing a 300-m buffer surrounding the Study Area was created using the “Buffer” tool and two additional confluences that fell within that buffer were also included: one to the east of the Study Area along Cormack Run and one to the north of the Study Area along Mine Run.

The “Multiple Ring Buffer” tool was used to create buffer polygons at 300-m increments around the confluence points. An interval of 300-m was used for the confluence buffer areas so that the areas surrounding the confluence are included rather than just the precise point of intersection. A similar example of this can be seen in the North Carolina Office of State Archaeology’s “Archaeological Investigation Standards and Guidelines” that defines a stream confluence as located within 200 m of the meeting point of two streams (NC OSA 2023 Appendix B:10). Three sets of distance to stream buffer polygons were created, including 0–300 m, 300–600 m, and greater than 600 m (capped at 1,200 m to limit size while still encompassing the entire Study Area), and the maximum distance of each of those ranges is recorded in the "buffer distance" attribute field the tool creates. The ring buffers were then converted to a raster using the "Feature to Raster" tool, with the "buffer distance" field used to assign values to the output raster and the cell size set to 15 m. Next, the "Reclassify" tool was used to convert the distance to stream confluence measurements into values, with 0–300 m assigned a value of three, 300–600 m assigned a value of two, and greater than 600 m assigned a value of one. The "distance to confluence" variable has been used in archaeological site predictive modeling (Ahlman et al. 2005, Harris et al. 2015, Klein et al. 2015, Madry et al. 2006, NCDOT 2017), but it is not very common. Stream confluences were attractive to people during the Archaic and Woodland periods for all the same reasons streams were attractive, but confluences are natural intersections along waterways that were likely used for navigational purposes.

The resulting raster (Figure 6) was used in conjunction with the percent slope raster and soil drainage raster to calculate the overall desirability score for all areas within the Study Area using the "Raster Calculator" tool. The final "Distance to Stream Confluence" model was categorized into three areas, with the "high" probability areas (red) receiving a total desirability score of 8–9, the "medium" probability areas (yellow) receiving a total desirability score of 6–7, and the "low" probability areas (green) receiving a total desirability score of 3–6.

<u>Distance to Wetland</u>

The wetland dataset was downloaded from the Orange County, Virginia GIS data catalog on the Open Orange Data website. The county wetland data were derived from the USGS Wetland dataset that was created by the United States Fish and Wildlife Service's National Wetlands Inventory (NWI). The most substantial wetlands within the Study Area are along

Cormack Run in the central portion and Mine Run along the western boundary. To capture wetlands adjacent to but outside the Study Area, a polygon representing a 300-m buffer surrounding the Study Area was created using the "Buffer" tool, and the "Clip" tool was used to isolate the wetland polygons immediately surrounding the Study Area.

As with the other two hydrologic features, the "Multiple Ring Buffer" tool was used to create buffer polygons at 100-m increments around the wetland polygons. A shapefile with three "distance to wetland" buffer polygons was created, including 0–100 m, 100–200 m, and greater than 200 m (capped at 400 m to limit size while still encompassing the entire Study Area), and the maximum distance of each of those ranges is recorded in the "buffer distance" attribute field the tool creates. The ring buffers were then converted to a raster using the "Feature to Raster" tool, with the "buffer distance" field used to assign values to the output raster and the cell size set to 15 m. Next, the "Reclassify" tool was used to convert the distance to wetland measurements into values, with 0–100 m assigned a value of three, 100–200 m assigned a value of two, and greater than 200 m assigned a value of one. Utilizing "distance to wetland" is more common than proximity to stream confluence, but not as common as distance to streams (Harris et al. 2015). Wetlands are sometimes included as a water source and included in a "distance to water source" variable (Madry et al. 2006, NCDOT 2017). Wetlands were important the precontact period in Virginia because of the high biodiversity they support which is valuable for resource procurement.

I considered assigning the wetlands themselves a low score because they are typically inundated but ultimately decided to assign them the highest value (like the areas within 0–100 m of the wetlands) for multiple reasons. Firstly, areas likely to be inundated are already receiving a low score through the soil drainage classification variable. Second, while discussing my research question with a fellow CRM archaeologist, he suggested that I read an archaeological survey report written by New South Associates (NSA) for a wetland mitigation tract within the coastal plain region of southeast Georgia (Joseph et al. 2018). The first research question posed in that report pertains to whether archaeological sites would be identified in the wetland and floodplain portions of the tract, and if they were, could information from those sites alter the currently accepted settlement pattern model that suggests well-drained upland locations were preferred (Joseph et al. 2018:42). To effectively test for archaeological deposits in areas that are seasonally inundated, the fieldwork took place during three different sessions over the course of ten months (Joseph et al. 2018:1). The conclusions of that report state that numerous sites were identified in and adjacent to the wetland and floodplain portions of the tract, including some of the largest and most diverse assemblages recovered (Joseph et al. 2018:172). The physiological settings of that investigation and the Study Area of this thesis are different, but NSA's research suggests that wetlands should not be unequivocally considered areas of low potential for archaeological sites. Lastly, while the wetland polygons from the NWI are fairly accurate, they are not 100% reliable. I decided that it was best to assign a high score to the wetland areas because micro-landforms that are dryer or more elevated (such as hummocks) within or along the edges of the mapped wetlands may exist.

The resulting raster (Figure 7) was used in conjunction with the percent slope raster and soil drainage raster to calculate the overall desirability score for all areas within the Study Area using the "Raster Calculator" tool. The final "Distance to Wetland" model was categorized into three areas, with the "high" probability areas (red) receiving a total desirability score of 8–9, the "medium" probability areas (yellow) receiving a total desirability score of 6–7, and the "low" probability areas (green) receiving a total desirability score of 3–6.

While all three of the hydrologic features examined in this thesis have been used as variables in previously created Precontact site location models, they are not typically isolated from one another to test their individual predictive power. Testing these variables against one another can not only improve the predictive power of future models, but also contribute to subsistence strategy and settlement pattern theory building by examining what temporal periods or site types are being identified near different hydrologic features.

Model Creation

All three models began with the soil drainage raster and the percent slope raster, with their cell values based on their theoretical favorability to contain Precontact archaeological sites. Three separate models were created by including either the distance to stream raster, the distance to stream confluence raster, or the distance to wetland raster. The "Raster Calculator" tool was used to sum the scores of each cell, with a total score of nine signifying the most desirable areas and a total score of three signifying the least desirable areas. High probability areas are defined as the areas with total values of 8–9, the medium probability areas are areas with total values of 6–7, and the low probability areas are areas with total values of 3–5. The Precontact archaeological site polygons were then overlayed on the three probability models and Kvamme's gain was used to measure the predictive power of each model.

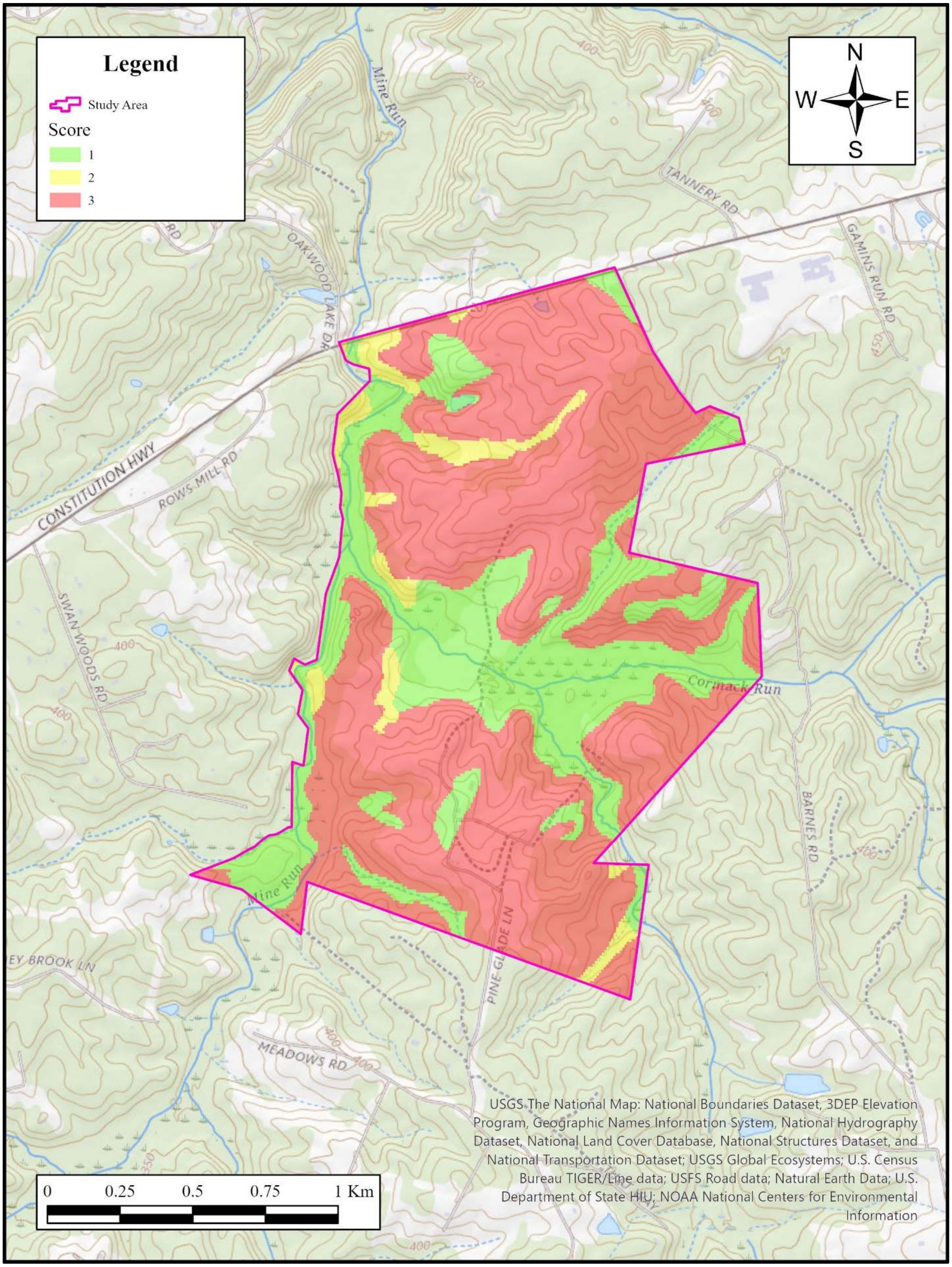

Figure 3. Soil drainage raster (original work of the author).

