PENNSYLVANIA DEPARTMENT OF TRANSPORTATION
ARCHAEOLOGICAL PREDICTIVE MODEL SET

TASK 1: LITERATURE REVIEW

CONTRACT #355I01

ARCHAEOLOGICAL PREDICTIVE MODEL SET

Category #05 – Environmental Research

October 2013

URS

PENNSYLVANIA DEPARTMENT OF TRANSPORTATION ARCHAEOLOGICAL PREDICTIVE MODEL SET TASK 1: LITERATURE REVIEW

CONTRACT #355I01

Prepared for
Pennsylvania Department of Transportation
Bureau of Planning and Research
Keystone Building
400 North Street, 6th Floor, J-East
Harrisburg, PA 17120-0064

Prepared by
Matthew D. Harris, Principal Investigator

URS Corporation
437 High Street
Burlington, NJ 08016-4514

October 2013

ABSTRACT

This report is the documentation for Task 1 of the Statewide Archaeological Predictive Model Set project sponsored by the Pennsylvania Department of Transportation (PennDOT). This project was solicited under Contract #355I01, Transportation Research, Education, and Technology Transfer ITQ, Category #05 – Environmental Research. The goal of this project is to develop a set of statewide predictive models to assist the planning of transportation projects. PennDOT is developing tools to streamline individual projects and facilitate Linking Planning and NEPA, a federal initiative requiring that NEPA activities be integrated into the planning phases for transportation projects. The purpose of Linking Planning and NEPA is to enhance the ability of planners to predict project schedules and budgets by providing better environmental and cultural resources data and analyses. To that end, PennDOT is sponsoring research to develop a statewide set of predictive models for archaeological resources to help project planners more accurately estimate the need for archaeological studies.

The objective of Task 1 is to review literature from Pennsylvania and the Eastern Woodlands pertinent to the practice of archaeological predictive modeling (APM). Based on this review, examples of successful and less than successful modeling methods are evaluated, and the findings synthesized. Within the context of the statewide predictive model project, this task will utilize past studies to help determine best practices and avoid pitfalls.

The foundational references for this task were drawn from the Pennsylvania Historical and Museum Commission's (PHMC) Environmental Review (ER) archives, URS's extensive library of CRM reports and research documentation, and other repositories of similar research. From the ER files of the PHMC, a total of 47 archaeological reports was identified using key words such as "predictive," "model," and "modeling," as well as reports with "Predictive Model" as the report type. These reports were scanned and reviewed for content. From these, 32 reports that contained formalized predictive models were evaluated for this synthesis. Nine of these reports exhibited creative methodologies or were otherwise seen as making significant contributions to the study of APM. These nine reports were evaluated and assessed in great detail to explore their methods and findings.

The synthesis of these reports led to the creation of a modeling methods typology. This report explores the characteristics and assumptions of each model type, the efficiency and performance of each model type, and the range and efficacy of environmental variables. Guidelines for when each type of model is most effective are also presented. Ultimately, these reports show that each physical environment and data set call for a modeling approach that is tailored to that situation and that the results of each model are comparable only given the situation and the goals of the model itself.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This report details a literature review focused on the published sources that have defined the methods of Archaeological Predictive Modeling (APM) and the test-cases where applied to the Eastern Woodlands, with a focus on Pennsylvania. This research not only considers the application of models within the Eastern Woodland, but also the key methodological developments from around the world that influence the practical applications. Through the inclusion of methodology as well as application a broader understanding of the advantages and drawbacks of each type of modeling is gained. Contained in this review is a brief history of methodological development, a synopsis and assessment of models applied to the region, an evaluation of the utility of each model and the environmental variables used, a synthesis of the evaluation results, and finally recommendations on how these types of models may successfully be implemented. The sources drawn from in this research include URS's library of computational archaeology literature, online research databases, academic journals, and the holdings of the Pennsylvania Historical and Museums Commission (PHMC).

This task is the first step in the creation of a Pennsylvania state-wide set of predictive models. The contribution of this task to the overall project is to allow history to inform the process. Modeling is by its nature an exercise in trial and error, and by studying the history of modeling in the region, hopefully we may learn from the successes and avoid the pitfalls. The understanding gained from these evaluations will influence the types of models that are created for different regions, the variables that are utilized in our attempt to identify a pattern in prehistoric site locations, and the standards of performance by which we can measure our achievements. Ultimately, the reports included in this summary provide a context for the understanding of how the real-world application of these abstracted models can accomplish the reciprocal goals of protecting cultural resources and aiding in the effective and efficient completion of transportation projects.

SCOPE OF THIS PROJECT

The scope of this task is to evaluate APM from Pennsylvania and the Eastern Woodlands region. The purpose of this evaluation is to gain a better understanding of the methods that have been used and which approaches have proven the most successful. In order to achieve this, this project drew from the Environmental Report (ER) archives of the PHMC and URS's in-house library of Cultural Resource Management (CRM) studies and research literature.

The report that follows is organized into chapters: Chapter 1 contains project specific information and background information on APM. Chapter 2 contains a description of the methods by which models were selected and a more lengthy description of the technical and methodological approaches utilized within the APM studies reviewed herein. Chapter 3 begins with a description of the APM study data set and a technical description of how each model is evaluated, followed by the individual evaluations. Each individual project evaluation contains information on the project's region of application, the significance of the model to this study, the model type, the variables used in the model, an evaluation of the model's performance, and a final assessment that determines if the model was successful in achieving its goals and what contributions it made to the overall project. Chapter 4 consists of the results and findings of the evaluation. This chapter is a synthesis of the reporting style, variables used, and methods observed within the APM reports. Chapter 5 offers concluding observations and general recommendations regarding how and when each type of model and certain variables could be applied with success. Finally, Chapter 6 is the references cited section.

BRIEF HISTORY OF APM

In order to evaluate the success of various modeling approaches, it is important to understand the origins of APM, the methods employed throughout time, and the benefits and drawbacks associated with them. The historical development of this field sheds light on which techniques have withstood the test of time and what APM is able to tell us about the reality of archaeological site locations and the systems that created them.

While many authors cite Willey's (1953) settlement pattern analysis in Peru as the tap root of today's APM, it was not until the 1960s that the use of the term "predictive" made headlines in the archaeological debate of new methods and systems thinking (Bayard 1969). With the methodological and computer-aided focus growing in the New Archaeology paradigm, the use of the term "predictive model" became more commonplace in archaeological literature in the 1970s (e.g., Judge 1973; Engelbrecht 1974; Smith 1974; Jochim 1976). Throughout the 1980s, the methods of numerous modeling techniques were developed and applied to archaeology. The seminal papers, including Kvamme (1983, 1984, 1988), Kohler and Parker (1986), and Judge and Sebastian (1988), each contributed to the elevation of a generalized modeling methodology that used measurements of the environment taken at the location of known archaeological sites to develop a pattern that was projected into unsurveyed regions to find landforms of similar measures.

The popularity of this method stemmed from its intuitiveness to archaeologists, accessible statistical methods, reliance on a body of identified sites, and suitability to be projected across large tracts of land. With this emphasis on projecting from the level of sites to the region and reliance on statistical models (mostly linear or logistic regression), this approach was referred to as an "inductive" approach. Developing concurrently with the adoption of these methods, alternate approaches were created in an attempt to address drawbacks of the inductive approach, namely being theory-neutral, often using biased samples, lacking in explanatory power, and seen as environmentally deterministic (Borillo 1974; Salmon 1976; Bettinger 1980). Alternate models often relied more on the use of theory and hypothesis to construct models of where sites *should* be located based on knowledge, as opposed to where sites *could* be located based on existing samples; this was called a "deductive" approach. The tension between these two approaches, inductive vs. deductive, has defined the archaeological literature of APM for almost 30 years.

Throughout the 1990s and 2000s these two approaches were debated heavily, but APM in general fell from the mainstream archaeological literature, with the exception of grand simulation models championed by Kohler (Kohler and Gumerman 2000). Throughout that time, computer technology made statistics-based models more accessible and easy to create, but no more accurate. Many deductive approaches were offered with mixed success though none gained widespread use. It is only in the most recent generation of APM literature that researchers have gone beyond the dichotomy of inductive vs. deductive, realizing that it is a false argument (Whitley 2005), and have begun to accept that every model is some combination of both approaches and that each orientation has valuable applications in specific realms such as management and research (Verhagen and Whitley 2012). Understanding the theoretical orientation of a specific model is vital in using it correctly. Without knowledge of the model's focus, blind spots, and intention, its use as a management device is flawed.

The models reviewed in this study mirror the broader developmental trends of the field within the United States and abroad. The studies selected for evaluation here contain examples of methodological and theoretical approaches that were clearly informed by the national debate. Even within some studies, the tension between utilizing archaeological experience and theory versus statistical inference and computer technology is very evident. This false dichotomy of experience versus technology—along with those of inductive versus deductive, hypothesis versus empirical observation, research versus management, academic versus CRM, etc.—builds the scaffolding that frames the use of APM methods, but does little to help advance the field.

While any given model may be more slanted toward a specific theoretical orientation, they are all a combination of many approaches and cannot be assigned to specific camps. To do so serves only to ignore useful methods and theory that may be the best fit for the data; it is the data that should inform the selection of the model and not the other way around. The report evaluations within this study are undertaken within the historical context of the development of the APM field. No specific theoretical approach or camp is adhered to or viewed as the best approach. It is hoped that this research can contribute to the growth of APM models that are less restricted by theoretical restraints and more open to achieving successful results.

2. CHARACTERISTICS OF APM MODELS IN PENNSYLVANIA

Numerous archaeological predictive models from throughout the Commonwealth of Pennsylvania have been evaluated to better understand the history, methodology, and success of various techniques. The APM models evaluated here cover the full range of time that archaeologists have been practicing APM building within the context of cultural resources management, from the early 1980s to the present. Further, the methods used to construct the models evaluated here showcase a rather wide range of techniques and applications. The following section will describe the methods, model variables, and outcomes of each of 32 models evaluated within the Commonwealth.

MODEL SELECTION

The selection of models for this evaluation was based primarily on a keyword search of the Pennsylvania SHPO's Environmental Review (ER) files database. A total of 47 reports were obtained from this archive and serve as the overall database for this evaluation. Each of these reports was scanned in its entirety and evaluated to gain an understanding of the settlement analysis or model that was undertaken. These reports were identified by searching the ER report database for archaeological projects listed as "Predictive Model" as a report type, and reports that contained the term "Predictive" in the title. The reports generated through this search method are judged to constitute a relatively complete and representative sample of what is available in this archive.

From these reports, a number of studies were culled because they were contextual or synthetic documents, as opposed to predictive models (Table 1). From the APM reports, 32 were chosen because they contained predictive models and a means to evaluate their results. The models chosen for inclusion in this study are those that have served as reference points for other APM studies, those that advanced the methodological and interpretive development of APM studies in the region, and those that presented cogent technical details. This includes models that are large and small in aerial scope, those that succeeded and some that failed, both GIS and non-GIS based models, and those that were innovative as well as models that used tried-and-true methods. Further, the models selected for evaluation are those that resulted in assessment of sensitivity or probability of archaeological resources for specific and often continuous areas. This is to differentiate models that resulted in broad and nonspecific assessments of sensitivity for idealized landforms or regions. The selection of models that resulted in continuous surfaces, GIS or otherwise, was done to draw a division between more specific APMs and less specific settlement pattern analyses or landscape sensitivity studies. The former, if well documented, allow for the verification of results and methods, while the latter are frequently hypothetical in nature and not as comparable other APMs and the models created through this project.