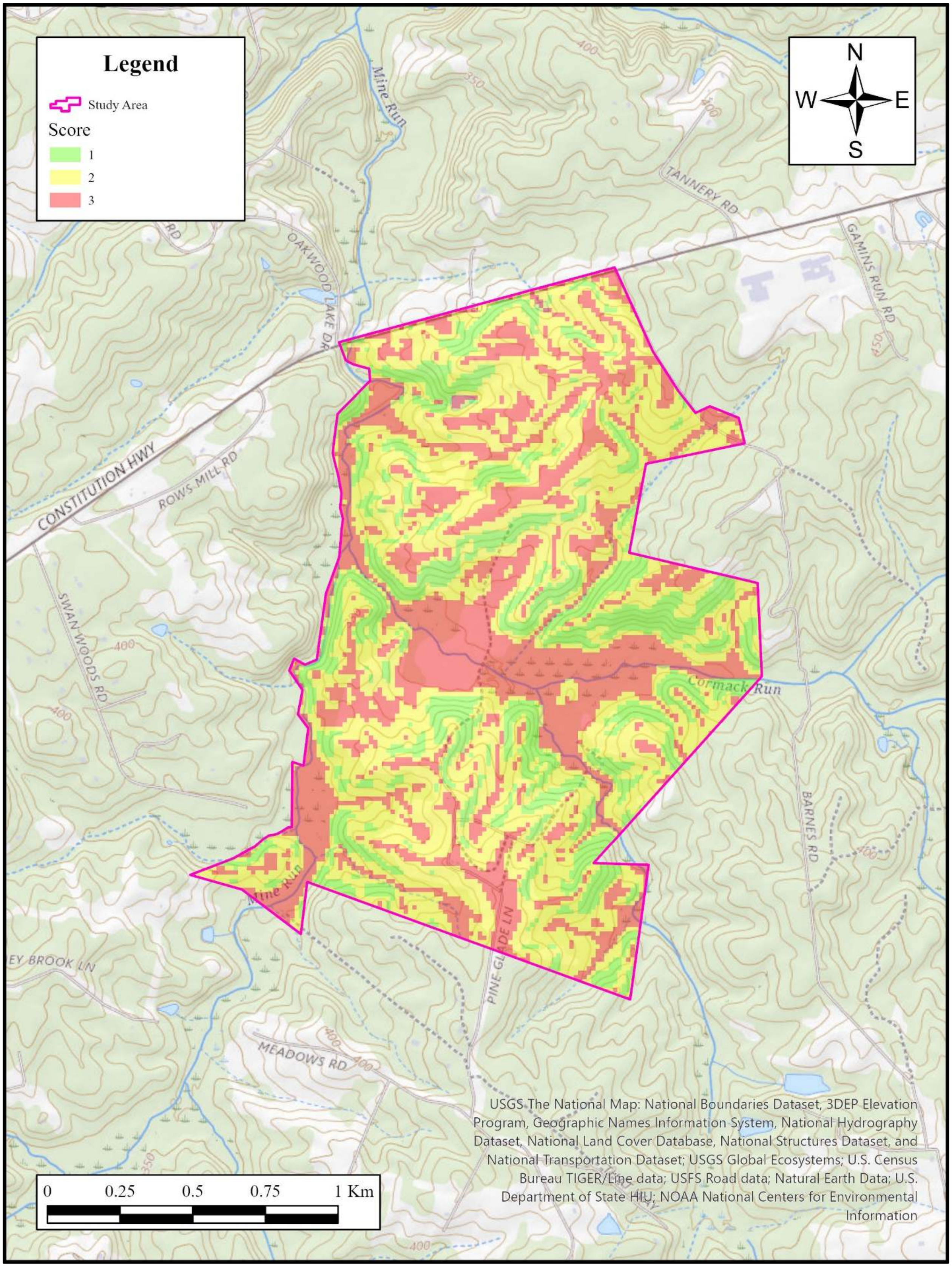

Figure 4. Percent slope raster (original work of the author).

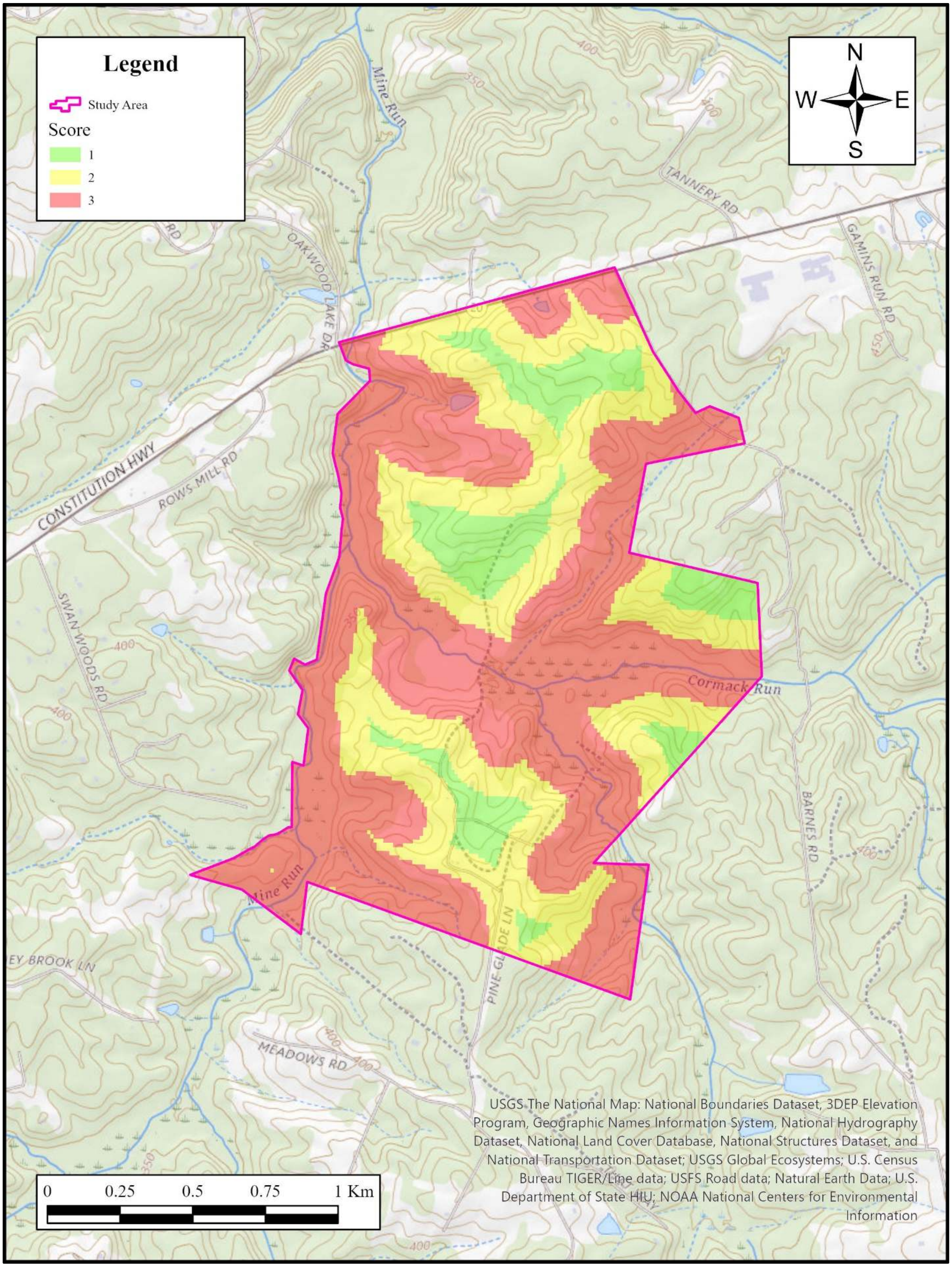

Figure 5. Distance to stream raster (original work of the author).

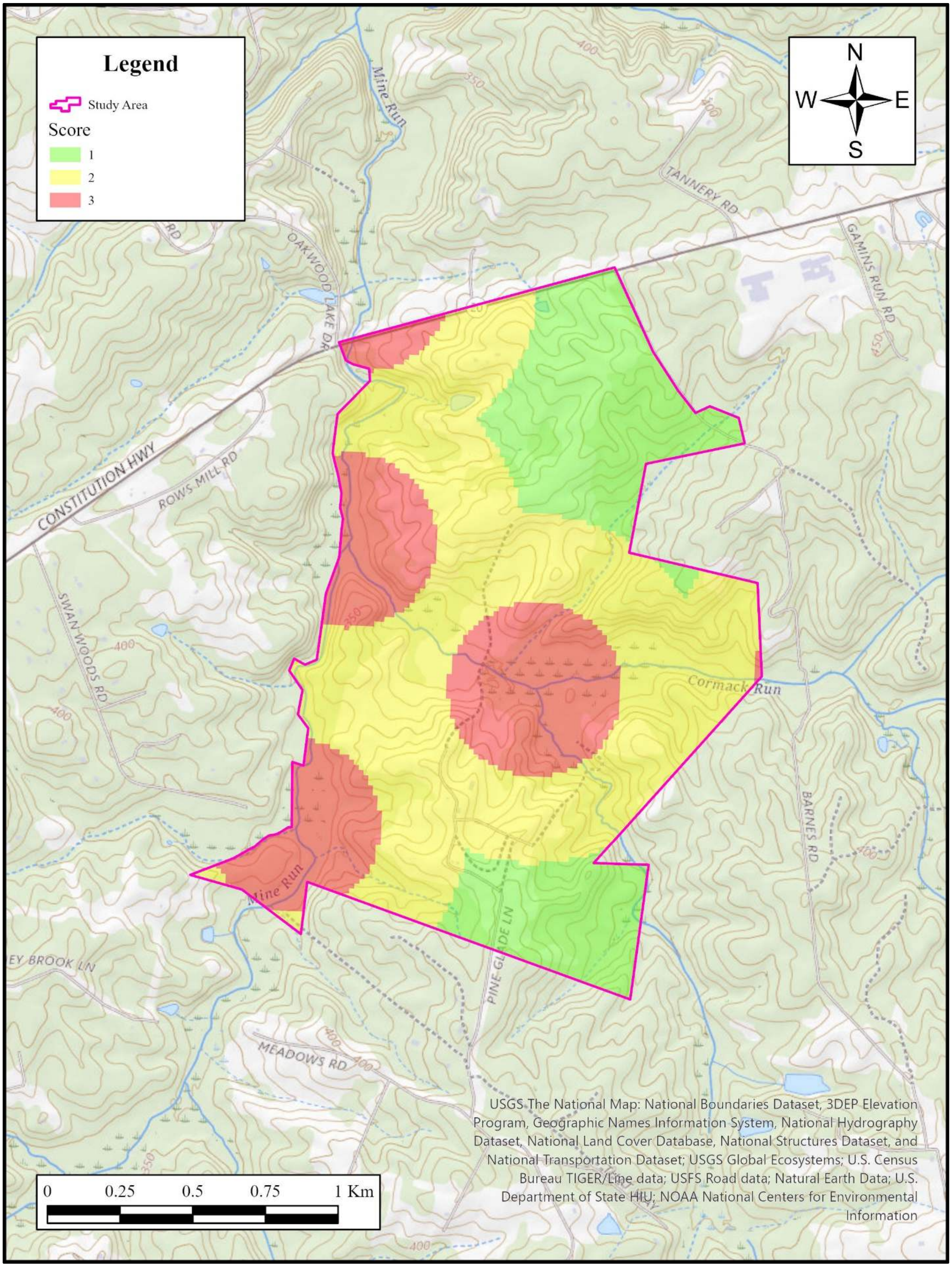

Figure 6. Distance to stream confluence raster (original work of the author).