Table 1 - Reports Excluded from Study

Author(s)	Date	Report Title	Reason for Exclusion
Corrie	1984	Predictive Archaeological Model Study, Third Street to Ferry Street Redevelopment Parcel, Easton, Northampton County, Pennsylvania. E.R. 1984-1641-095-B	Study of historical archaeological site locations.
Davis	1989	Archaeological Land Use History of the Pittsburgh Technology Center Site, Pittsburgh, Pennsylvania. E.R.1989-1053-003-A	This report is a land use history for the survey for historical industrial sites.
Vento	1994	Volume IA, Genetic Stratigraphy: The Model for Site Burial and Alluvial Sequences in Pennsylvania. E.R. 1994-R001-042-A	A detailed geomorphological study.
Heberling and Associates	1995	Phase 1 Archeological Investigation, West Fairview Borough park, Cumberland County, Pennsylvania. E.R 1985-1323-041-B	Phase I survey, no discernible archaeological predictive model.
GAI Consultants	1998	Abbreviated Technical Report Phase I Cultural Resources Survey Proposed Knowledge Parkway Project Harborcreek Township, Erie County, Pennsylvania. E.R. 1992-0329-018	Phase I survey, no discernible archaeological predictive model.
VandenBosch, Siemon, and Johnson	2000	Phase 1 Archaeological Survey of the Proposed East Side Access Highway, wintergreen George Bridge Project Area. SR4034-A91, Harborcreek Township, Erie County, Pennsylvania. E.R. 1992-0858-049-F	Phase I survey, no discernible archaeological predictive model.
Chiarulli, Kellogg, Kingsley, Meyer, Miller, Perazio, and Siegel	2001	Prehistoric Settlement Patterns in Upland Settings: An Analysis of Site Data in a Sample of Exempted Watersheds. E.R 2001-R001-042-A	A detailed prehistoric data synthesis and settlement pattern analysis for upland settings, but no discernible archaeological predictive model.
Weed	2002	Prehistoric Context Study (Chaper Three) in Support of Data Recovery at Site 36AL480, Leetsdale, Allegheny County, Pennsylvania. E.R. 1999-2661-003-T.	A prehistoric context and settlement pattern analysis, but no discernible archaeological predictive model.
Lawrence, Weinberg, and Hayes	2003	Alternative Mitigation to the Interstate Fairgrounds Site (36BR210). S.R. 1056, Section 001, Athens Bridge Replacement Project, Athens Township, Bradford County, Pennsylvania. E.R. 2000-8029-015-R.	Synthesis and unpublished reports, but no discernible archaeological predictive model.

Author(s)	Date	Report Title	Reason for Exclusion
MacDonald, Lothrop, and Cremeens	2003	Pennsylvania Archaeological Data Synthesis: The Raccoon Creek Watershed (Watersheds D, Subbasin 20) Bridge Replacement Project T-319, Beaver County Bridge No. 36 (Link Bridge), Independence Township, Beaver County, Pennsylvania. E.R. 1996-8232-007-G.	A very detailed prehistoric data synthesis at the watershed scale, but no discernible archaeological predictive model.
MacDonald, Mahoney, and Dugas	2003	Pennsylvania Archaeological Data Synthesis: The Upper Juniata River Sub-Basin 11 (Watersheds A-D) Walter Industrial Park: Mitigation of Adverse Effects, Grennfield Township, Blair County, Pennsylvania. E.R. 2000-2888-013-P.	A very detailed prehistoric data synthesis at the sub-basin scale, but no discernible archaeological predictive model.
Diamanti	2006	Addendum to Phase l a Sampling Design for Urban Archaeological Resources Mon/Fayette Transportation Project S.R. 51 TO I-376 Section Allegheny County, Pennsylvania. E.R. 1987-1002-042-B86	Sampling strategy for historic urban sites.
MacDonald	2006	Pennsylvania Archaeological Data Synthesis Subbasin 9: The Central West Branch Susquehanna River Watersheds A (Pine Creek), B (Kettle Creek) & C (Bald Eagle Creek) With a focus on Great Island, Clinton County, Pennsylvania. E.R. 2004-1413-035-H	A very detailed prehistoric data synthesis at the sub-basin scale, but no discernible archaeological predictive model.
Wall, Sara, Schmidt, and Ross	2008	Phase 1 Survey for the Armenia Mountain Wind Energy Project, Tioga and Bradford Counties, Pennsylvania. E.R. 2007-1478-042-D	Chapter 4 references previous models for the study area and considers the environmental variables that were seen as important in these studies, but does not develop an actual model as defined in this study.
Coppock	2009	Phase I Archaeological Survey, US 219 Improvement Project, Meyersdale to I-68, Somerset County, Pennsylvania and Garrett County, Maryland. E.R. 2002-8042-111-Q.	Phase I survey, no discernible archaeological predictive model.

Undoubtedly, additional reports are within the ER files that contain sensitivity or APM models, but in most cases these models will be fashioned for a specific project area and are often smaller in scale and do not attempt to advance modeling techniques. While this may be the most common type of model that is created for the scoping of field studies in CRM projects in Pennsylvania, the exclusion of many of them from this evaluation is not a deficit. Models of this type—ad hoc, non-statistical, very rarely field tested—do not offer much to this study in the way of new techniques or innovative approaches.

TYPES OF MODELS

The brief background section presented in Chapter 1 describes a number of the dichotomies and approaches used in the construction of APMs around the world. The broad methodological trends and theoretical issues discussed in the background section set the context for the specific implementations discussed here. Through the formalization of different theoretical approaches and actualization of various classes of methods, the models discussed here can be grouped into more specific model types. The more specific model types group the APMs into classes that share many common characteristics, both technical and theoretical, but also are open enough to allow for a good deal of diversity with each type. This typology serves as a convenient way to characterize models of similar structure to facilitate in-kind evaluation. Conversely, the types are not arranged as a hierarchy, and APMs may borrow from multiple types and have characteristics that blur type boundaries. As this evaluation will show, each of the model types recognized in this study are capable of identifying the location of archaeologically sensitive areas, and no one model type is the best for every situation. Figure 1 presents a schematic of model types and associated characteristics. Figure 1 divides the models evaluated for this study into two broad classes, those that use weighted variables and those that do not, and identifies eight model types defined by the methods each uses to establish the sensitivity of specific locations.

Depicted below the eight model types are five characteristics that apply to each model type. These characteristics are arranged on scales that span the various model types, but are not directly tied to any one type. For example, a *qualitative* model is generally built from poorer data quality and non-statistical methods as compared to a *correlative* model. On the other hand, this figure is not intended to imply the reverse: a direct relationship from each characteristic to a model type. Non-statistical models with poor data quality need not always be of the *qualitative* type. Any model can use poor data, but in this study models on the left end of the spectrum tend to have poorer data quality. These model types and their general characteristics represent the variety of models evaluated for this study within Pennsylvania. However, research into APMs from around the world shows that this typology is a reasonably accurate road map to understanding the variety of techniques applied across the discipline.

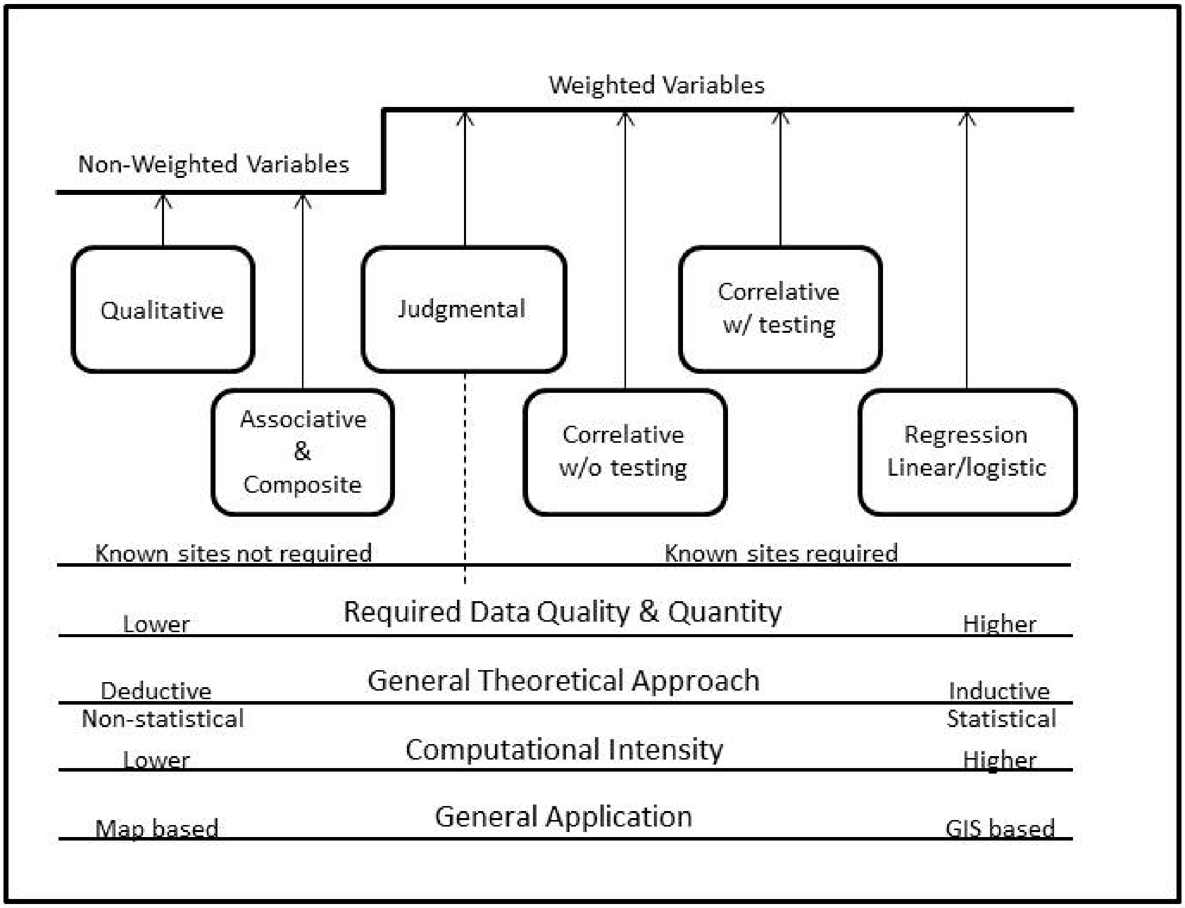

Figure 1 - Schematic of model types and characteristics within Pennsylvania APM.

Qualitative Models

The qualitative model is the most simplistic type of model encountered in this study. This model may be based on a deductive or empirical understanding of environmental and cultural factors that affect site location decisions. The distinguishing feature of this type is that the application of the site location preference model to the real world landscape does not involve taking measurements of the landscape, in the field or remotely. That is to say that the implementation of the model is achieved through "guesstimates" of the value or location of a variable on the ground. For example, if the modeler hypothesizes that areas of lesser slope are more sensitive than higher slope angles, these areas will be judgmentally delineated as opposed to measured using a digital elevation model (DEM) or topographic map. Likewise, areas of high resource productivity may be located by delineating areas where a map appears to have a range of habitats in close vicinity, as opposed to calculating a diversity index or other quantification of resource availability. This method was used predominately in the earlier APM studies when GIS data were not readily available and field measurements of the study area were impractical. Within these earlier studies the method was implemented in two ways: 1) hand-drawn based on a qualitative assessment, and 2) generalization of environmental variables across large survey blocks (square kilometer or more) into a range of sensitivity.

Characteristically, this type of model is often couched as deductive in its approach due to the assessments of sensitivity being based on the model creators' theories of settlement location. In fact, this approach is more aptly characterized as *pseudo*-deductive since the modelers' theories of settlement sensitivity are based on their knowledge of site locations. Because the correlation of known site locations to environmental variables are not measured by this type of model, the method does not require the knowledge of sites within or near the project area. As practiced, these maps are often non-GIS and difficult to validate because of the qualitative assessments and lack of measured variables.

Albeit simplistic, this model type is useful over large areas or in regions where data is very sparse. Today, however, data are seldom sparse and measuring environmental variables is a much easier task than in the past. For models covering large areas, it is impractical to hand draw every sensitive landform and equally impractical to employ survey areas on the scale of square kilometers. Further, the process used to create these models is often arbitrary and difficult to document or recreate. Therefore, this type of model is not useful or appropriate for large area models intended to assess archaeological sensitivity for planning purposes.

Associative and Composite Models

The associative model type is the most common type of model within the sample studied for this report. The composite model type, on the other hand, is the least common type within this study. These two model types are quite different in their implementation, but they are classified together because they share many of the same characteristics.

The associative model type covers a wide range of implementations, but is centered on the methods of quantifying environmental measures and associating sensitivity with the presence, absence, or proximity to those variables. This method differs from the qualitative models, in that real world variables are measured, and differs from judgmental models, in that the classes of variable measures are not weighted, arbitrarily or mathematically. Associative models most often derive site location theories based on broad regional studies that synthesize variables that appear to affect settlement. These models often describe the choice of environmental variables as deductive. As with qualitative models, this is often a pseudo-deductive approach because the choice of variables is structured upon the location of sites discovered through regional survey. Because of this, these models often do not utilize site locations within their study area to derive environmental correlations. Essentially, this model type recognizes regionally applicable associations between site locations and the presence/absence or distance from environmental features, and then applies those associations to the study area being modeled. This application is done through the measuring of the same variables within the study area and assessing sensitivity based on the presence/absence or magnitude of these measures. As noted above, this process does not involve the weighting of variable measures for their relative contribution to the overall sensitivity.