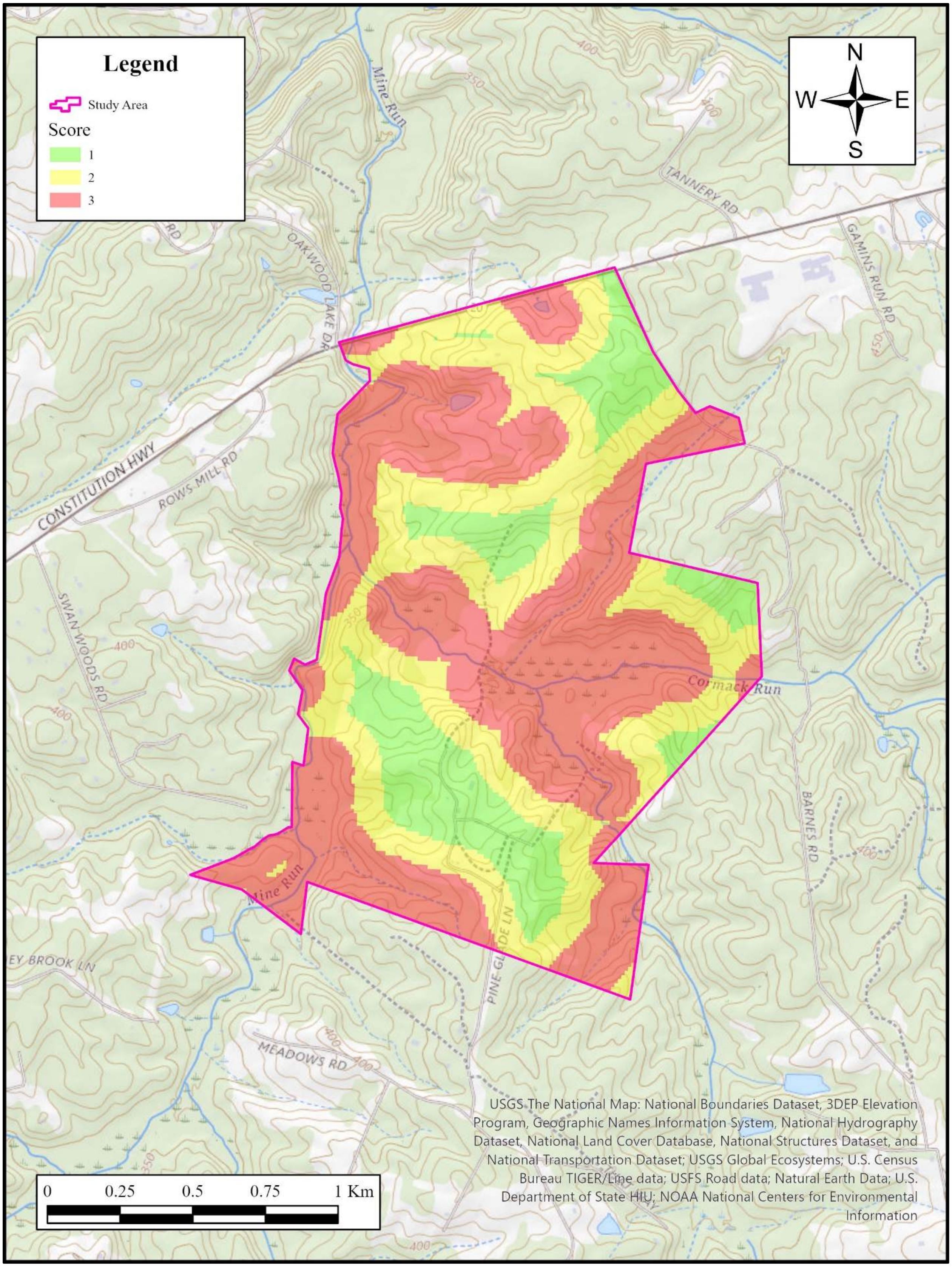

Figure 7. Distance to wetland raster (original work of the author).

Model Testing

Predictive models used in CHRM should be both accurate and precise (Verhagen 2008:286). A model can be very accurate by using a large “high probability” area that encompasses all the archaeological sites within a project area, but the lack of precision means the model is unlikely to save a significant amount of time or money when compared to simply surveying the entire project area as “high probability.” While precision is key to a predictive model’s value in CHRM, many archaeological survey guidelines specify that the methods used should attempt to identify all archaeological sites within a project area, and a model that prioritizes precision over accuracy is more likely to not identify archaeological sites than one that prioritizes accuracy. Kvamme’s gain is frequently used to measure the predictive power of an archaeological site location model because it combines both accuracy and precision into one easily calculated measure (Verhagen 2008:286). Kvamme’s gain equation is $1 - (p_a/p_s)$, with p_a equaling the percentage of total area covered by the model and p_s equaling the percentage of total sites within the model area (Kvamme 1988:329). The closer the gain value is to 1, the more predictive utility the model has; if the gain is near or approximately 0, then the model has little to no predictive utility (Kvamme 1988:329). If a model results in a negative (<0) gain value, then it has reverse predictive utility, meaning it is more likely that sites exist outside those areas than within them (Kvamme 1988:329). In Chapter 5, the calculated Kg values for all three different predictive models are presented to compare their usefulness in a CHRM context. A discussion of each model’s accuracy, precision, value in a CHRM context, limitations and shortcomings, and possible future revisions is presented in Chapter 6.

Chapter 5: Results

Predictive models for Precontact site location are frequently used in CHRM contexts, but they are usually created for large-scale undertakings such as the management of large areas of land or planning large-scale infrastructure development. There are few, if any, Precontact site location predictive models that do not utilize some form of the "distance to water" variable in their predictions. Tests and reviews of models often show that distance to water is one of the strongest predictors for Precontact site location, a conclusion that is typically reached through statistical analysis but is not often explained in a theoretical framework. By comparing these three models based on different hydrologic features, I hope to shed light on why these features are good predictors of Precontact site location rather than just showing that they are statistically relevant. My review of published predictive models showed that these models are typically created for high budget, large-scale, and long-term CHRM projects, but most CHRM undertakings are not of that scale. I hope that the models presented here demonstrate that predictive models can be created in a timely and cost-effective manner and be applied to smaller-scale CHRM projects without being detrimental to the quality of the investigation.

Results of Creating the Predictive Models

The three models presented began with two data sets: percent slope and soil drainage classification (see Figures 3 and 4). Then, each of the three hydrologic feature data sets, including distance to streamline, distance to stream confluence, and distance to wetland (see Figures 5–7), were overlayed with the percent slope and soil drainage datasets to create three separate predictive models (Figures 8–10). The “Raster Calculator” too was used to sum the overlapping weighted values for all areas within the Study Area, with a total score of nine signifying the most desirable areas and a total score of three signifying the least desirable areas.

Results of Testing the Predictive Models

Kvamme’s gain equation was used to measure the predictive power of each of the three models created. Kg values were calculated for each of the three probability areas (high, medium, and low) within each model. A raster is a matrix of cells organized into a grid, and each cell contains a value, or in this case, a desirability score (ESRI 2024). Since the probability areas were created using raster data sets, the total area of each model does not exactly match the total area of the smooth-edged Study Area polygon. The total areas used to calculate the Kg values are presented for each model, all of which are within one acre of the actual Study Area acreage. A total of 23 archaeological sites with Precontact components were identified within the Study Area (Millis and Hayden 2018:145), and the polygon shapefile for the site areas was used to determine which probability area each site would be encompassed by. Sites that fell within multiple probability areas were included with the higher-level (high > medium > low) area if 25% or more of its area was covered by the higher-level probability area. This was done because it is assumed that more intensive investigation strategies will be used within higher probability areas, and those more intensive strategies are likely to result in identification of a site even if only applied to a portion of the site.

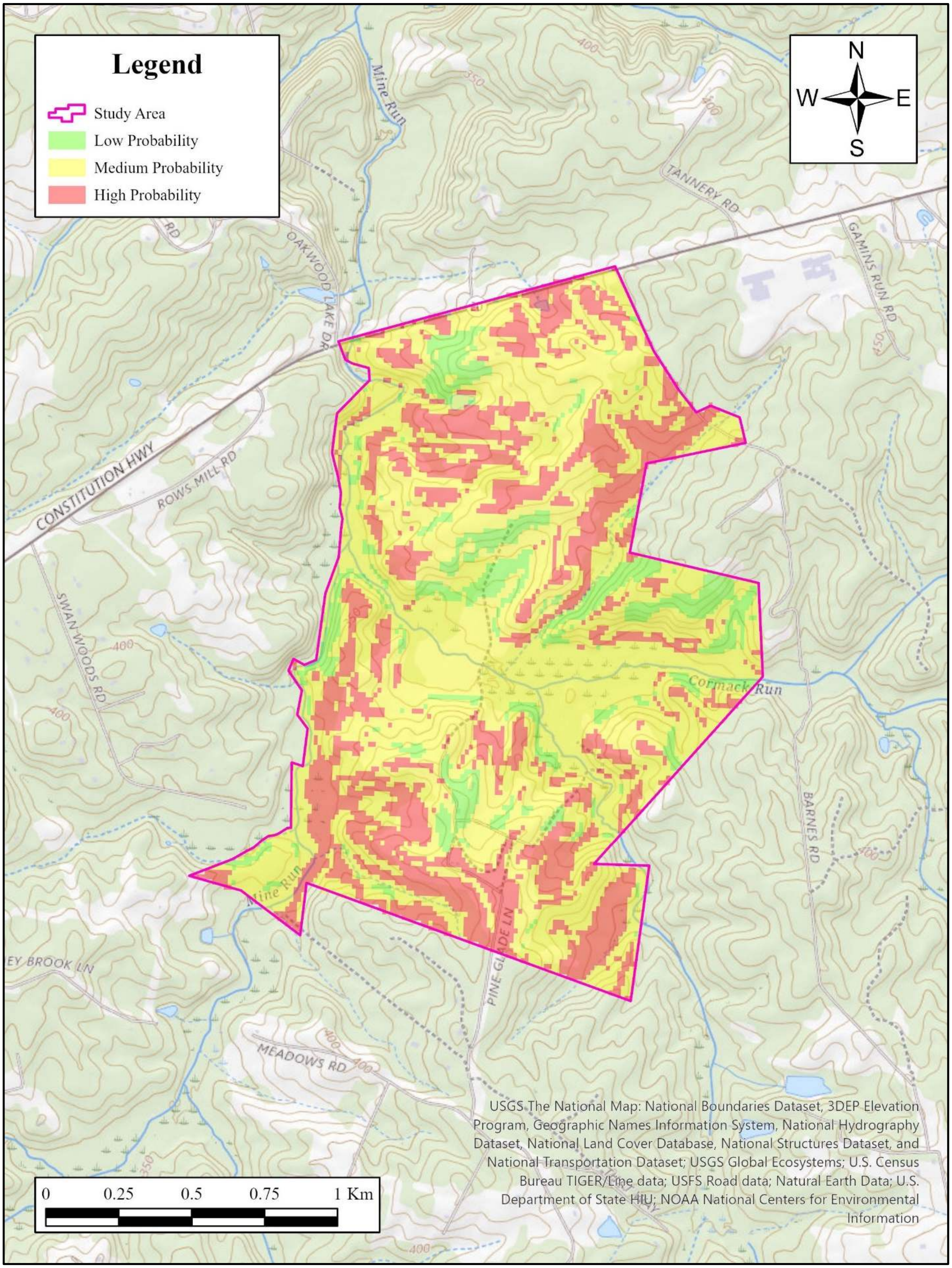

Figure 8. Distance to stream model (original work of the author).