Models of associative type are among the most common because they are the easiest to apply with limited data, are intuitively compatible with archaeological settlement theories, and are rigorous enough to be quantified and repeatable. The associative models are present throughout the time span of models reviewed here, but have an origin in the earlier portion of predictive modeling. This is because that time period, the early to late-1980s, coincided with the advent of GIS, the increasing accessibility of digital data, and the availability of numerous broad regional settlement studies and surveys. With the power of GIS to measure digital environmental data and the numerous volumes of settlement pattern analysis, this type of model quickly filled the void between subjective qualitative sensitivity assessments and rigorous statistical regression models.

Composite models on the other hand are only represented by a single example within this evaluation, but share many of the same characteristics as described above. The composite model type is defined as an assessment of sensitivity derived from the composite location of many sub-models that seek to locate areas desirable to Native American settlement. These areas may be locations of specific resources or locations of cultural significance. This model type is the closest method in this study to achieving a purely deductive approach. With this type, any number of models is used to locate areas that may hold desirable resources diachronically or on a seasonal basis. These models may include the location of deer habitat, fish migrations, prime agricultural land, and specific plant communities. The specific models used are derived deductively from archaeological theories about settlement in that region. Once generated, each of these models is overlain and the total sensitivity is a composite of all of the models. Typically, the highest sensitivity areas are those that have the highest ranking for all of the combined sub-models. This could be done on any time scale from seasonal, to yearly, to all of prehistory without relying on temporal data from archaeological sites.

Because this method derives the environmental or cultural variables and the quantification of sensitivity from an essentially deductive approach, known sites with the study area are not required. Unlike the associative model, the mathematical rigor and computational intensity can be as simple or complex as the underlying models of resources are made to be. However, this exposes a weakness of this model type: the inaccuracies and variance of the underlying models are compounded into the final assessment of prehistoric sensitivity. It is easy to recognize that modeling the presence/absence and change in plant or animal communities formed thousands of years ago may be as difficult as modeling the settlement locations that are the ultimate objective of the process. The uncertainties introduced into the numerous underlying models are multiplied in the composite. The outcome of this type model is only as good as the product of the models it is composed from. Validation of this type of model can be accomplished using known prehistoric site locations within the study area without the fear of bias through model circularity.

The associative and composite model types are conceptually quite different. The associative type is very common, is often ad hoc, and it models environmental relationships at a very basic level. On the other hand, the composite model is rarely attempted, requires a high degree of planning, and models a wide range of environmental relationships in a variety of ways. The similarities include the fact that neither model requires the knowledge of archaeological site locations within the study area, is more or less deductive in approach, and does not rely on the relative weighting of variables to calculate sensitivity.

Judgmental Models

Proportionally, judgmental models likely represent a much larger percentage of APMs created for CRM than is reflected in this evaluation sample. This is because this model type, like the associative type, is a very flexible set of methods that are easy to implement, conceptually easy to understand, and easy to repeat or apply to different areas. Models of this type are very commonly used to create quick and broadly applicable sensitivity assessments to aid in scoping field work effort and budgets for larger projects. Because of this, most of the applications of these models are tucked within Phase IA reports or the front matter of survey reports and not drafted specifically as predictive modeling reports. The smaller proportion of judgmental models in this evaluation may not represent how often these models are used relative to the other types, but it does cover the various uses of this method well.

Judgmental models are relatively simple models that use weights to boost the relative contribution of certain variables or classes of a variable to account for a greater portion of the overall assessment of sensitivity. Typically, each variable is overlain within a GIS in a regular grid pattern and the weights assigned to them are summed to create the overall sensitivity, where the highest sum is interpreted as the most sensitive area. This model type can combine the use of presence/absence or distance to or from variables to contribute to the sensitivity. Variations include using negative weights to reduce the overall sensitivity (such as in areas of disturbance) or weighting factors that multiply variable weights that are seen as more important than the other variables. This method is very flexible and can be applied in many different ways.

This model type is called judgmental because the weights assigned to each variable are chosen based on expert judgment and observation. This is opposed to correlative models, which assign weights based on measuring the variables at each known site, and often non-site, location.

Known sites may be used from the modeling study area, but in many cases weights are derived from regional site syntheses and hypotheses.

Within the sample of reports evaluated here, there are two main types of weighting systems used. The first, referred to as *basic weighting*, is a method by which the presence/absence or distance classes from a variable (e.g., presence/absence of a drainage divide or distance from a river) are weighted with their relative contribution to the overall sensitivity. Presence/absence variables are given a single weight, generally only for presence, and distance variables are given a range of weights based on the desired breakdown of classes. For example the distance to a river may be broken down into classes of every 100 feet. Depending on the examples from regional studies or the judgment of the modeler, each class is given a weight corresponding to its contribution to sensitivity within that variable. Since weights are assigned relatively within a variable, each variable is assumed to contribute equally to the overall sensitivity of the landscape. Conversely, *factor weighting*, assumes that each variable may contribute differently to the overall landscape sensitivity and assigns them a weighting factor accordingly. Within this method, the classes within a variable are weighted the same as before, but are then multiplied by the contribution factor. For example, each 100-foot distance class from a river is assigned a relative weight, but then it is multiplied by the weight assigned to the proximity to a river relative to the other variables. In this example, if proximity to a river is assigned a factor of 2 and proximity to a drainage divide is a factor of 1, the river variable is considered to contribute to the overall sensitivity twice as much as the drainage divide.

This method is very versatile and allows a lot of room for experimentation and blending of weighting schemes. The popularity of this approach is based on this versatility, simple mathematics, and the idea that some variables are greater attractors for settlement that others. However, this approach does have some drawbacks, chief among which is the effect of the central tendency of summing weight. As an example, a model may be set up to contain four environmental variables, each with three classes. Each of the three classes is assigned a weight of 3, 2, or 1 for high, moderate, and low sensitivity. In this example there are no factor weights. With three possible values for four different variables, there are a total of 81 different combinations of weights possible. When the weights of each of the four variables layers is summed in the GIS, to derive the overall sensitivity layer, each of these 81 possible combinations of weights is summed into one of nine different overall sums. If four cells of a weight of 1, for low sensitivity, are summed, the total sum is four—the lowest possible. If four cells of weight 3, for high sensitivity, are summed, the overall sum is 12—the highest possible. The other 79 possible combinations between these two are summed to a value of 5, 6, 7, 8, 9, 10, or 11 (Figure 2). The large number of possible combinations is being summed into a smaller number of possible outcomes. The distribution of the 81 combinations into 9 outcomes approximates a normal distribution. Therefore, as you sum the sensitivity weights, the resulting values will tend toward the mean of the overall sensitivity.

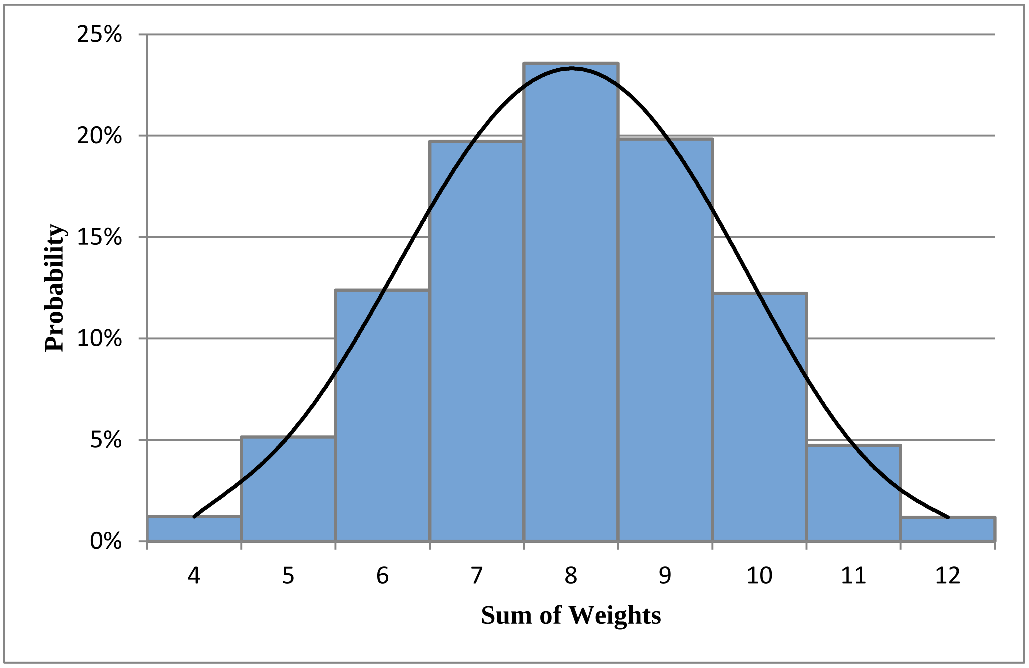

Figure 2 - Example of central tendency of summing weights.

As shown by Table 2, there is only one combination of weights that will give you a total sensitivity value of 1 or 12. Conversely, there are 19 combinations of weights that will give you a value of 8, the mean of the distribution. In a set of randomly distributed weights, you have a 1% chance of a total value of 1 and a 24% chance of a summed value of 8. Based on the assumption that the archaeological sensitivity of the study area is not random, the example model will likely be biased toward higher or lower sensitivity. However, once summed, the weights are rescaled into a more normal distribution, moving the weights toward the mean. This has the effect of clumping most of the weighted combinations toward the mean, which is moderate sensitivity. The end result is overestimating the middle, while underestimating the highs and lows. This can be countered somewhat be re-establishing high, moderate, and low sensitivity cut-off points from the final summed distribution, but for each value that the cut-off moves toward the center, it takes in an increasingly large portion of the model space, thereby diluting the model’s performance.

Table 2 - Tabular Example of Central Tendency of Summed Weight

Summed Value	Probability	Weight Combinations
4	1%	1
5	5%	4
6	12%	10
7	20%	16
8	24%	19
9	20%	16
10	12%	10
11	5%	4
12	1%	1
Total		81

Correlative with and without Testing Models

Correlative models choose variables and assign weights through correlating site locations with each variable. Typically, the correlation is observed between the number of site locations within a given distance from a variable. These models differ from judgmental type models in that weights are calculated empirically and the relationship between sites and the variables is better understood. The two varieties of correlative models discussed here are distinguished by those with testing of the correlations against background values and those without testing. The correlation of a site location to a variable signals a potential relationship between the two. However, testing this correlation against environmental background values can detect those correlations that are by chance and those that are capable of discriminating site locations from the background values. This type of model, particularly with background testing, is the first step into the realm of statistical APM.

Both forms of the correlative model type are the same in many respects. Each model begins with the choice of environmental variables that the researcher has available or feels to be useful in predicting sensitive locations. A measure of each variable is taken at each known site location, preferably within the study area, and classified into a range. This is typically visualized as a histogram (Figure 3). From this, the sensitivity of each class of the variable (e.g., 0-5% slope) is assessed based on the number or percentage of sites that are found within it. Based on the frequency of site locations within each class, they can be weighted for their contribution to overall sensitivity in a variety of ways. One method is to weight each class judgmentally, as with the previous model type, and assign arbitrary weights relative to the portion of sites within each class. Factor weights may also be used. Alternatively, weights can be assigned in a way that utilizes the information generated through the correlation step. Weights for each class may be assigned as the proportion of sites that are found within each class. Additionally, the factor weights may be set based on the strength of the correlation. This last weighting method is a more objective reflection of the data.

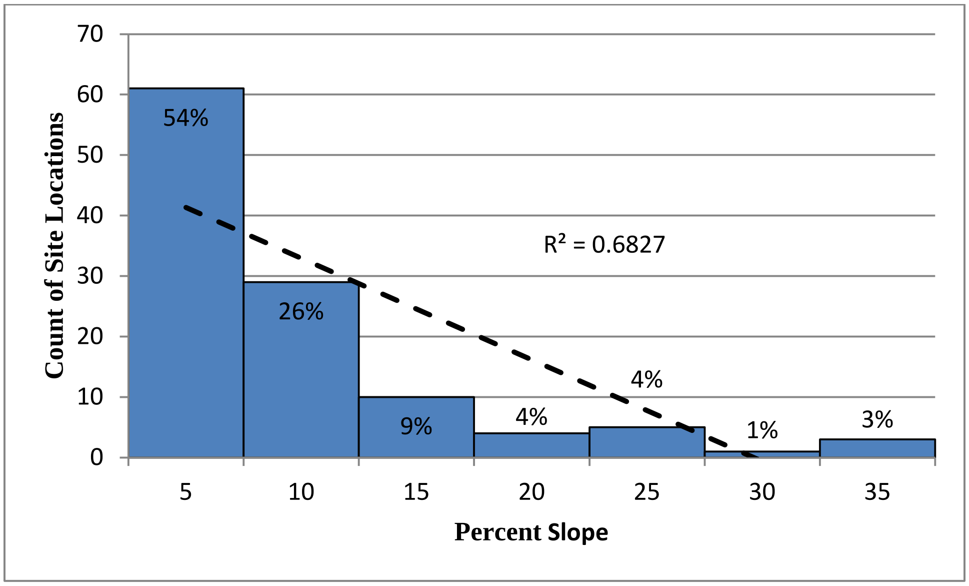

Figure 3 - Example of single variable histogram and linear fit.