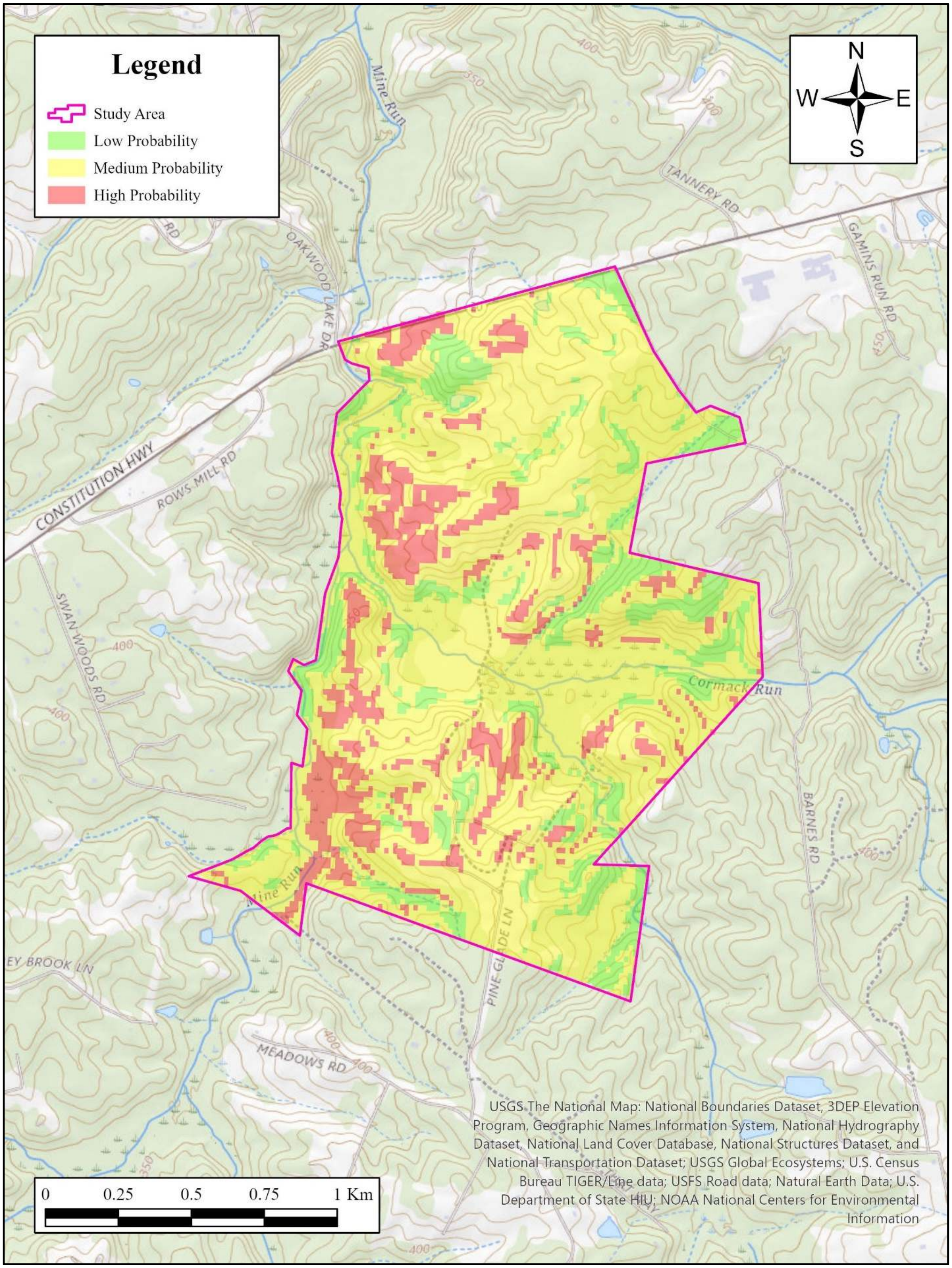

Figure 9. Distance to stream confluence model (original work of the author).

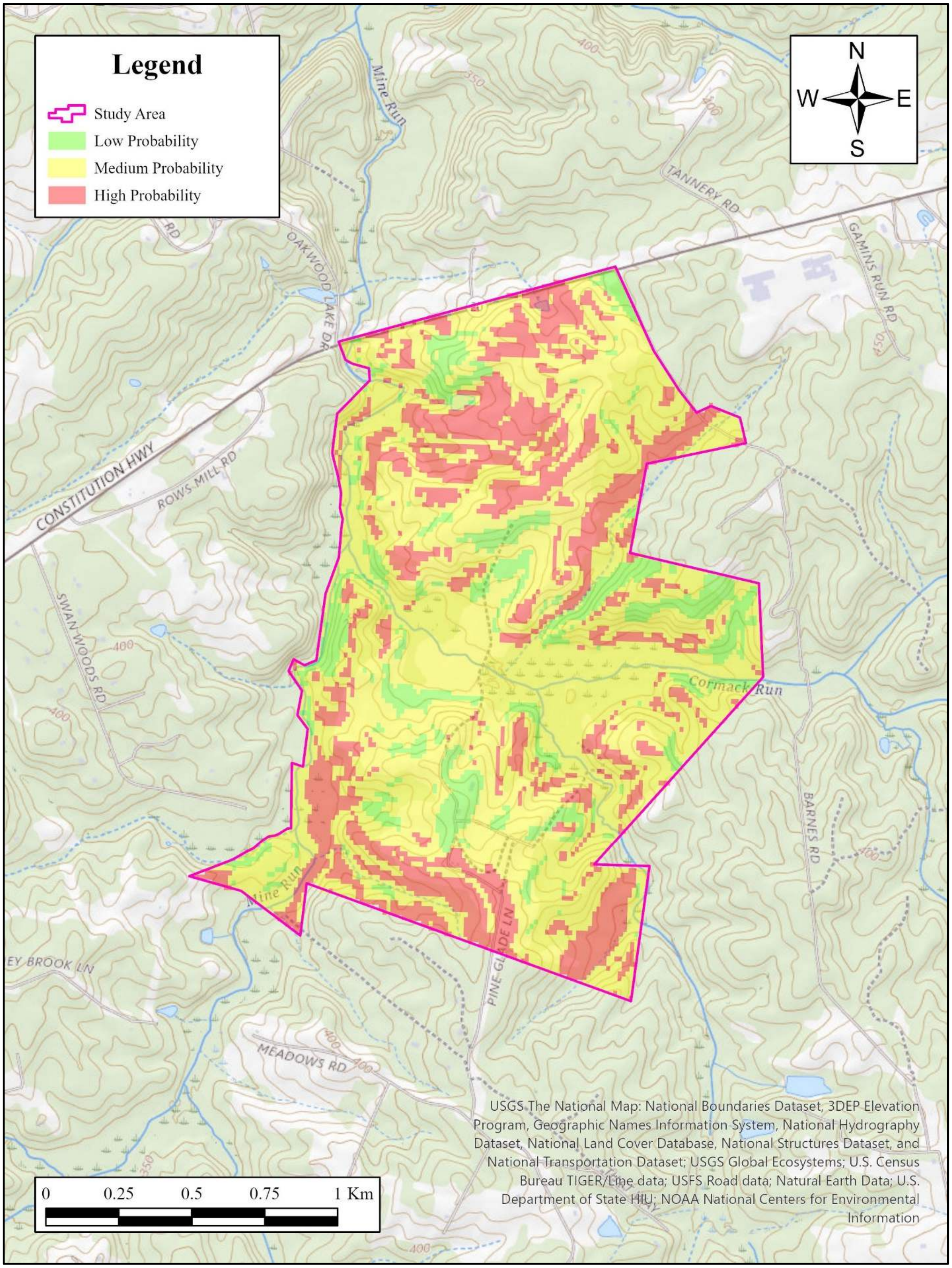

Figure 10. Distance to wetland model (original work of the author).

Distance to Stream Model

The model that utilized the distance to stream variable encompassed a total area of 686.9 acres. The high probability areas intersected with 17 of the sites, the medium probability areas intersected with five of the sites, and the low probability areas intersected with one of the sites. The high probability area totaled 202.7 acres within the Study Area, the medium probability areas totaled 413.3 acres, and the low probability areas totaled 70.9 acres. The percentages used and the calculated Kvamme gain for each area are presented in Table 2.

Table 2. Kvamme's gain for the distance to stream model.

Probability area	**Percent of total area**	**Percent of total sites**	**Kvamme's gain**
Low	10.3%	4.3%	-1.395
Medium	60.2%	21.8%	-1.761
High	29.5%	73.9%	0.601

Distance to Stream Confluence Model

The model that utilized the distance to stream confluence variable encompassed a total area of 687.6 acres. The high probability areas intersected with 15 of the sites, the medium probability areas intersected with seven of the sites, and the low probability areas intersected with one of the sites. The high probability area totaled 99.9 acres within the Study Area, the medium probability areas totaled 479.5 acres, and the low probability areas totaled 108.2 acres. The percentages used and the calculated Kvamme gain for each area are presented in Table 3:

Table 3. Kvamme's gain for the distance to stream confluence model.

Probability Area	**Percent of total area**	**Percent of total sites**	**Kvamme's gain**
Low	15.8%	4.3%	-2.674
Medium	69.7%	30.5%	-1.285
High	14.5%	65.2%	0.777

Distance to Wetland Model

The model that utilized the distance to wetland variable encompassed a total area of 686.9 acres. The high probability areas intersected with 17 of the sites, the medium probability areas intersected with six of the sites, and the low probability areas intersected with none of the sites. The high probability area totaled 176.2 acres within the Study Area, the medium probability areas totaled 427.1 acres, and the low probability areas totaled 83.6 acres. The percentages used and the calculated Kvamme gain for each area are presented in Table 4:

Table 4. Kvamme's gain for the distance to wetland model.

Probability Area	Percent of total area	Percent of total sites	Kvamme's gain
Low	12.2%	0%	N/A
Medium	62.2%	26.1%	-1.383
High	25.6%	73.9%	0.653

Conclusion

The three models created using the methods outlined in Chapter 4, with each model utilizing a unique hydrologic feature (streams, stream confluences, and wetland), are presented in this chapter. All three models began with raster datasets representing the soil drainage classifications and percent slope of areas within the Study Area. The different soil drainage classifications and ranges of percent slope were assigned values ranging from one to three, with a value of one assigned to the areas least likely to contain Precontact archaeological sites and a value of three assigned to areas most likely to contain those sites. The likelihood that these areas contain Precontact sites is based on archaeological theories about Precontact subsistence strategies and data presented in published reports on the effectiveness of predictive models created over the past 20 years (Ahlman et al. 2005, Harris et al. 2015, Klein et al. 2015, Madry et al. 2006, NCDOT 2017). Next, three separate models were created, each one utilizing a different raster created to represent proximity to either streams, stream confluences, or wetlands. The three hydrologic feature rasters were overlayed with the soil drainage and percent slope rasters, and the “Raster Calculator” tool was used to sum the three sets of values from each raster to create a final model for each hydrologic feature. Site boundary polygons for sites with Precontact components within the Study Area were then overlayed on each model. To test the predictive power of each model, Kvamme’s gain equation was applied to each model and a “gain” score was calculated for the three probability areas within each model. An analysis of the results of this test is presented in the next chapter.

Chapter 6: Analysis

This chapter presents an analysis of the results from Chapter 5. It begins by presenting the Kg value for each of the three probability areas within each of the three models created and an analysis of what each score implies about the predictive power of the models. Next, a discussion of the pros and cons associated with using Kvamme's gain to test the performance of predictive models is presented. Following that is a summary of the advantages and disadvantages of using the chosen Study Area as a test case for the predictive models. Finally, the three models are compared with one another to determine which hydrologic feature, when utilized in conjunction with other environmental variables, was the strongest predictor for Precontact archaeological sites within the Study Area.

Analysis of Test Results

As mentioned in Chapter 4, the closer the calculated Kg value is to 1, the more predictive utility the model has; if the gain is near or approximately 0, then the model has little to no predictive utility (Kvamme 1988:329). If a model results in a negative (<0) gain value, then it has reverse predictive utility, meaning it is more likely that sites exist outside those areas than within them (Kvamme 1988:329). Kg values were calculated for each of the three probability areas (high, medium, and low) within each of the three models (stream, stream confluence, and wetland). The high probability areas within all three of the models yielded Kg values greater than 0.5, and the medium and low probability areas all yielded negative gain values (Tables 2–4).

Distance to Stream Model

The high probability area for this model contained 73.9% of the sites within 29.5% of the total Study Area, equaling a Kg value of 0.601. This was the lowest gain value of a high probability area from the three models. The medium probability area resulted in a Kg value of -1.761 and the low probability area resulted in a Kg value of -1.395, meaning sites are more likely to occur outside those areas than within them.

Distance to Stream Confluence Model

The high probability area for this model contained 65.2% of the sites within 14.5% of the total Study Area, equaling a Kg value of 0.777. This was the highest gain value of a high probability area from the three models. The medium probability area resulted in a Kg value of -1.285 and the low probability area resulted in a Kg value of -2.674, meaning sites are more likely to occur outside those areas than within them.

Distance to Wetland Model

The high probability area for this model contained 73.9% of the sites within 25.6% of the total Study Area, equaling a Kg value of 0.653. This was the second highest gain value of a high probability area from the three models. The medium probability area resulted in a Kg value of -1.383, meaning sites are more likely to occur outside those areas than within them. A Kg value for the low probability area could not be calculated because none of the sites occurred within that area.

Kvamme's Gain

Kvamme's gain is frequently used to measure the predictive power of an archaeological site location model because it combines both accuracy and precision into one easily calculated measure (Verhagen 2008:286). Kvamme's gain equation is $1 - (p_a/p_s)$, with p_a equaling the percentage of total area covered by the model and p_s equaling the percentage of total sites within the model area (Kvamme 1988:329). To restate, as the Kg value of an APM approaches 1, the more predictive utility it has, and if the Kg value is near 0, it has minimal predictive utility; a model with a negative (<0) Kg value has reverse predictive utility, meaning it is more likely that sites are present outside those areas than within them (Kvamme 1988:329). This method of model testing is appropriate in CHRM contexts because models with higher gain values are ones that predict a larger number of sites in a smaller area, meaning they can be used to focus survey efforts and thus reduce costs.

Verhagen notes one of the drawbacks of using Kvamme's gain to assess the predictive power of a model, which is that equal gain values can be calculated with different values for accuracy and precision (Verhagen 2008:286). In other words, a Kg value of 0.5 can be achieved by both a model that includes 50% of the total sites in 25% of the total area and one that includes 90% of the sites in 45% of the area. Verhagen goes on to argue that a criterion for model quality that is rarely considered in the Netherlands is "does it achieve the goals set by either the authorities or developers (Verhagen 2008:286)?" A model may result in an adequate Kg value, but the model defines 45% of the total area as high probability, therefore it is unlikely that the model would be useful in achieving goals such as cost saving or planning to avoid sensitive areas. Verhagen suggests that to solve this problem, the entities creating archaeological predictive models should include precision limits as a criterion for acceptable models.

The high probably areas for all three models presented yield a precision value of less than 30% (see Tables 2–4), meaning none of the high probability areas encompass more than 30% of the total Study Area. Being able to reduce the intensity of survey methods on 70% or more of a Study Area or limiting areas where avoidance is recommended to less than 30% during project planning suggests that applying these models to this Study Area would have value in CHRM contexts.

The Study Area

The Study Area chosen to test the predictive models I created presents several characteristics that make this research relevant in both academic and CHRM contexts. Firstly, the Study Area is not massive. Most archaeological predictive models that have been published for public use were created for very large-scale undertakings, often encompassing multiple counties or entire states. It is understandable that an expensive endeavor like creating a viable predictive model would typically only be done for a large-scale well-funded project, but those types of projects are not the norm in CHRM. This thesis attempts to answer the question "proximity to which hydrologic feature is the strongest predictor of Precontact period site location?" While that is the focus, the results of testing these models imply that a model that performs sufficiently well can be created for smaller-scale project areas in a reasonably short amount of time, meaning they would not be considered prohibitively expensive to create. Complex models created for large-scale undertakings may be more statistically sound than these models, but those types of models have not been created for all areas where CHRM projects occur, they are sometimes only applicable to a specific environmental setting, and the methods used to create them may be proprietary. The growing number of undertakings requiring compliance with cultural resource laws and the diminishing number of people working in the industry means that time, labor, and cost saving measures like utilizing predictive models are going to become more common moving forward, meaning there is value in attempting to make models more accessible and applicable rather than more complex and specific.

Another advantage the Study Area presents is that it is a sufficiently sized sample area that has been completely investigated. A criticism of using site data derived from CHRM archaeology to create predictive models is that the data is fragmented across a patchwork of project areas where a variety of survey methods were employed, meaning the data contains survey bias. The Study Area used to test the models created has been completely investigated using a single survey methodology.

Lastly, the Study Area contained all three of the hydrologic features being used to create the separate models, which presents both advantages and disadvantages in this analysis. If all three of the hydrologic features being examined did not exist within the Study Area, it would be difficult to compare their predictive potential using the data from this survey. At the same time, one could argue that one or both features not being examined in a model are drawing Precontact people to the area rather than the feature included in the model. In other words, was the whole area attractive to Precontact Native Americans because of its biodiversity, or did the individual hydrologic features draw people to this area? It should be noted that these three hydrologic features are related to one another: confluences are intersections of streams and wetlands are often located along streams. Therefore, if an area contains streams, it is likely that it also contains confluences and wetlands.

One of the drawbacks of the Study Area used to test these models is its relatively small footprint when compared to others that have been used to test models in the past. As mentioned above, the Study Area contains streams, confluences of those streams, and wetlands. Would the models perform as well on smaller areas that do not contain all three, or any, of these hydrologic features? Would the models perform as well on larger areas containing major rivers, confluences of higher order streams, and more extensive wetlands? A future direction of this research would be to apply the models to other areas of the Virginia Piedmont, including both smaller and larger Study Areas, to see how the different models perform in different settings.

Another drawback of the Study Area is the methodology used to investigate it. A CRSSP was utilized to conduct the survey of this Study Area that was created using, in part, soil drainage classifications and percent slope, both derived from the USDA NRCS soil data (Millis and Hayden 2018:30). To account for the CRSSP used during site identification, both of those datasets were included in all three of the models. The soil drainage classifications are the same as those used to create the CRSSP utilized to conduct the survey of the Study Area, but the slope data was improved upon. The elevation data used to calculate the percent slope of the Study Area for this thesis was published in 2020 (USGS NGP 2020), meaning it was not available when the survey was conducted in 2018. Using LiDAR data to calculate percent slope is significantly more precise than using slope percentages from the USDA NRCS soil data, thus improving upon the data used to create the models tested in this thesis while still holding true to the datasets applied during the identification of the archaeological sites within the Study Area. While it is true that the locations where Precontact sites were identified during the survey contain bias relating to the CRSSP used at that time, that can be said about all site location data. The methods used during site identification are rarely accounted for when creating probability models from site location data because it is too time consuming and costly to account for the wide array of survey methods that have been applied in the past (Verhagen 2008:289). Site location data is riddled with bias, which is one of the major arguments against the inductive approach to creating archaeological predictive models (Wheatley 2004a and 2004b). Bias is present within the site location data used to test the models presented in this thesis, but it represents only one survey methodology rather than dozens, or hundreds, of different methodologies.

Conclusion

An analysis of the results of this thesis project is presented in this chapter. The Kg values calculated for each of the three probability areas within each of the three models are presented again along with a discussion of what each score implies about the predictive power of the models. That section is followed by a discussion on the use of Kvamme's gain to assess the performance of predictive models, including advantages and criticisms of the practice. Lastly, the pros and cons of using the presented Study Area to test the models are discussed.

Chapter 7: Conclusion

The use of predictive modeling in CHRM archaeology has become commonplace since its foundational principals were established in the 1980s. The advancement of technology coinciding with the decrease in cost of GIS applications over time has contributed to their expanded use in archaeology. That said, criticisms of the practice persist, often centered around their lack of theory and dehumanization of the archaeological record. Proximity to water, typically expressed in the United States as distance to streamline data from the NHD, is one of the most utilized variables when creating predictive models for Precontact period sites, but how does the variable "distance to streamline" compare to other hydrologic variables? The association between Precontact site locations and proximity to streams has been well established, but statistical correlations do not provide an explanation as to why that is the case. I was unable to find a model that compares the predictive power of various types of hydrologic features beyond the initial statistical correlations with known site locations during the selection of variables in the model building process, which is what inspired me to conduct this research. The primary goal of this thesis is to determine which of the three selected hydrologic features (streams, stream confluences, and wetlands) is the best predictor of Precontact site locations, but I also hope to demonstrate that models based on archaeological theory instead of statistical correlations have value in CHRM contexts. Creating APMs based on archaeological theory allows for those theories to be tested in real world situations, which benefits both academic and CHRM archaeologists alike.

The previous chapter presents an analysis of the results of testing the APMs created, while this chapter serves to discuss the relevance of this research. It begins with the research question and the answer to that question my research supports. Then, a summary of the methods used throughout the research process is presented. That section is followed by a comparison of the three APMs created for this thesis. Next, future directions of research pertaining to theory driven APMs and possible methodological changes are suggested. Finally, a summary of the important themes of this research is discussed.

Research Question

Proximity to water is one of the most used variables when creating APMs. In the United States, this is typically expressed though distance to streamline data from the NHD. It is understandable that the NHD data are used because they are available for the entire United States, are free to access and download, and streams are often considered permanent water sources. Other sources of water were also available to Precontact Native Americans, such as wetlands. Also, specific areas along streams, such as their confluences with other streams, present environmental conditions that may not be found along the entirety of the stream. This thesis is focused on the question "how do distance to stream confluences and distance to wetlands compare to distance to streamline when attempting to predict Precontact site locations in the Virginia Piedmont?" Based on my research, the answer to that question is that stream confluences are the best predictor in the context of this study. Interestingly, the distance to wetland APM also outperformed the distance to streamline model. The results of the performance tests suggest that a greater variety of water sources should be considered as variable when predicting Precontact site locations within the Virgina Piedmont. My research has also demonstrated that deductive APMs are useful in CHRM contexts because all three models performed well (Kg values over 0.5), they were relatively simple and quick to create when compared to other APMs created for CHRM undertakings, and they can be applied to locations anywhere within the Virginia Piedmont because they are based on Precontact settlement pattern theory relevant to the entire region.