Where the two forms of the correlative model type differ is in the understanding of variable to site location correlations versus environmental background values. For example, site locations may show a high correlation for being within 0-5% slope, but this may only be a reflection of the topography of the project area and the fact that much of the land is relatively level. The objective of most of these APM methods—that is, to find variables that help to distinguish the pattern of site locations within a given environment—is best served by using variables that discriminate between site locations and background values. There are a number of ways to test between measures at site locations and background values, but the most common way is to either visually inspect histograms of each or use statistical testing. Further, in either of these approaches, the background may be represented by “non-site” locations, random points, large random samples, or the full set of background values, each having their own pros and cons. Using the visual method of testing, a histogram of site location measures is overlain on a histogram of background values classified in the same manner (Figure 4). If the two histograms appear to differentiate site locations from the background, the variable may be assumed to be useful in model building. However, this method is prone to error. The less biased approach is to use the distribution of measures for both sites and background values within a statistical test. Commonly used tests for this step include the Kolmogorov-Smirnov test (K-S test) for equality of probability distributions and the Two Sample T-Test for difference of the mean. In most cases, a nonparametric test such as the K-S test or Mann-Whitney test is preferable because the data rarely conform to the normal distribution, an assumption of the T-test. Using the K-S test or similar test, if the range of a variable where archaeological sites are found is significantly different to the background values it can be said that the variable discriminates site location and therefore may be very useful in model building. An example of this is demonstrated in Figure (Figure 5), which compares the Empirical Cumulative Distribution Functions (ECDF) of the percent slope for site locations and 2,000 random points within Blair County. The outcome of the K-S test indicates that the distribution for slope recorded at site locations differs significantly from slope throughout the study area (D=0.4858, $p < 0.05$). Once each variable is accepted or rejected based on its ability to discriminate site locations, the classes of each variable can be weighted in the same manner as described in the paragraph above.

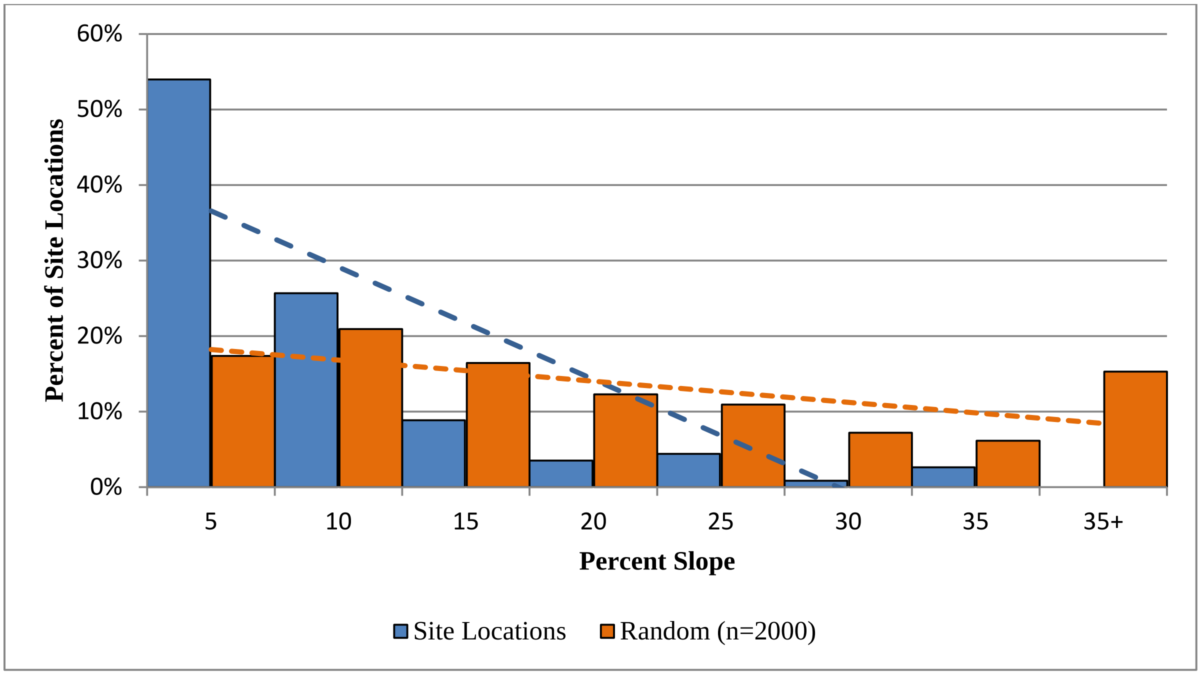

Figure 4 - Example of histograms comparing distributions of slope for site locations and background values.

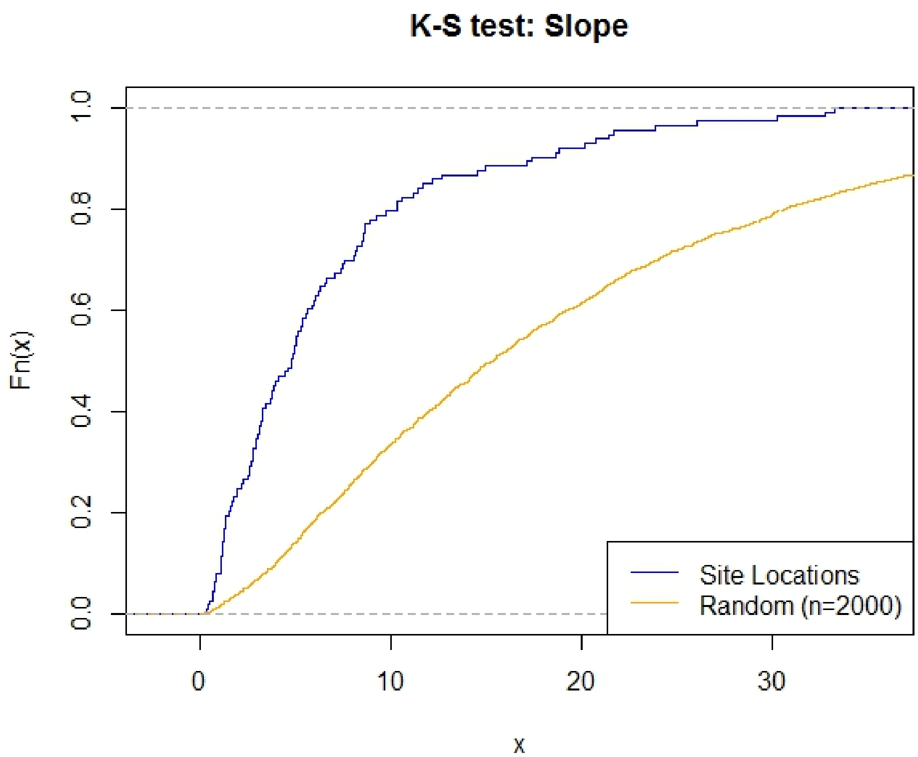

Figure 5 - Example of K-S test results comparing slope for site location and background values.

Following the steps of identifying correlations and weighting variables, this model type proceeds in a similar manner to the judgmental model type in that the overall sensitivity is most often generated by summing the sensitivity of each variable. A potential downside to this is the trending of values toward the mean, as discussed above, and the existence of multicollinearity within the data. However, the correlative model type goes much further that the previously discussed types in identifying or statistically assessing the strength and nature of the correlations between site locations and variables. This extra step removes many of the arbitrary decisions of other model types and creates a more transparent, justifiable, and repeatable model with more explicit assumptions.

Linear and Logistic Regression Models

By a simplified definition, regression models are used in APM to characterize the relationships between site presence/absence and variables and to model those relationships in order to forecast the probability of site presence/absence across a study area. While an in-depth treatment of the statistical field of regression is beyond the scope of this report, the basics of the approaches often utilized in APM will be covered.

In the most basic form ($y = \alpha + \beta x$), the use of linear regression attempts to calculate the value of site sensitivity, known as the dependent variable or *y*, based on an independent explanatory variable referred to as *x*. The slope of the regression line (β) is calculated from the two variables as the best-fit by minimizing the sum of the squared residuals, the deviations from the line. Finally, α is the y-intercept, the point at which the regression line intersects the y axis. In order to estimate the archaeological sensitivity of a location (*y*) based on the measure of a variable (*x*), the slope, y-intercept, and value for *x* would be substituted into the equation and solved for *y*. Figure 6 is a graphical example of a simple ordinary least squares (OLS) linear regression. The inclusion of more than one independent variable requires the calculation of a multiple linear regression:

$$(y = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}).$$

This method builds from the simple linear regression by modeling the relationship between explanatory variables and the response.

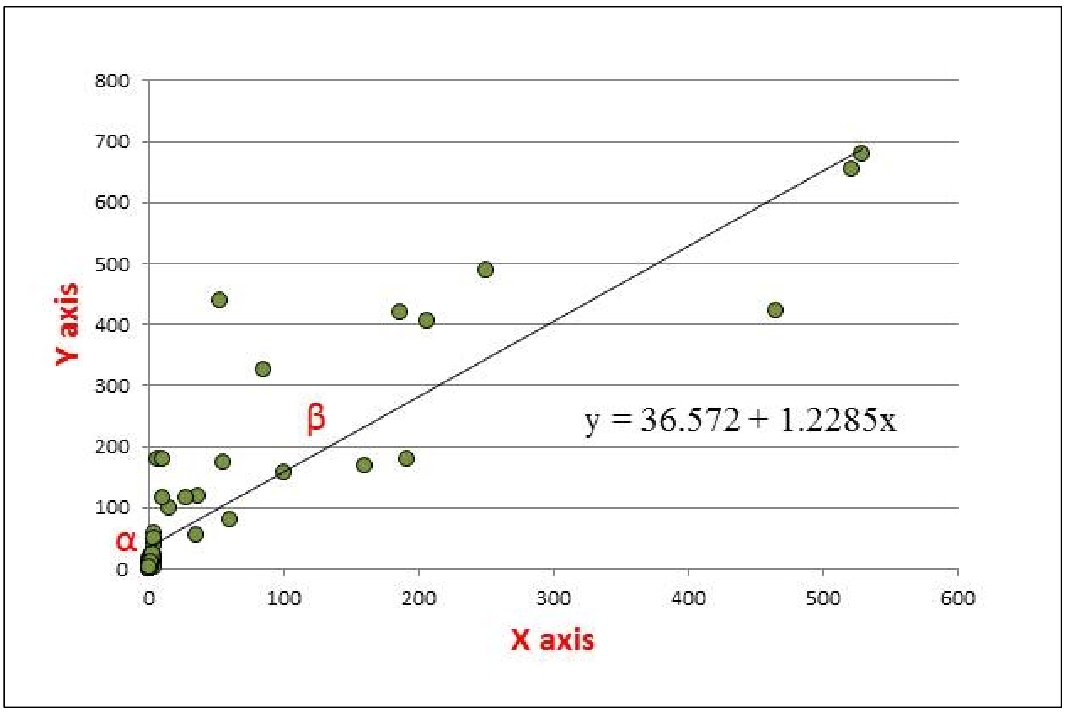

Figure 6 - Example of OLS simple linear regression and equation components.

The most prevalent issue with attempting to model archaeological site locations with this technique is that the location of archaeological sites is most often calculated as presence or absence, or numerically as 0 or 1. Because of this, it is difficult to fit a linear model to dichotomous data. Logistic regression is a special case of the generalized linear model family that utilizes a sigmoidal shaped logistic function, and its inverse the logit, to make continuous predictions from the dichotomous dependent variable. The logistic equation follows as:

$$(y = 1/(1 + Exp(\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}))$$

Figure 7 shows a simple example of the logistic function underlying the predicted values (red dots) between the binary dependent variable of site presence/absence. The resulting continuous prediction must then be segmented to represent presence or absence; this can be done in a number of ways.