The Research Process

The process of answering the research question began with a review of settlement pattern theory pertaining to Precontact Native Americans in the Virginia Piedmont. Based on that research, hydrologic features that could be expressed through environmental data were selected to create three separate APMs. All three models began with soil drainage classifications and percent slope as a baseline. Next, proximity to one of the three hydrologic features selected, including streamlines, stream confluences, and wetlands, were overlayed with the baseline model to create three separate APMs that could be compared to one another.

Once the three APMs were established, they were then applied to a Study Area that has already been archaeologically investigated. The Study Area I selected to test the models was chosen because it has been fully investigated, meaning that it contained Precontact site location data and did not contain gaps in survey coverage, and because it is characterized as a sizable, but not excessively large, area that is representative of project areas commonly encountered by CHRM professionals. Also, the Study Area contains all three of the hydrologic features I wanted to compare, which is convenient for testing purposes but not unexpected for an area as large as the Study Area within the Virginia Piedmont. A CRSSP was used to investigate the Study Area in 2018 (Millis and Hayden 2018), which is not ideal but also not entirely problematic. The probability areas within the APMs I created are significantly different than the one used to investigate the Study Area. The CRSSP used in 2018 was intended to predict both Precontact and Postcontact sites. Also, probability models are used in Virginia to guide survey efforts, meaning that an inductive model based on previously recorded site data would also contain that bias, but at least in this case only one methodology was used to identify the sites within the Study Area.

Once all three APMs were applied to the Study Area, the polygon shapefiles representing Precontact site areas were overlayed. A total of 23 archaeological sites with Precontact components were identified within the Study Area, and their boundaries were used to determine how many of the sites were located within each probability area. Sites that fell within multiple probability areas were included with the higher level (high > medium > low) area if 25% or more of its area was covered by the higher level probability area.

Next, each model's performance was tested using Kvamme's gain. Kvamme's gain equation is $1 - (p_a/p_s)$, with p_a equaling the percentage of total area covered by the model and p_s equaling the percentage of total sites within the model area (Kvamme 1988:329). The Kg values for each of the three probability areas within each of the three models were calculated. The resulting values were used to evaluate which model was best at predicting Precontact site locations. Based on the Kg values of the high probability areas within each model, it was determined that the distance to stream confluence variable performed the best at predicting Precontact site locations. This research process effectively describes how to select variables to include in an APM using a deductive approach, how to create an APM utilizing publicly available environmental datasets that reflect those selected variables, and how to test the performance of the resulting models using Kvamme's gain equation.

Comparison of the Models

Three models were created to compare the predictive power of streams, stream confluences, and wetlands in relation to Precontact site location. Kg values were calculated for each of the three models to test their predictive utility. The high probability areas within each model yielded Kg values over 0.5, meaning they have predictive utility (Tables 2–4). The medium and low probability areas all yielded negative Kg values, meaning it is more likely that sites occur outside those areas than within them. Negative Kg values for low probability areas are indicative of a quality model, but not necessarily for medium probability areas. In these cases, the medium probability areas within each model resulted in negative Kg values because they all encompass more than 60% of the Study Area while predicting less than 1/3 of the total sites (see Tables 2–4).

Many archaeological predictive models only define high and low probability areas, with high probability areas often recommended for survey or avoidance during planning and low probability areas recommended for no survey, less intensive investigation, or use during planning. Adjusting the goals or recommendations to fit the trichotomy of areas within the models presented can justify the negative Kg values of medium probability areas in a CHRM context. For example, the medium probability areas could be recommended for standard survey strategies, while the high and low probability areas could be recommended for more and less intensive survey strategies, respectively. Another option would be that high probability areas could be surveyed using standard survey strategies, medium probability areas could be surveyed using slightly less intensive strategies (like extending shovel test intervals from 15 meters to 30 meters but still investigating the entire area), and low probability areas could be surveyed using the least intensive survey strategies (pedestrian survey with limited but targeted shovel test excavation). As stated by Verhagen, an archaeological predictive model should not only perform well at predicting site locations but also meet the goals of authorities or developers (Verhagen 2008:286). The goal of identifying sites within an area while also reducing costs can be achieved even if the defined medium probability areas yield a negative Kg value.

The highest Kg value for high probability areas was achieved by the distance to stream confluence model. Distance to streamlines has been shown to be one of the strongest predictors of Precontact site location, so it is not surprising that the intersection of multiple streams would also be a good predictor. Streams are natural navigation routes; therefore, it is understandable that the intersection of two routes would have more activity than along either of the individual routes. In other words, the sum is greater than its parts. More expansive floodplains form at the confluences of rivers (Seramur et al. 2007:34), which have been established as favorable locations for sites throughout the Paleoindian, Archaic, and Woodland periods. Rapids also form at stream confluences, which support diverse aquatic habitats that were exploited for their resources (Seramur et al. 2007:44). All of these are theories that may explain why proximity to stream confluences was the best predictor for Precontact site location within the Study Area. The high probability areas within the distance to stream confluence model were the least accurate of the three models, intersecting with 15 of the 23 identified Precontact sites, but they were the most precise, encompassing only 99.9 acres of the Study Area. I was selective in which stream confluences were included in the model, suggesting that the model may have been less precise if I had included every intersection of the flowlines. This seems to support my assumption that the confluences of perennial streams were more attractive locations than those of ephemeral streams.

The next highest Kg value for high probability areas was achieved by the distance to wetland model, but that model performed only slightly better than the distance to stream model.

All the wetlands within the Study Area are situated adjacent to the streams, so it is not surprising that the Kg values for those two models are similar. Theories about why wetlands were attractive locations to Precontact Native Americans typically highlight the diverse flora and fauna they support as well as being sources of water. The high probability areas defined in the distance to wetland and distance to stream models were equally accurate, with both containing 17 of the 23 identified Precontact sites, but the distance to wetland high probability area was more precise, encompassing 26.5 acres less than the distance to stream model's high probability areas. Since the wetlands within the Study Area are all located along the streams and proximity to wetlands proved to be a more precise predictor of site location, it could by hypothesized that distance to stream has a strong correlation to Precontact site location not because of the steams themselves but because of the wetlands located along them. Perhaps model creators could improve the precision of their models by utilizing wetland data instead of streamline data.

Tests of Precontact site location models of the past typically conclude that proximity to streams is one of the strongest variables when attempting to predict the locations of sites, but that was not the result of this test. Surprisingly, the high probability areas within the distance to stream model performed the poorest of the three. The high probability areas within the distance to stream model were equally as accurate (17/23 sites) as those within the distance to wetland model and more accurate than those within the distance to stream confluence model (15/23 sites), but they were also the least precise (202.7 acres). Proximity to streams is clearly a strong indicator of Precontact site location, especially considering that confluences and wetlands are both associated with streams and therefore are often encompassed by a stream buffer, but utilizing that variable may include a larger area than is desired. Streamline data is widely available, can be easily incorporated into a model, and has abundant theoretical support for its use, but the results of this thesis suggest that a specific features along a stream may have been more attractive than a stream in its entirety.

Future Research Directions

APMs created using a deductive approach are not ever-changing like inductive APMs that can be "trained" using newly recorded archaeological sites. That said, archaeological theories are constantly being reevaluated and new theories are continuously being developed. Using those theories to create APMs is one strategy of testing their validity in a real-world situation. As new theories are developed, they should be used to create new models. Also, more specific theories, such as those pertaining to specific temporal periods, could be used to create APMs with more specific research goals.

In a similar vein, archaeological theories focused on different regions could be applied to APMs. The models created in this thesis were created for use within the Virginia Piedmont and therefore may not perform as well if applied to other regions. For example, percent slope is a commonly used variable in areas with a greater variability of topography, but that variable may not be as useful in areas with less topographic variability like the Coastal Plain.

Another direction this research could take is to isolate variables other than hydrologic features to test their predictive power. The APM creation process outlined in this thesis is not excessively complicated or time consuming, meaning it can be used by a wider range of CHRM professionals, not just GIS specialists. The soil data used to create the three models presented in this thesis is a good example. One could compare the predictive power of different soil attributes, such as texture, farmland classification, or frequency of flooding.

It is estimated that over half of the wetlands that existed in the lower 48 states prior to European contact have been lost (USEPA 2001). This presents a problem to the distance to wetland APM because the size and number of wetlands represented in the NWI dataset are likely not an accurate reflection of the wetlands that existed during the Precontact period. The paleolandscape of wetlands in Virginia appears to have been extremely complex and therefore difficult to reconstruct accurately. For example, wetlands are formed or expanded through pooling above beaver dams, and those landscapes would have experienced constant change as new dams were constructed and older ones deteriorated, which is a process that would be exceedingly difficult to reconstruct accurately, especially considering that we do not know how large the beaver population in Virginia was prior to European contact (Pollock et al. 2015:5–6). As our understanding of paleoclimates improves, information from that understanding could be applied to the APM to more accurately represent the extensiveness of wetlands before European contact.

The emergence of Artificial Intelligence (AI) technologies is another factor that could improve upon the practice of creating APMs in the future. In simple terms, AI "is technology that enables computers and machines to simulate human intelligence and problem-solving capabilities" (IBM 2024). AI technologies have already been applied to many areas within the field of archaeology, including the use of software tools as a creative stimulus for exhibition organization, as an analytical tool for artifacts, to assist in the reconstruction of fragmented artifacts and documents, the cataloging and study of human remains, and the detection of terrestrial or submerged archaeological sites (Mantovan and Nanni 2020:1). Three ways I think AI could contribute to my research are: creating environmental datasets that are not already available (such as the stream confluence data points I had to create), summarizing vast bodies of archaeological theory to assist in the selection variables to include in a deductive approach APM, and reconstructing paleolandscapes using geospatial data along with information from our understanding of paleoclimates. AI technologies are becoming increasingly useful in a wide range of fields and the benefits they present should be applied to archaeological research.

Methodological Adjustments

There are methodological changes that could contribute to a better performing APM or more reliable tests of the APMs. The APMs could be applied to areas that were investigated in a uniform way. In other words, to determine the site and non-site areas more accurately, the Study Area should be investigated using the same methodology on all portions of the area, like shovel tests being excavated at the same intervals across the entire area. Finding areas that have been investigated in this manor is difficult because many states, including Virginia, allow for the exclusion of shovel testing in certain areas based on environmental conditions, like areas of greater than 15% slope or areas inundated with water. This would be best achieved by conducting investigations that are specifically intended to test the APMs. Since most of the available site data is the result of investigations conducted in CHRM contexts, it can generally be assumed that the areas that are allowed to be skipped were skipped because CHRM projects are awarded based on a competitive bidding process.

The methodology of the performance test could be improved by using shovel test data to assess the accuracy of the different probability areas. Instead of using site area polygons to determine which probability areas intersect with sites, one could use shovel test points and calculate how many of the shovel tests that contained artifacts were situated within the different probability areas. Archaeological site polygons are not all created using the same methodology. For example, the *Guidelines for Conducting Historic Resources Survey in Virginia* define what an archaeological site is, but do not give a singular method for determining a site's boundary (VDHR 2017:40–41). This is good practice because archaeological sites are not monoliths; their boundaries are defined in a wide variety of ways. That said, this is why using strictly dichotomic shovel test points, recorded as either positive or negative, could be seen as a more accurate and statistically appropriate representation of site location.

Final Conclusions

The primary focus of this thesis was to determine if using hydrologic features other than streams, specifically stream confluences and wetlands, to express the distance to water variable would improve the performance of an APM. I demonstrated that, yes, other hydrologic features may be better predictors of Precontact site locations in the Virginia Piedmont. Both the distance to stream confluence and the distance to wetland APMs outperformed the distance to stream APM when tested using Kvamme's gain.

Secondarily, I hoped to show that an APM created using a deductive approach, where variables were selected based on archaeological theory and tested using previously recorded site location data, would perform well enough to be considered appropriate for use in CHRM contexts. The high probability areas of all three of the APMs I created yielded Kg values high enough to be considered as having predictive utility (Kvamme 1988:329). This demonstrates that all three of the APMs I created could be considered appropriate to use to guide survey efforts in a CHRM context.

Bibliography

Ahlman, Todd M., Gail L. Guymon, and Nicholas P. Herrman
2005 *Archaeological Overview and Assessment of the Cumberland Gap National Historic Park, Kentucky, Tennessee, and Virginia*. Prepared for the National Park Service, Cumberland Gap National Historic Park. Archaeological Research Laboratory, Department of Anthropology, the University of Tennessee, Knoxville.

Altschul, Jeffrey H.
1988 Models and the Modeling Process. *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, U.S. Department of the Interior, Bureau of Land Management Service Center, Denver, pp. 61–96.

Anderson, David G.
2001 Climate and Culture Change in Prehistoric and Early Historic Eastern North America. *Archaeology of Eastern North America,* Vol. 29, pp. 143–186.

Anderson, David G., Lisa D. O'Steen, and Kenneth E. Sassaman
1996 Environmental and Chronological Consideration. In *The Paleoindian and Early Archaic Southeast*, edited by David G. Anderson and Kenneth E. Sassaman, pp. 1–15. University of Alabama Press, Tuscaloosa.

Anderson, David G., and Glen T. Hanson
1988 Early Archaic Settlement in the Southeastern United States: A Case Study from the Savannah River Valley. *American Antiquity,* Vol. 53, No. 2, pp. 262–286.

Barber, Michael B.
2010 Settlement Pattern Modeling in Virginia's Blue Ridge: A View from Arnold, Rockbridge and Botetourt Counties, Virginia. *Quarterly Bulletin of the Archaeological Society of Virginia* 65(3):103–118.

Barfield, Eugene B., and Michael B. Barber
1992 Archaeological and Ethnographic Evidence of Subsistence in Virginia During the Late Woodland Period. In *Middle and Late Woodland Research in Virginia: A Synthesis,* edited by Theodore R. Reinhart and Mary Ellen N. Hodges. Council of Virginia Archaeologists. Special Publication No. 29 of the Archaeological Society of Virginia, Richmond. pp. 225–248.

Boyd, Cliff
2020 A Review of Paleoindian Research in Virginia and a View Towards the Future. In *The Archaeology of Virginia's First Peoples*, edited by Elizabeth A. Moore and Bernard K. Means, Ch. 3, pp. 31–42. The Archaeology Society of Virginia. Layout and Production by Jessica Davenport. Independently published, Richmond.

Carter, J.R., J.W. Willis, and W.E. Cummins
1971 *Soil Survey of Orange County, Virginia*. U.S. Department of Agriculture, Soil Conservation Service, in Cooperation with the Virginia Agricultural Experiment Station, U.S. Government Printing Office, Washington, D.C.

Church, Tim, R. Joe Brandon, and G. R. Burgett
2000 GIS Applications in Archaeology: Method in Search of Theory. *Practical Application of GIS for Archaeologists: A Predictive Modeling Kit,* Taylor and Francis, London, Ch. 9, pp. 135–155.

Custer, Jay F.
1980 Settlement-Subsistence Systems in Augusta County, Virginia. *Quarterly Bulletin of the Archaeological Society of Virginia* 35:1–27.

Ebert, David
2004 Applications of Archaeological GIS. *Canadian Journal of Archaeology,* Vol. 28, No. 2, pp. 319–341.

Ebert, James I., and Timothy A. Kohler
1988 The Theoretical Basis of Archaeological Predictive Modeling and a Consideration of Appropriate Data-Collection Methods. *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, U.S. Department of the Interior, Bureau of Land Management Service Center, Denver, pp. 97–172.

Egghart, Christopher
2020 State Plan and Research Design Middle Archaic (6500 BC–2500 BC). In *The Archaeology of Virginia's First Peoples*, edited by Elizabeth A. Moore and Bernard K. Means, Ch. 5, pp. 53–70. The Archaeology Society of Virginia. Layout and Production by Jessica Davenport. Independently published, Richmond.

Egghart, Christopher, and C. Neil Manson
2016 Prehistoric Settlement along the Nottoway River Fall Zone: Analysis of the Robert Ogle Projectile Point Collection. *Archaeological Society of Virginia, Quarterly Bulletin,* 71(2):55–92, Richmond.

Egghart, Christopher, Cliff Boyd, Michael B. Barber, Carole Nash, Bernard K. Means, Elizabeth A. Moore, Christopher J. Shephard, Martin D. Gallivan, and Keith Egloff
2020 *The Archaeology of Virginia's First Peoples*, The Archaeology Society of Virginia. Layout and Production by Jessica Davenport. Independently published, Richmond.

Environmental Systems Research Institute, Inc. (ESRI)
2024 Raster. In *support.esri.com GIS Dictionary.* https://support.esri.com/en-us/gis-dictionary/raster. Accessed January 2024.

Fiedel, Stuart J.
2000 The Peopling of the New World: Present Evidence, New Theories, and Future Directions. *Journal of Archaeological Science* 8:39–103.

Gallivan, Martin D.
2003 *James River Chiefdoms: The Rise of Social Inequality in the Chesapeake*, University of Nebraska Press, Lincoln.

Ganopolski, Andrey, Claudia Kubatzki, Martin Claussen, Victor Brovkin, and Vladimir Petoukhov
1998 The Influence of Vegetation- Atmosphere-Ocean Interaction on Climate During the Mid-Holocene. *Science* 280(5371):1916-1919.

Gardner, William M.
1974 The Flint Run Paleo-Indian Complex: A Preliminary Report 1971–73 Seasons (editor). Catholic University of America, Department of Anthropology, Archaeology Laboratory, *Occasional Publication* 1, Washington, D. C.

1977 Flint Run Paleo-Indian Complex and its Implication for Eastern North American Prehistory. In *Amerinds and Their Paleoenvironments in Northeastern North America,* edited by Walter S. Newman and Bert Salwen. Annals of the New York Academy of Sciences 288, pp. 257–263.

Godfrey, Michael A.
1997 *Field Guide to the Piedmont: The Natural Habitats of America's Most Lived-in Region, From New York City to Montgomery, Alabama.* The University of North Carolina Press, Chapel Hill and London.

Green, Ernestine L.
1973 Location Analysis of Prehistoric Maya Sites in Northern British Honduras. *American Antiquity* Vol. 38, pp. 279–293.

Harris, Matthew D., Robert G. Kingsley, and Andrew R. Sewell
2015 *Archaeological Predictive Model Set*. Prepared for the Commonwealth of Pennsylvania Department of Transportation, Bureau of Planning and Research. URS Corporation, Burlington.

Hay, Conran A., Alan N. Snavely, Thomas E. Scheitlin, Catherine E. Bollinger, and Thomas O. Maher
1982 Archaeological Predictive Models: A New Hannover County Test Case. *North Carolina Archaeological Council Publication Number 18*, North Carolina Archaeological Council and the Archaeology Branch Division of Archives and History North Carolina Department of Cultural Resources, Raleigh.

Haynes, C. Vance, Jr.
2000 New World Climate: Dramatic Climate Shifts Welcomed and Bedeviled the First Americans. In *Scientific American Discovering Archaeology*, Vol. 2, No. 1, pp. 37–39.

Hodges, Mary Ellen N.
1991 Late Archaic and Early Woodland Periods in Virginia: Interpretation and Explanation within an Eastern Context. In *Late Archaic and Early Woodland Research in Virginia: A Synthesis,* edited by Theodore R. Reinhart and Mary Ellen N. Hodges. Council of Virginia Archaeologists. Special Publication No. 23 of the Archaeological Society of Virginia, Richmond. pp. 221–242.

Hoffman, Michael A., and Robert W. Foss
1980 *Man in the Blue Ridge – An Archaeological and Environmental Perspective*. Paper Presented at the 1977 Annual Meeting of the Society for American Archaeology, New Orleans (Ms. on file at the Laboratory of Archaeology, University of Virginia, Charlottesville).

Holland, C. G.
1978 Albemarle County Settlement: A Piedmont Model. *Quarterly Bulletin of the Archaeological Society of Virginia*, 38:1–42.

International Business Machines (IBM)
2024 What is Artificial Intelligence (AI)? In *Think: Tech News, Education and Events.* IBM.com. Online document, https://www.ibm.com/topics/artificial-intelligence.

Joseph, J. W., Laura Kate Schnitzer, and Scot J. Keith
2018 *Phase I Archaeological Survey of the Pierpont Wetland Mitigation Tract, Bulloch County, Georgia*. New South Associates, Stone Mountain, Georgia. Submitted to the Georgia Department of Transportation, Atlanta, Georgia.

Kavanaugh, Maureen
1983 Prehistoric Occupation of the Monocacy River Region, Maryland. In *Piedmont Archaeology,* edited by J. Mark Wittkofski and Lyle E. Browning, pp. 40–45. Special Publication No. 10 of the Archaeological Society of Virginia, Richmond.

Kincaid, Chris
1988 Predictive Modeling and its Relationship to Cultural Resource Management Applications. *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, U.S. Department of the Interior, Bureau of Land Management Service Center, Denver, pp. 549–570.

Klein, Michael J., and Thomas Klatka

1991 Late Archaic and Early Woodland Demography and Settlement Patterns. In *Late Archaic and Early Woodland Research in Virginia: A Synthesis,* edited by Theodore R. Reinhart and Mary Ellen N. Hodges. Council of Virginia Archaeologists. Special Publication No. 23 of the Archaeological Society of Virginia, Richmond. pp. 139–184.

Klein, Mike, Emily Calhoun, Marco González, and Earl E. Proper

2015 *Archaeological Background Review and Predictive Model for the Washington, D.C. to Richmond, Virginia, Southeast High Speed Rail Corridor*. Prepared for the Virginia Department of Rail and Public Transportation. DC2RVA Project Team, Richmond.

Kohler, Timothy A.

1988 Predictive Locational Modeling: History and Current Practice. *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, U.S. Department of the Interior, Bureau of Land Management Service Center, Denver, pp. 19–60.

Kohler, Timothy A., and Sandra C. Parker

1986 Predictive Models for Archaeological Resource Location. *Advances in Archaeological Method and Theory,* Vol. 9, pp. 397–452.

Kvamme, Kenneth L.