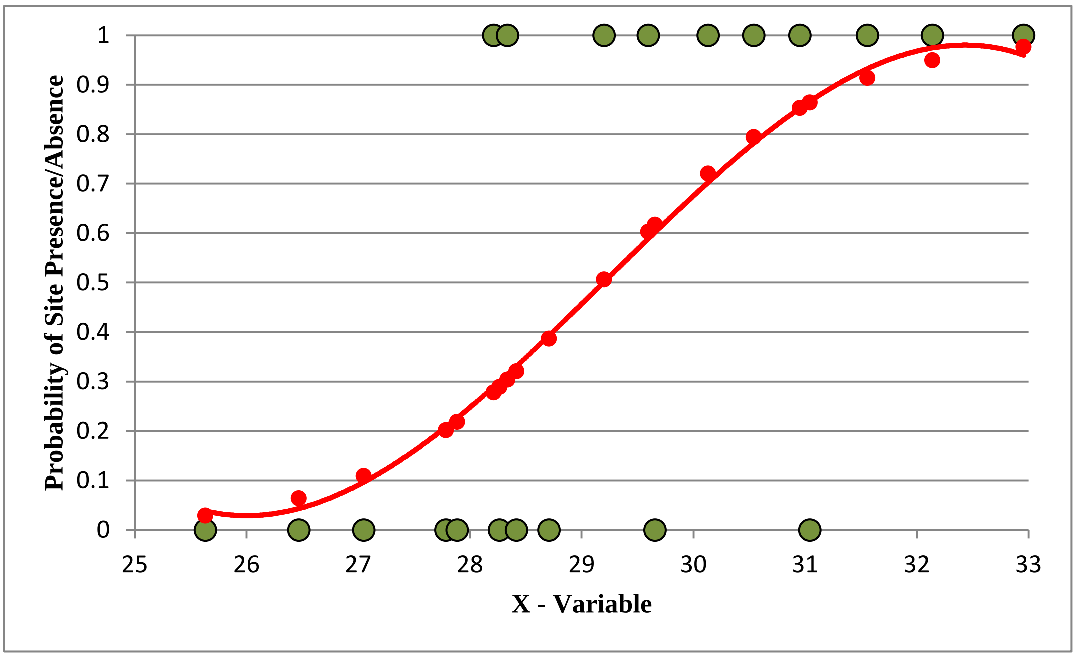

Figure 7 - Schematic of logistic function fitting site/non-site location as a binary response variable.

Expressing the results of a linear or logistic regression in map form requires the calculation of the equation for each grid cell of each variable. This is most often done in a GIS using raster images that contain a data value in each of the regularly spaced cells that cover the study area. Each of these data values are treated as a predictor value for each of the explanatory variables (x_1, x_2, x_3 ... x_i). These values are substituted into the regression equation and multiplied by the slope coefficient of each explanatory variable (β_1, β_2, β_3 ... β_i) and summed across the explanatory variable as the y-intercept. The result is a predicted value that corresponds to the scale of the dependent variable *y*. In the case of logistic regression, the result of this arithmetic is probability of being in one of the dichotomous groups, site present or site absent.

Despite a lengthy history of use within APM studies (see Kvamme 1988 for examples), regression techniques were only attempted four times within the reports evaluated here. Three of the four attempts resulted in regression being used as a way to help evaluate the relationship of variables that were then assembled in a correlative model. In only one instance (Hart 1994) was a logistic regression used as the primary model to forecast archaeological sensitivity. In this case, Hart utilized a stepwise logistic regression, in conjunction with K-S tests, to estimate the probability of site presence within each of 8,000 grid cells measuring 100 m square. The stepwise method of this model describes a technique that seeks to find the combination of explanatory variables that best describe the outcome. This is accomplished by creating a preliminary model that utilizes all of the explanatory variables, removes the least significant variable, and then repeats the model minus left-out variables until a stable model is reached. There are many different criteria for assessing which combination of variables to retain. In the case of Hart, it appears that the t-statistic was used to eliminate variables. After using a GIS to calculate the regression equation for 8,000 individual cells, Hart segmented the probability outcomes into high, moderate, and low potential.

The example described above is a good illustration of how the published method and theory within the APM literature differ from practice within CRM. The methods of rigorous statistical techniques, principally logistic regression, have been published in numerous examples (Altschul 1988; Kvamme 1988; Parker 1985; Warren 1990). Yet, models of the judgmental and correlative types are rarely published. In the gray literature of CRM, the inverse is true. Clearly the statistically advanced methods of regression are seen as more "scientific" based on their ability to quantify patterns, deliver responses that can be measured relatively, and are repeatable due to the creation of an equation. These are qualities that many researchers would value in studies such as APM. However, with such rigor comes the difficulty of implementing such models, the tedious calculations, and the application of numerous assumptions and qualifications. The effort of navigating the cost and benefit of using statistical models is the likely reason why they are not often used in CRM studies.

3. EVALUATION OF APM MODELS IN PENNSYLVANIA

A total of 32 APM reports were evaluated in this study. Of these, nine were selected for detailed evaluation based on their use of an innovative or original methodology, inclusion of adequate data to evaluate their outcomes, and ability to serve as reference points for models that followed. Most of the 32 models included in the study did not qualify for detailed evaluation and are presented in summary, by model type. The purpose of the detailed evaluations is to determine what methods have been attempted in the past and which of those attempts were successful. From this understanding, the successes of the past will be incorporated into the models created throughout this project. Similarly, the failures of the past will be recognized and examined to better understand their shortcomings.

DATA SET

The 32 APM reports included in the study were obtained from the Environmental Review (ER) archive of the Pennsylvania Historical and Museums Commission (PHMC). These reports were selected as a representative and relatively complete, albeit not exhaustive, sample of cultural resources survey projects that incorporated a prehistoric archaeological predictive model within the research design. As described in the previous chapter, a number of different methodological approaches were used in the creation of these models. Similarly, a variety of internal and external model testing strategies led to a range of outcomes regarding the ability of the model to achieve the goal of identifying areas of resource sensitivity.

The ER reports studied range in publication date from 1980 to 2010 (Table 3). The creation of APM reports in Pennsylvania began in 1980 and reached its numerical peak in the decade between 1995 and 2005 (Figure 8). There is undoubtedly a larger number of project-specific sensitivity models produced for project scoping and Phase IA studies than documented in Figure 8. However, many of these models were not included in this evaluation because they most often used a judgmental or associative model approach and do not contribute to the overall study of modeling methodology. More accurately, this figure shows the rise and fall of experimentation with new modeling techniques and application of models over large areas as the ability of computers and availability of digital data increased through the end of the twentieth century. Interestingly, this temporal distribution can be compared to national trends in APM development. Figure 9 shows the field of archaeology's "interest" in APM throughout time by counting the number of articles that mention the terms "Predictive" and "Predictive Model" within the title and abstracts of articles printed in the journal *American Antiquity*, published by the Society for American Archaeology (SAA). Also included on this figure is the same data from Figure 8 put in the national context. Overall, the literature concerning APMs peaked within the national discussion between 1985 and 1995, with still robust mention in the five years before and after. A secondary peak occurred between 2005 and 2010. It appears that the initial use of APMs in Pennsylvania began as the national discussion was in full swing and continued in use for the next 20 years. Generally, the decline in APM methodological development followed the national trend, with some exceptions that will be addressed below.

Table 3 - List of APM Studies Reviewed in this Report, in Chronological Order

Author(s)	Date	Report Title
Bailey and Dekin	1980	A Survey of Archaeology, History and Cultural Resources in the Upper Delaware National Scenic and Recreational River, Pennsylvania and New York States. E.R. 1981-0311-42-A
Johnson, Athens, Fuess, Jaramillo, and Ramos	1989	Late Prehistoric Period Monongahela Culture Site and Cultural Resource Inventory. E.R. 1989-R015-042
Neusius and Neusius	1989	A Predictive Model for Prehistoric Settlement in the Crooked Creek Drainage. E.R. 1989-R016-042-A
Stewart and Kratzer	1989	Prehistoric Site Locations on the Unglaciated Appalachian Plateau
Nass, Wright, Frye, and Krupp	1992	Phase I Historic Properties Investigations, Youghiogheny River Lake Project, Fayette and Somerset Counties, Pennsylvania and Garrett County, Maryland. E.R. 1981-0150-042-P
Whitley and Bastianini	1992	The Design and Testing of a Mathematical Archaeological Predictive Model for the APEC, DCQ, and Storage and Transport Project Areas, Pennsylvania. E.R. 1992-R001-042-A
Diamanti, Miller, Dinsmore, and Hay	1993	Predictive Model for Archaeological Resources, U.S. Route 202, Section 700, Bucks and Montgomery Counties. E.R. 1991-1019-042-KK
Hart	1994	Development of Predictive Models of Prehistoric Archaeological Site Location, for the Lake Erie Plain and Glacial Escarpment in the Erie East Side Access Project Area Erie County, Pennsylvania. E.R. 1992-0858-049-E
Perazio	1995	East Stroudsburg Area School District, Bushkill Road School Complex Project, Cultural Resources Sensitivity Study. E.R. 1995-0370-103-C
Duncan, East, and Beckman	1996	Allegheny and Washington Counties Mon/ Fayette Transportation Project Interstate 70 to Route 51. Evaluation of Crooked Creek Predictive Model. E.R. 1987-1002-042-A02 & A03
Becher, Weed, Warner, and Walsh	1997	Phase I Cultural Resources Investigations of Columbia Gas Transmission Corporation’s proposed market expansion project: Artemas storage A and B line 29520 loop in Mann, Southampton, Monroe Townships, and Bedford Counties, Pennsylvania. Volume 8, E.R. 1996-2683-009-B

Author(s)	Date	Report Title
Polglase	1997	Letter Report. Archeological Predictive Model for the ANR Independence Pipeline Project. E.R. 1984-1506-042-G
Means	1998	Phase 1 and Phase 2, Archaeological Investigations, U.S. 219 Meyersdale bypass Project S.R. 6219, Section B08, Somerset County, PA. Volume 1, E.R. 1992-0237-111-A19
Duncan and Schilling	1999a	Fayette and Washington Counties Mon/Fayette Expressway Project Uniontown to Brownsville, Archaeological Predictive Model Development. E.R. 1987-1002-042-B03
Duncan and Schilling	1999b	Northumberland, Snyder and Union Counties. Central Susquehanna Valley Transportation Project. S.R. 0015, Section 088. Archaeological Predictive Model. E.R.1997-0475-042-Q
Duncan, East, and Schilling	1999	U.S. Route 15 Improvement Project, Tioga County, Pennsylvania. S.R. 6015, Sections G20 and G22, Steuben County, New York. E.R. 1997-2018-117-H
Coppock, and Heberling	2001	Predictive Model for Archaeological Resources, US 219 Improvements Project S.R. 6219, Section 020, Somerset County Pennsylvania. E.R. 2001-8012-111-C
Duncan	2002	Centre and Clearfield Counties, Pennsylvania. S.R. 0322, Section 802 Corridor Project. Phase IA, Archaeological Investigation and Predictive Model Summary. E.R. 1999-2755-033-M
Katz, Branigan, Schopp, and Biondo	2002	S.R. 0228, Section 290 Cranberry, Adams, and Middlesex Townships, Butler County, Marshall, Pine, and Richland Townships, Allegheny County, Pennsylvania. Volume 1, E.R. 1999-6127-019-H
Lawrence, Herbstritt, Branigan, and Schopp	2002	Susquehanna Beltway Project S.R. 0220, Section 077 Woodward, Piatt, and Porter Townships and Jersey Shore, Lycoming County, Pennsylvania. Volume 1, E.R. 2002-8006-081-K
Miller	2002	Archaeological Predictive Model Report and Recommendations, PA 23 EIS Project, SR 0023, Section EIS, Lancaster County, Pennsylvania. E.R. 2003-8015-071-G
A.D. Marble & Company	2003	S.R. 1056, Section 001 Athens Bridge Replacement Project Athens Township, Bradford County, Pennsylvania. Volume 1, E.R.2000-8029-015-R
Baublitz, Richmond, and Shaffer	2003	Archaeological Predictive Model, S.R. 0830, Section 590, DuBois-Jefferson County Airport Access Project, Jefferson and Clearfield Counties, Pennsylvania. E.R. 1993-0231-065-M
Coppock, Heberling, Krilov, and Carthy	2003	Phase. I Archaeological Survey U.S. 219 Improvement Project Meyersdale to I-68 Somerset County, Pennsylvania and Garrett County, Maryland. E.R. 2001-8012-111-C

Author(s)	Date	Report Title
Mooney, Moore, Perazio, Rinehart, and Davis	2003	Phase I Cultural Resource Investigations of the Planned Bushkill Road Schools Complex, project area, Lehman Township, Pike County, Pennsylvania. E.R. 1995-0370-103-H
Baublitz and Shaffer	2004	Archeological Predictive Model South Central Centre County Transportation Study Centre County, Pennsylvania. E.R. 2000-8003-027-N
Miller and Kodlick	2006	Archaeological Predictive Model Field Results, PA 23 EIS Project, SR 0023, Section EIS, Lancaster County, Pennsylvania. E.R. 2003-8015-071-G
Blades, Vento, and Brett	2007	Pennsylvania Archaeological Data Synthesis: Deer Creek Watershed (Watershed A of the Lower Allegheny River Sub basin 18) Allegheny River Bridge Replacement, Pennsylvania Turnpike. Harmar Township, Allegheny County, Pennsylvania. E.R. 2004-0897-003-K
McIntyre	2009	East Resources Inc. Troy Pipeline Project, Lycoming and Bradford Counties, Pennsylvania. E.R. 2009-0922-042-B
Glenn	2010	Archaeological Overview and Sensitivity Models Erie National Wildlife Refuge Crawford County, Pennsylvania. E.R. 2012-1218-042-A
Reinbold	2010	Talisman Energy USA Pipelines Ostrander to Longenecker Pipeline located in Jackson Township., Tioga County and Wells Township, Bradford County, Pennsylvania and Yurkanin to Boor Pipeline, Columbia Township, Bradford County, Pennsylvania. Phase I A, Archaeological Survey and Predictive Model. E.R. 2010-1506-042-B
Yamin, Harris, McVarish, and Ziesing	2010	Independence National Historical Park Archaeological Sensitivity Study (Phase IA Archeological Assessment, Independent Living History Center, North Lot

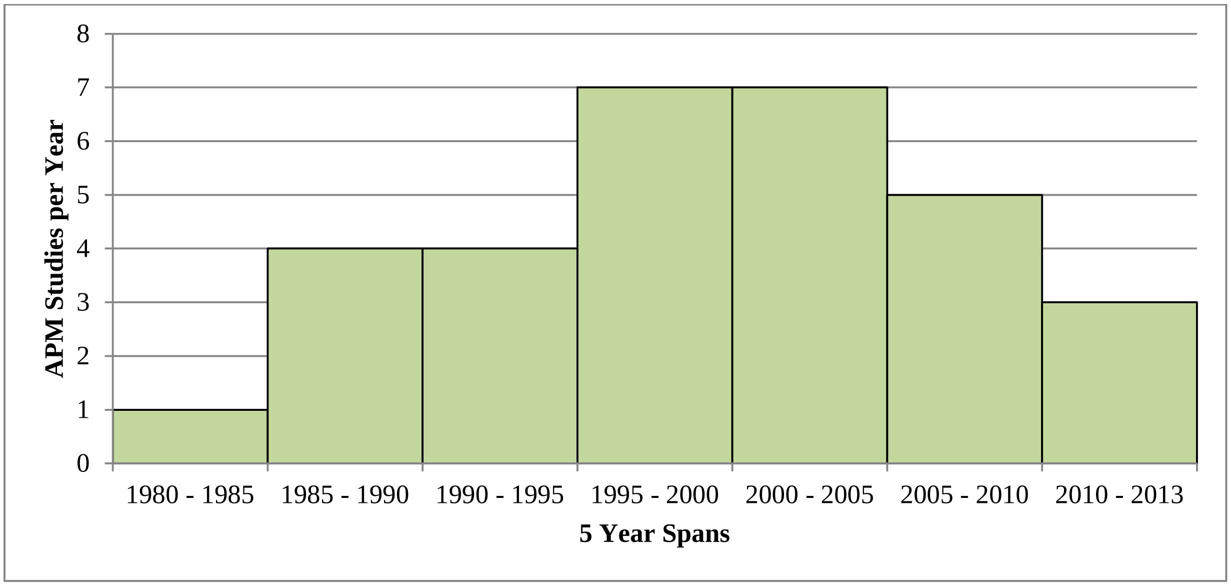

Figure 8 - Distribution of published APM reports in this study.

Figure 10 depicts the geographic distribution of report study areas throughout the state. This distribution reflects the fact that many of these models focus on transportation projects and areas with large tracts of relatively undisturbed land. That the APM study areas favor central and southwestern Pennsylvania may affect the state-wide applicability of environmental factors the authors felt most useful; however, statistical testing will be used to evaluate variables within each study region. Methodologically, it is unlikely that the geographic bias will have any effect on evaluating the usefulness of various techniques and approaches.

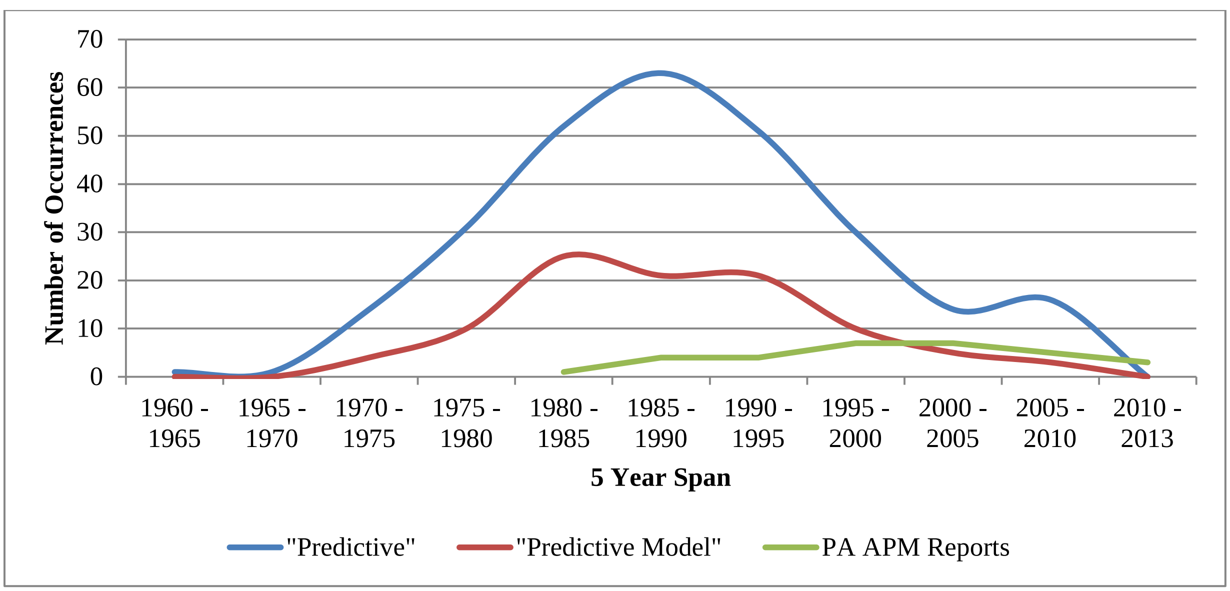

Figure 9 - Distribution of occurrences of APM terms within *American Archaeology* journal compared to reports in this study.

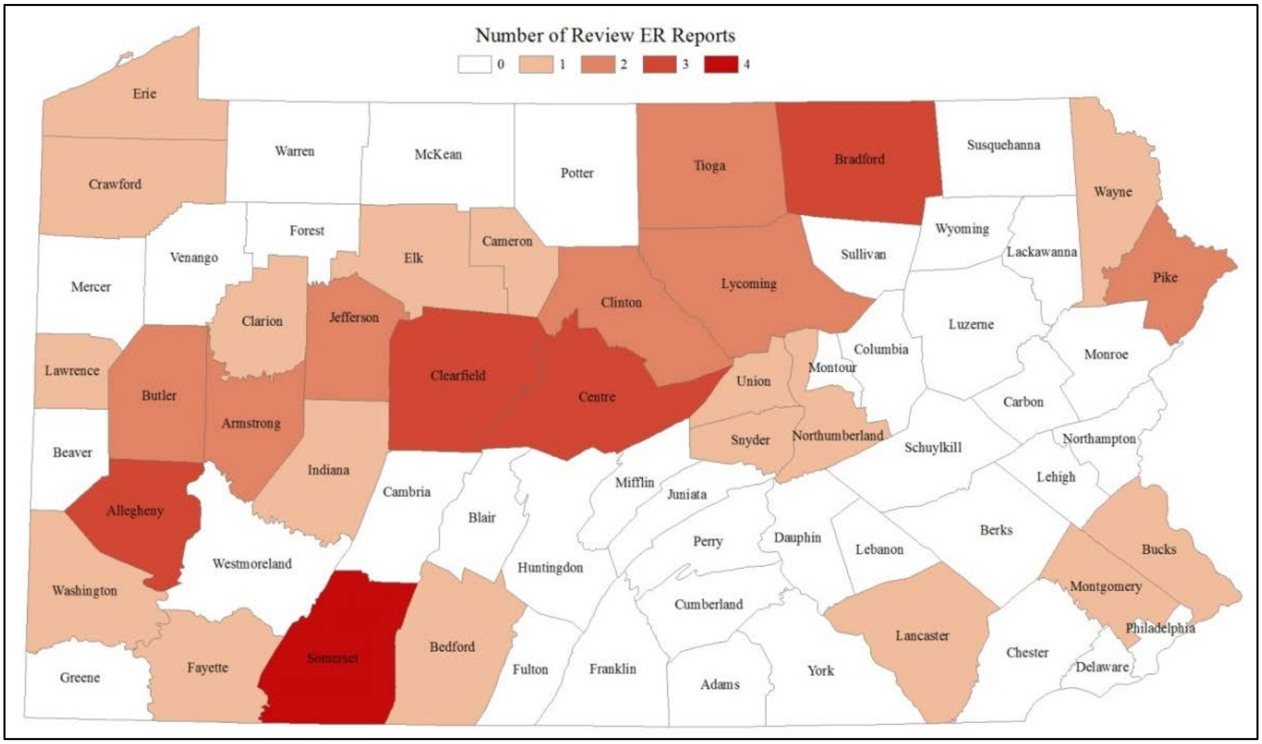

Figure 10 - Map of APM study area distribution for reports in this study.

EVALUATION METHODS

The evaluation of APMs within ER reports considers the following: 1) model type; 2) modeling steps and methods to achieve the outcome; 3) variables used to build the model and their usefulness; 4) internal and external model validation; and finally 5) model performance, that is, the ability of the model to correctly classify the location of archaeological sites and achieve its goals. This last consideration contains two very important components, the first is the model's ability to correctly classify site locations (*classification error*), and the second is its ability to achieve the modeler's goals through classification (referred to as the model's *efficacy*). The completeness of each evaluation, based on the above five points, depends on the detail, accuracy, and methodological description provided by the original author. In some cases, simply understanding the methods used to create each model requires many assumptions and the reverse engineering of data to understand how it was created. Frequently, the original report fails to provide the information required to fully evaluate the model methods and the ability of the model to predict site locations. The evaluations below will indicate where reports lack in the completeness and precision needed to fully assess their methods and findings. All attempts will be made to counter the deficient information with assumptions based on documented methods and estimations of their predictive power based on the data provided.

In the following evaluations, the description of model types will follow the typology established within Chapter 2. Many reports use hybrid model types or multiple model types within a single study area. The blending of model types and particular implementation strategy will be discussed. The model steps will be presented to show how the original author chose to approach the model type and achieve the outcome. For each of the model types, there are numerous ways to proceed from the base data to the resulting sensitivity assessment, and each approach has its merits. For each report, a table is provided that lists the environmental and cultural variables used in the model. In many cases, a larger number of variables are explored by the model builder with a smaller selection chosen for model construction. The evaluations below will list all variables, identify those chosen to represent the model, and give their relative weight if applicable. Further, the attempts of the model builder to validate their model, either internally using model data or externally using set-aside or field data, will be documented. Finally, the overall ability of the model to correctly classify areas for the presence and absence of archaeological sites will be discussed. Taken together as a measure of performance, the model's classification error will be balanced by its efficacy in achieving its cultural resources management goals. As previously noted, the measures and model data needed to accurately assess classification power are seldom reported in the sample evaluated here. However, attempts will be made to estimate the appropriate measures and evaluate the outcome relative to the model goals.