1988 Development and Testing of Quantitative Models. *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, U.S. Department of the Interior, Bureau of Land Management Service Center, Denver, pp. 325–428.

Madry, S., M. Cole, K. Kvamme, S. Seibel, S. Gould, and J. Broush

2006 *North Carolina Archaeological Predictive Modeling Project: Results of Task 2, Cabarrus, Chatham, Forsyth, Granville, Guilford, Randolph, and Wake Counties*. Environmental Services, Inc. Raleigh. Prepared for North Carolina Department of Transportation, Raleigh.

Mandel, Rolfe D., and E. Arthur Bettis, III

2001 Use and Analysis of Soils by Archaeologists and Geoscientists: A North American Perspective. *Earth Sciences and Archaeology,* edited by P. Goldberg, V. T. Holliday, and C. R. Ferring, pp. 173–204. Kluwer Academic/Plenum Publishers, New York.

Mantovan, Lorenzo, and Loris Nanni

2020 The Computerization of Archaeology: Survey on Artificial Intelligence Techniques. *SN COMPUTER. SCIENCE* 1, 267. Available online, https://doi.org/10.1007/s42979-020-00286-w.

Martin, Paul S.
1984 Prehistoric Overkill: The Global Model. In *Quaternary Extinctions: A Prehistoric Revolution,* edited by Paul S. Martin and Richard G. Klein, pp. 354-403. University of Arizona Press, Tucson.

Martin, Paul S., and Richard G. Klain (editors)
1984 *Quaternary Extinctions: A Prehistoric Revolution*. University of Arizona Press, Tucson.

Means, Bernard K.
2020 Late Woodland Archaeology of Northern Virginia and Adjacent Regions. In *The Archaeology of Virginia's First Peoples*, edited by Elizabeth A. Moore and Bernard K. Means, Ch. 9, pp. 161–184. The Archaeology Society of Virginia. Layout and Production by Jessica Davenport. Independently published, Richmond.

Millis, Heather, and Philip Hayden.
2018 *Phase I Cultural Resources Survey for the Sol Madison Solar Project, Orange County, Virginia*. TRC Environmental Corporation, Chapel Hill, North Carolina. Submitted to Cypress Creek Renewables, LLC, Asheville, North Carolina.

Monaghan, G. William, Daniel R. Haynes, S. I. Dworkin, and Eric Voigt
2004 Geoarchaeology of the Brook Run Site (44CU122): An Early Archaic Jasper Quarrt in Virginia, U.S.A. *Journal of Archaeological Science* 31 (8): 1083–1092.

Moon, Heather
1993 *Archaeological Predictive Modeling: An Assessment.* Prepared for the Earth Science Task Force Resources Inventory Committee, Archaeology Branch of the Ministry of Tourism and Ministry Responsible for Culture, Victoria, BC.

Nash, Carole L.
2009 *Modeling Uplands: Landscape and Prehistoric Native Settlement Archaeology in the Virginia Blue Ridge Foothills*. Ph. D. dissertation, Department of Anthropology, Catholic University, Washington D.C.

2020 Middle Woodland Research in Virginia: A Review of Post-1990 Studies. In *The Archaeology of Virginia's First Peoples*, edited by Elizabeth A. Moore and Bernard K. Means, Ch. 8, pp. 123–160. The Archaeology Society of Virginia. Layout and Production by Jessica Davenport. Independently published, Richmond.

National Historic Preservation Act
1966 36 CFR Part 800 See Title 36, Part 800 of the Code of Federal Regulations (36 CFR Part 800). Available online. https://www.achp.gov/sites/default/files/regulations/2017-02/regs-rev04.pdf.

Nebbia, Marco

2023 Predictive Modeling in Archaeology. *Encyclopedia of Archaeology,* edited by Efthymia Nakita and Thilo Rehren. Second edition, Vol. 2A: Fieldwork, pp. 222–230. Elsevier, Inc., Amsterdam.

North Carolina Department of Transportation

2017 *Terrestrial Archaeological Resources Predictive Model for Administrative Action State Environmental Impact Statement Kinston Bypass, Lenoir, Jones, and Craven Counties, North Carolina*. Revised October 2017. North Carolina Department of Transportation, Raleigh.

North Carolina Office of State Archaeology

2023 *Archaeological Investigation Standards and Guidelines for Background Research, Field Methodologies, Technical Reports, and Curation*. Revised November 2023. North Carolina Department of Natural and Cultural Resources, Raleigh.

Parker, Scott K.

1990 Early and Middle Archaic Settlement Patterns and Demography. In *Early and Middle Archaic Research in Virginia: A Synthesis,* edited by Theodore R. Reinhart and Mary Ellen N. Hodges. Council of Virginia Archaeologists. Special Publication No. 22 of the Archaeological Society of Virginia, Richmond. pp. 99–117.

Parsons, Jeffrey R.

1972 Archaeological Settlement Patterns. *Annual Review of Anthropology,* Vol. 1, pp. 127–150.

Pollock, M.M., G. Lewallen, K. Woodruff, C.E. Jordan and J.M. Castro (Editors)

2015 *The Beaver Restoration Guidebook: Working with Beaver to Restore Steams, Wetlands, and Floodplains*. Version 1.02. United States Fish and Wildlife Service, Portland, Oregon. 189 pp. Available online, http://www.fws.gov/oregonfwo/ToolsForLandowners/RiverScience/Beaver.asp. Accessed July 2024.

Roberts, C., and C.M. Bailey (Cartographers)

2000 Physiographic Map of Virginia Counties [Map]. College of William & Mary. Retrieved from https://training.fws.gov/courses/CSP/CSP3200/resources/documents/Physiographic_Map_of_Virginia.pdf. Accessed December 2022.

Seramur, Kieth C., Ellen A. Cowan, Loretta Lautzenheiser, and Jane M. Eastman

2007 A Model of Distribution and Preservation of Archaeological Sites Along Piedmont Stream: The Deep River, North Carolina. *Southeastern Archaeology,* Vol. 26, No. 1, pp. 32–46. Taylor & Francis, Ltd.

Stewart, R. Michael

1992 Observations on the Middle Woodland Period of Virginia, A Middle Atlantic Regional Perspective. In *Middle and Late Woodland Research in Virginia: A Synthesis,* edited by Theodore R. Reinhart and Mary Ellen N. Hodges. Council of Virginia Archaeologists. Special Publication No. 29 of the Archaeological Society of Virginia, Richmond. pp. 1–38.

1995 The Status of Woodland Prehistory in the Middle Atlantic Region. *Archaeology of Eastern North America,* Vol. 23, pp. 177–206.

Turner, E. Randolph, III

1994 Paleoindian Settlement Patterns and Population Distribution in Virginia. In *Paleoindian Research in Virginia: A Synthesis,* edited by J. Mark Wittkofski and Theodore R. Reinhart. Council of Virginia Archaeologists. Special Publication No. 19 of the Archaeological Society of Virginia, Richmond. pp. 71–94.

United States Department of Agriculture, Natural Resources Conservation Service (USDA NRCS)

2024 Soil Survey of Orange County, Virginia. Electronic Document, https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx. Accessed January 2024.

United States Environmental Protection Agency (USEPA)

2001 *Threats to Wetlands.* Electronic Document, https://www.epa.gov/sites/default/files/2021-01/documents/threats_to_wetlands.pdf. Accessed July 2024.

2024 *What is a Wetland?* Electronic Document, https://www.epa.gov/wetlands/what-wetland. Accessed July 2024.

United States Geological Survey (USGS)

2020 USGS one meter x29y413 VA Sandy 2014. 3D Elevation Program 1-Meter Resolution Digital Elevation Model. ScienceBase-Catalog, https://www.sciencebase.gov/catalog/item/5eace9a482cefae35a2497a1. Accessed January 2024.

United States Geological Survey, National Hydrography Dataset (USGS NHD)

2022 Virginia National Hydrography Dataset. Virginia Geographic Information Network. Geodatabase, https://vgin.vdem.virginia.gov/documents/5e929f8d9db44dadbe545f0ccbfc7fa7/about. Accessed January 2024.

Verhagen, P.

2008 Testing Archaeological Predictive Models: A Rough Guide. In *Layers of Perception*, edited by Posluschny, A., K. Lambers, and I. Herzog. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA). April 2–6, 2007, Berlin, Germany.

Verhagen, Philip and Thomas G. Whitley
2012 Integrating Archaeological Theory and Predictive Modeling: A Live Report from the Scene. *Journal of Archaeological Method and Theory,* Vol. 19, No. 1, pp. 49–100.

Virginia Department of Conservation and Recreation, Division of Natural Heritage (VDCR DNH)
2021 *Overview of the Physiography and Vegetation of Virginia.* Electronic Document, https://www.dcr.virginia.gov/natural-heritage/natural-communities/document/ncoverviewphys-veg.pdf. Accessed December 2022.

Virginia Department of Historic Resources (VDHR)
2017 Conducting Archaeological Investigation. *Guidelines for Conducting Historic Resources Survey in Virginia,* Ch. 6, pp. 40–56.

2022 Virginia Cultural Resource Information System (V-CRIS). Online Database, https://vcris.dhr.virginia.gov/VCRIS/. Accessed December 2022.

Voigt, Eric
2001 *Preliminary Results of the Excavations at the Brook Run Quarry*. Paper presented at the annual meeting of the Archaeological Society of Virginia, Williamsburg, October 19–21.

Voigt, Eric, Jeffrey Flenniken, and J.T. Rasic
2004 *Archaeological Data Recovery the Brook Run Jasper Quarry (Site 44CU0122) Associated with the Proposed Route 3 Improvement*. The Louis Berger Group, Richmond, Virginia.

Webb, Harry W., and Palmer C. Sweet
1992 Interesting Uses of Stone in Virginia – Part I. *Virginia Minerals,* Vol. 38, No. 4. Division of Mineral Resources, Charlottesville.

Wescott, Konnie L.
2000 Introduction. *Practical Applications of GIS for Archaeologists: A Predictive Modeling Kit,* Taylor and Francis, London, Ch. 1, pp. 1–4.

Wheatley, T. G.
2004a *Making Space for an Archaeology of Place.* Internet Archaeology 15. https://-intarch.ac.uk/journal/issue15/wheatley_index.html. Accessed June 2024.

2004b Causality and Cross-purposes in Archaeological Predictive Modeling. In *Enter the Past: The E-way into the Four Dimensions of Cultural Heritage*, edited by A. Fischer Ausserer, W. Börner, M. Goriany, and L. Karlhuber-Vöckl. Computer Applications and Quantitative Methods in Archaeology 2003. BER International Series 1227 (Oxford 2004), pp. 236–239.

Woods, Alan J., and James M. Omernik
1999 *Level III and IV Ecoregions of Delaware, Maryland, Pennsylvania, Virginia, and West Virginia.* Unites States Environmental Protection Agency (USEPA), Corvallis.

Woods, Alan J., James M. Omernik, and Douglas D. Brown
2003 Level III and IV Ecoregions of EPA Region 3 [Map]. Map preparation and development of digital files provided by Jeffrey A. Comstock, Sandra H. Azevedo, M. Frances Faure, and Suzanne M. Pierson (OAO Corp.). Scale 1:500,000. U.S. Environmental Protection Agency, Washington, D.C.