Performance: Evaluating Model Classification Error and Efficacy

Along with understanding methodology and variable choice, the most important outcome of this evaluation is assessing the performance of previous models. As previously stated, a measure of a model's performance needs to consider two very important aspects: 1) classification error; and 2) success in achieving the model's goals. The first measure of performance, classification error, is the calculation of the *actual* (or "observed") presence/absence of archaeological sites versus the *predicted* presence/absence generated from the model. Classification error can be presented through standardized tables of classification rates or shorthand in the form of the Kvamme Gain Statistic (K_g) (Kvamme 1988). While the gain statistic is convenient for comparing model results, both it and classification errors are needed to evaluate a model's performance.

The K_g is the standard measure of model efficiency used within APM literature. It is a relative measure of a model's balance between the correct prediction of site locations versus the size of the area within which sites are predicted. This balance may be described as the difference between *completeness* and *efficiency*. It is critical to note that any given model can have a wide range of K_g values depending on how the model is segmented into site-likely/site-unlikely or high, moderate, and low sensitivity areas (or any other scheme of dividing up the model's outcome). A more detailed discussion of how the statistic is derived and applied to a given model is presented in Appendix A.

The equation used to derive the statistic is the percent of the modeled area *predicted* as likely to contain sites divided by the percent of the site sample that is predicted *correctly* within that area, subtracted from 1. A K_g ranging between 0.5 and 0.8 is typical for models that are parsed so as to balance the correct prediction of sites against the disadvantages of predicting an unduly large area to contain sites. A negative gain for any area thought likely to contain sites is universally considered a failure because it actually signifies a negative prediction. Often, if the statistic falls below a K_g of 0.5, the classification of site-likely area may be too large and should be reconsidered or the model may not be a good reflection of the site pattern. Conversely, a result above a K_g of 0.8 may either indicate a very successful model or perhaps a model that is over-fit and does not consider other areas beyond those that contain the most typical sites in the settlement pattern.

The K_g is well suited to showing the success of changes made within a model or changes made to the classification of a single model. For comparing models, the K_g is best applied to models that have similar characteristics such as location, methodology, and site sample. To compare models of dissimilar characteristics, the K_g can give clues to each model's general ability to correctly classify sites based on how it is portioned into site-likely and unlikely areas. When the K_g is combined with the use of classification errors, a more accurate understanding of a model's ability is gained.

The gain statistic has drawbacks, most importantly that while it does provide a single measure of correct site predictions, it does not distinguish between the completeness or efficiency of a model's classification. A model biased toward completeness may encompass all of the known archaeological sites but do so only because the site-likely area covers a large region. On the other hand, a model biased toward efficiency may minimize the region of site-likely, but also correctly classify fewer known sites. It is possible to have very different models of a given area with the same gain statistic, but very different classification rates. In practice, the K_g is most useful when considered in tandem with the false-negative and false-positive classification error rates.

In the model evaluations presented below, classification errors are specified when they can be derived from the data in the reports. Since a minority of the reports actually contain the information necessary to derive classification error rates, however, the K_g is used as an overall measure. As noted above, the best application of the K_g is as a measure of efficiency between different iterations of the same model and to determine the most efficacious parsing of completeness versus efficiency combined with the lowest possible classification error rates. It is used in these model evaluations because it is the most common measure in the body of APM research literature and because it is flexible enough to be derived from many models even when they lack performance details. Throughout the rest of the Pennsylvania predictive model set project, the K_g will be used within the modeling process to assess how models react to different environmental variables, to determine how to most efficiently classify model results, and to compare models within modeling regions.

MODEL EVALUATIONS

In the following section, the selection of APM models chosen for detailed evaluation are presented in chronological order of publication. These evaluations will contain pertinent information on the project for which the modeling was undertaken, model type, variables, methods used, validation, and overall assessment. Where possible based on the data available in each of the reports, the validation section includes tables with values used to derive the K_g. In order to understand the content of these tables, the reader is referred to the detailed discussion in Appendix A.

Bailey, Douglas L., and Albert A. Dekin
1980 A Survey of Archaeology, Bi Story and Cultural Resources in the Upper Delaware National Scenic and Recreational River, Pennsylvania and New York States. E.R. 1981-0311-42-A. Prepared for; United States Department of Interior National Park Service Mid-Atlantic Region, Philadelphia, Pennsylvania. Public Archaeology Facility, Department of Anthropology, State University of New York, Binghamton, New York.

Region:

Initial model: Callicoon, New York, and Damascus, New York, USGS quadrangles.
Final model: Upper Delaware National Scenic and Recreational River, between Hancock, New York and Port Jervis, New York.

Significance:

The significance of this model is in its developmental approach from a simpler associative model developed for previous projects, to a more refined judgmental model that was tested and applied to a different region.

Model Type:

Judgmental

Variables:

The variables chosen for this model were all based on measures of the environment within the project area (Table 4). The particular selection of variables was influenced by previous studies within the Southern Tier of New York State, principally those of the I-88 corridor, Tioga River valley, and Elmira-Lowman Highway Corridor studies (Versaggi 1979a, 1979b).

Table 4 - Model Variables used by Bailey and Dekin (1980)

Variable
Stream Rank (Strahler systems; highest rank stream only)
Confluence (highest rank confluence only)
Slope (proportion of hex area with less than 5% slope)
Bog (largest bog only)
Lake (largest lake only)
Valley Train (largest exposure only)
Alluvial Fan (largest exposure only)
Kame (largest kame only)
Islands (in lake or river; largest island only)

Model Methodology:

The goals of this model were to: 1) test the general applicability of a form of weighted "land-suitability scoring" model for the prediction of prehistoric site locations; and 2) aid in the planning and resource management of the Upper Delaware National Scenic and Recreational River between Hancock, New York and Port Jervis, New York.

The findings of the earlier models and expert judgment were used to inform the selection and weighting of variables (Table 5). From this selection of variables, an initial model was created and applied across the Callicoon, New York, and Damascus, New York 7.5-minute USGS quadrangles. This model was applied to a 1-km hexagonal grid totaling 24 cells. These two USGS quadrangles were chosen as an ideal location to test this model methodology because each of the quads was well surveyed and a large number of sites had been identified in a variety of upland and lowland settings.

Table 5 - Variable Weights from Bailey and Dekin (1980)

Variable	Criteria	Score
Highest stream rank in hex (Strahler systems)	Delaware River	6
	After confluence of two rank 2 streams	3
	After confluence of two rank 1 streams	2
	Feeder stream	1
Highest rank stream confluence with other body of water	Incoming Stream	Rank or that stream
Slope (flat terrain of slope < 5%)	75-100% of hex	4
	50-74% of hex	3
	25-49% of hex	2
	1-24% of hex	1
	0% of hex	0
Floodplain	75-100% of hex	4
	50-74% of hex	3
	25-49% of hex	2
	1-24% of hex	1
	none	0
Physiographic features (bog, lake, kame, valley train, alluvial fan, islands	Diameter 24mm on 7.5′ USGS quad	4
	Diameter 16-24mm on 7.5′ USGS quad	3
	Diameter 8-15mm on 7.5′ USGS quad	2
	Diameter 1-7mm on 7.5′ USGS quad	1
	none	0

Based on the weights in the table below, the highest weight represented by each variable per cell was selected and summed for variables within each cell. The identification and measurement of these variables was done through measuring the USGS quadrangles by hand. Additionally, a method of using "proximity effects" was used by adding half the score of any component of the physiographic features variable that were found in a neighboring cell. In a sense, this approach anticipates and addresses the proximity issues raised as a result of Stewart and Kratzer's (1989) model (see below). The encouraging results from the application of this model to the Callicoon and Damascus quadrangles allowed the authors to apply it in the same way to the Upper Delaware River study area from Hancock to Port Jervis, New York. The variables, weights, and procedures were applied in the same manner within the two separate study areas.

Model Classification, Efficacy, and Performance:

Bailey and Dekin (1980) reported the findings of the initial test model applied to the Callicoon and Damascus USGS quadrangles. These results were based on the number of previously known sites that were not directly used to create the model weights, but the knowledge of which most likely has some influence on the authors' judgment. Results for the Upper Delaware National Scenic and Recreational River, between Hancock and Port Jervis study area were not provided. This is because the model was applied to this region as a planning tool and not for a specific survey.

The way in which the percentage of hexagonal cells and site percentages were reported does not allow for the calculation of classification rates. The total number of cells per sensitivity strata would need to be known, and trying to derive those from the information provided in the report results in conflicting estimates. However, from the information provided, K_g gain statistics can be calculated.

Bailey and Dekin stratified their model into four classes of sensitivity, Very High, High, Medium, and Low. Based on the information presented in their report, the following Table 6 can be created (note, the total of 101% of the area column is an error in the original report). Within the highest sensitivity ranking, the model included a total of 13% of the cells within the study area, which contained 61% of the known archaeological sites, for a $K_g = 0.787$. This classification appears to be very good.

With the information included in the report that can be used to identify classification errors, the true measure of this classification cannot be fully understood, but based on this information it looks quite successful. The high sensitivity stratum contained 15% of the area and 23% of the sites for a K_g of 0.348. In the medium sensitivity stratum, the classification appears to be much less successful with 48% of the area and 16% of the sites. Indeed the model classified in the correct direction, with fewer sites in larger areas. However, the jump in area was quite significant for a zone considered to have moderate sensitivity, and the negative K_g actually signals a model class with negative prediction power. The medium sensitivity class was actually less likely to contain sites than by chance alone. The low sensitivity area contained 25% of the area and zero percent of the site.

Table 6 - Kvamme Gain for Classification from Bailey and Dekin (1980)

Ranking	% of Area	% of Sites	Kg
Very High	13%	61%	0.787
High	15%	23%	0.348
Medium	48%	16%	-2.000
Low	25%	0%	n/a
Total	101%	100%	

A possible approach to remedying the very large percent of the study area present in the medium sensitivity zone would be to collapse the very high and high zones into site-likely, and medium and low zones into site-unlikely. In this segmentation, the site-likely area would contain 28% of the area and 84% of the sites, for a K_g of 0.666. The site-unlikely area would cover 73% of the study area and contain only 16% of the sites, K_g = -3.562. While a not-insignificant percent of sites are still located in site-unlikely, the percentage of the study area considered to have site potential is greatly decreased.

While there was a notable deficiency in the medium sensitivity stratum, the model overall appears to work quite well. With the data provided, it appears that the fit of percent sites in a site-likely area is quite good and reasonably balanced. However, additional information would be required to fully investigate the results.

Assessment:

As noted above, the classification of this model cannot be accurately determined, but, overall, the gain statistic for the Callicoon and Damascus, New York, test area appears promising. Following the model test application, the authors listed a series of model limitations, which included the difference in landscape between the Southern Tier of New York and the Delaware River Valley; the effect of these two regions on the classes of variables and weights; the need to ground-truth the USGS maps; and the effects of urbanization on site preservation. However, based on this finding, the authors felt that this “empirically-based test supports the applicability of the SUNY-Binghamton model for predicting prehistoric site locations throughout the Upper Delaware Valley Scenic and Recreational River” (Bailey and Dekin 1980:285).

The application of the model to the Upper Delaware National Scenic and Recreational River between Hancock and Port Jervis, New York, study area required a different stratification of the summed variable weight as compared to the test model. This was foreseen by the authors given the differences in study area physiography. While this model was not tested within the study area, the authors noted the general geographic distribution of high- and low-ranking cells through the area. They concluded with a few tentative hypotheses based on the mapping of sensitivity ranks within the study area. These observations centered around the known density of sites between the northern and southern halves of their study area and high terrain restriction; multi- versus single-component sites may have contributed to this pattern. While these observations do not aid in the assessment of the model results, they are important to the implementation and interpretation of any APM within a given study area.

It appears that the goals of this model were achieved with the use of these variables and judgmental weighting strategy. While the full diagnosis of performance cannot be achieved without more information on the model itself, it seems quite likely that this model would perform well in the intended Upper Delaware Valley study area. In essence, this model is a great example of basic "camping" strategy; searching for dry, level ground near water. The additions of weight for topographic features and the additional weighting for proximity to such features likely contribute to the model's success. The addition of these two methods set this model apart from many basic judgmental models. The use of topographic features in this case is pretty specific to the glaciated portions of Pennsylvania, and the use of stream orders is tailored around the Delaware River as the central feature. These variables and weights would require adjustment to tailor this model to other areas of the state.

Stewart, R. Michael, and Judson L. Kratzer
1989 Prehistoric Site Locations on the Unglaciated Appalachian Plateau. Pennsylvania Archaeologist, 59(1), pp.19-36.

Region:

Unglaciated Plateau in Armstrong County, Pennsylvania

Significance:

The significance of this model is that it was a transitional model from the earlier regional settlement pattern analysis studies to computerized APMs. As a study published in the journal *Pennsylvania Archaeologist*, this study is frequently cited and discussed.

Model Type:

Qualitative

Variables:

The variables utilized by Stewart and Kratzer (1989) were environmentally based. The variables listed in Table 7 represent qualitative assessments of a combination of environmental factors to which the authors associate high archaeological sensitivity. This association is based on regional survey results, experience, and analysis of 44 known sites within the region of the study (Stewart and Kratzer 1989:27).

Table 7 - Model Variables Utilized by Stewart and Kratzer (1989)

Variables
Slope < 15%
Prominent upland flats overlooking stream valley
Saddles between drainage divides and prominent upland flats
Areas near the head of active drainages
Areas near the head of inactive drainages
Upland flats adjacent to a first order stream or drainage and in proximity to a stream confluence

Model Methodology:

The model presented in this study established high and low sensitivity zones based on the environmental variable associations in Table 7. The presence and absence of these variables was determined from an analysis of 7.5-minute series USGS quadrangles of the Unglaciated Plateau physiographic section of the Kittanning areas of western Pennsylvania. The combinations of variables that contributed to an assessment of high sensitivity were established by the authors based on their intuition, regional studies, and empirically from 44 documented sites within the region.

Prior to the stratification of the study region into high and low sensitivity, the authors conducted a walking survey of portions of the area to confirm areas of disturbance and ground truth the USGS maps. Following this field check, the region was stratified based on the presence of the variable associations and field observation. Once stratified, the authors conducted a shovel test survey of the study area.

A total of 23 areas of undisclosed size along four alignments were surveyed. A 15-m shovel test interval was used within both high and low sensitivity areas. Additionally, areas with the potential for deep soils were soil augured to determine soil depth. Areas of steep terrain were walked to identify potential rockshelters. The results of this field survey were used to test the model

Model Classification, Efficacy, and Performance:

This model was not validated through the referencing of known sites. This may be because there were no known sites within the area of investigation. However, the model was field tested through the use of 15-m interval shovel test units in both high and low sensitivity areas. Areas of slope greater than 15% were inspected for the presence of rockshelters.

A total of 23 areas were surveyed for the presence of archaeological sites. Of these 23, the split between high and low sensitivity was nearly equal with 12 high and 11 low sensitivity areas. A total of 5 of the 23 survey areas contained archaeological material; 18 areas were absent of such evidence. Of the five areas containing archaeological material, two were stratified as low sensitivity and three as high sensitivity.

The classification results of this model were somewhat poor, but were likely biased to some degree by the small sample size of survey areas (Table 8). Three of the five archaeological sites were identified within high sensitivity survey areas, whereas the remaining two were in low sensitivity areas. The resulting classification indicated that 60% of the identified sites were correctly classified, but the remaining 40% were misclassified (Table 9) This was a rather large percent of misclassification. Calculating the gain based on a total of 47.8% of the survey consisting of high sensitivity containing 60% of the identified sites resulted in $K_g = 0.203$. This gain was quite low and a signal that this model classified site locations only slightly better than chance alone.

Table 8 - Probabilities of Assignment for Stewart and Kratzer (1989) Model

Probabilities of Assignment

		Model Prediction: Present (M)	Model Prediction: Absent (M')	Total	%
Site Observation	Present (S)	0.130 (3)	0.217 (2)	5	21.7%
Site Observation	Absent (S')	0.348 (8)	0.435 (10)	18	78.3%
	Total	8	12	23	1
	%	47.8%	52.2%	1	

Stewart and Kratzer (1989) designed this model to test assumptions regarding the usefulness of certain environmental variables for predicting site locations. Their intent was not necessarily to locate areas of high sensitivity for planning purposes, but more so to see if the variables referred to in many regional syntheses were truly good predictors. In this context, the end result of model classification was not the best measure of model success. The low K_g successfully addressed the goals of the model by suggesting that the variables chosen were not good predictors of site location. In this sense, the model was effective in achieving its goals. However, bias introduced through a small survey area sample and through the lack of quantifying the variables of interest cast doubt on the truth of the model results.

While this model performed somewhat well in its goal to identify the usefulness of certain environmental variables, the lack of quantification or ability to identify the contributions of each variable made the results less than useful for the current project. More to the point of the current project, this model performed only slightly better than chance in its ability to stratify areas of high archaeological sensitivity.

Table 9 - Classification Successes and Errors for Stewart and Kratzer (1989) Model

Conditional Probabilities

		Model Prediction: Present (M)	Model Prediction: Absent (M')	Total
Site Observation	Present (S)	0.60	0.40	1
Site Observation	Absent (S')	0.44	0.56	1

Assessment:

The poor classification performance of this model gives it little utility in the implementation of the statewide model for this project. However, the value in Stewart and Kratzer's (1989) study was their synthesis of regional survey findings and their own assessment of the model. As stated by the authors, the intention of this model was to blend empirically collected data with a more deductively oriented approach to variable selection. This was done in part out of a desire to draw more explanatory power from the data, but also because at the time gaining accurate measures of environmental variables was very time consuming and costly.

Stewart and Kratzer (1989:31) acknowledged that their model performed "reasonably well" given its qualitative limitations. While their assessment was arguable given the results detailed above, they did go on to provide a number of potential improvements. First, they suggested that using a cutoff of 8% slope would be more productive than that common use of 15%. Second, the authors observed that they were only testing a small portion of larger high sensitivity zones and that sites may exist in only portions of these zones. The authors stated that, "[s]ites are lacking in some areas simply because more attractive settings exist nearby. The density of prehistoric populations may have never reached the point where all potentially attractive environments are being utilized" Stewart and Kratzer (1989:31).

This is a very important observation and gets to the core of one issue with the interpretation of model results and the gain statistic. APMs most often seek to identify landforms that are similar to landforms that are known to contain archaeological sites. This does not imply that every high sensitivity landform will actually contain an archaeological site. The model essentially identifies the universe of locations from which Native Americans may have settled based on the modeled parameters. If prehistoric settlement density in the study area was quite low, then a perfectly good model may have a poor rate of successful classification. This situation would also lead to a relatively low K_g statistic because variations in the small site sample can have big consequences on the gain. The authors suggested that models may rank the zones of sensitivity relative to neighboring zones in order to combine the intrinsic landform sensitivity with that of the broader region. Finally, Stewart and Kratzer pointed out that their model would benefit from a wider range and quantification of environmental variables.

Neusius, Sarah W., and Phillip D. Neusius
1989 A Predictive Model for Prehistoric Settlement in the Crooked Creek Drainage. E.R. 1989-R016-042-A. Prepared for Pennsylvania Historic and Museum Commission, Harrisburg, Pennsylvania. Archaeology Program, Department of Sociology and Anthropology, Indiana University of Pennsylvania, Indiana, Pennsylvania.

Region:

Crooked Creek Drainage in Armstrong and Indiana Counties, Pennsylvania.

Significance:

Good example of earlier non-GIS expert judgment weighted model methodology. This model is frequently referenced by other models in this region.

Model Type:

Judgmental

Variables:

The variables selected for this model were based on basic environmental features such as slope, aspect, and distance to water, but also incorporated more complex variables such as the USDA soil ratings. In addition, this model incorporated the variable of proximity to documented historic Indian Paths (Wallace 1965). This variable was seen as a cultural or social variable, but the location of paths clearly had an environmental component. In the application of the model, the proximity to Indian paths variable was dropped due to difficulty of mapping these trails.

The selection of this set of variables was based in part on Jochim's (1976) assertion that there were three important goals that factored into hunter-gather decision on settlement location. These goals were: 1) proximity to economic resources; 2) shelter and protection from the elements; and 3) preference for a view shed that offered good defense and hunting vantage points. Neusius and Neusius focused on Jochim's second goal of shelter from the elements to aid in the establishment of the variables within this model. In order to approach aspects of the landscape that may have played a part in this goal, the authors chose the environmental variables listed in Table 10.

Table 10 - Initial Variable Selection for Upland Site Model (Neusius and Neusius 1989)

Variable
Slope
Topographic setting
Aspect
Proximity to Indian path
Proximity to water source
Agricultural suitability (USDA)
Openland wildlife suitability (USDA)
Woodland wildlife suitability (USDA)
Wetland wildlife suitability (USDA)

Model Methodology:

The goal of this model was to create "a means of predicting site occurrences which would be useful to the Pennsylvania Bureau of Historic Preservation (BHP) in its efforts to manage and protect cultural resources in this part of the state" (Neusius and Neusius 1989:3). To this end, the authors' chose to model the location of upland prehistoric sites within the Crooked Creek drainage in the Unglaciated Allegheny Plateau physiographic section. The methods chosen for this model were informed in part by this goal, but the authors were also seeking to produce a hybrid inductive/deductive approach that followed the approach discussed by Stewart and Kratzer (1989). To achieve this goal, the authors chose an expert judgment weighted sum model that was field tested.

The nine variables chosen to represent this model are listed in Table 11. These variables were selected based on data availability, previous surveys in the region, and the belief that they contributed to understanding prehistoric settlement choices. The classification and weighting of each variable was based on the same judgmental factors. The weighting scheme used a basic interval scale of weights that ranged from 0 to 3 based on the hypothesized sensitivity of each variable class (Table 11). For each quadrant within the study area the overall sensitivity score was based on the sum of each of the nine individual scores. The highest weight for each variable was obtained through the presence/absence of features or hand measurement from USGS quadrangle maps or USDA soil maps. The weight obtained for each variable was summed to produce a total sensitivity score for each quadrant. As stated by Neusius and Neusius:

> We assume that the influence of these factors is additive and that a sum of these variables can be used to score areas in terms of their attractiveness… Thus, the model will produce total scores for attractiveness which can be ranked and compared with site distributions to see if the expectation that more highly ranked areas are more likely to contain sites can be met [Neusius and Neusius 1989:29].

Table 11 - Judgmental Weight Assignment of Variables for Neusius and Neusius (1989)

Variable	Weight	Weight
Slope	0-3%	3
	4-9%	2
	10-15%	1
	>15%	0
Topographic Setting	terrace, floodplain, stream bench, river bluff	3
	saddle, hilltop, ridge top	2
	foot slope, toeslope	1
	hill slope, side slope	0
Aspect	south, southwest, east	3
	northeast, southwest, flat	2
	north, west, northwest	1
	no common direction	0
Proximity to Indian path	path located in block	3
	path in one or more adjacent blocks	2
	path one block distant	1
	no path in vicinity	0
Proximity to water source	water source located in block	3
	water source in adjacent blocks	2
	water source one block distant	1
	no water source in vicinity	0
Agricultural suitability (USDA)	class I	3
	Class II	2
	Class II and IV	1
	Class V, VI, VII, and VII	0
Openland wildlife suitability (USDA)	well suited	3
	suited	2
	poorly suited	1
	unsuited	0
Woodland wildlife suitability (USDA)	Score as for Open wildlife suitability	
Wetland wildlife suitability (USDA)	Score as for Open wildlife suitability	

This model was applied to the Elderton, Pennsylvania, 7.5-minute USGS quadrangle. The quadrangle was divided into 1,610 square quadrants measuring 300-m on a side. Each of the variables was measured for each of the quadrants and the total of the individual weights were summed to create a total sensitivity number for each cell. Within this encoding process, Neusius and Neusius dropped the variable measuring the proximity to Indian paths because they found that Wallace's (1965) maps did not translate well to the USGS base maps. The remaining eight variables were utilized. Upon deriving the summed sensitivity value for each of the 1,610 quadrants, those representing lowland settings were removed from the model. With the intention of modeling the location of upland sites, a total of 1,456 quadrants were selected for this model.

The eight variable weights were summed within 1,456 upland quadrants within the study area. The summed weights ranged from a low of 0 to a high of 13. This range was then divided into five sensitivity classes: Sites Highly Probable, Sites Probable, Sites Possible, Sites Improbable, and Sites Highly Improbable. The counts of quadrants assigned to each zone were 53, 193, 545, 532, and 133, respectively. The method by which the range of sensitivity values was divided into this five-class ranking was not disclosed.

To field test this model Neusius and Neusius selected a representative 1% sample of 15 quadrants from the total 1,456 upland quadrants. This field test was carried out through pedestrian surface survey and subsurface shovel testing. Following the guidelines on field survey established by the BHP, pedestrian surface survey was conducted in areas of greater than 50% surface visibility, and shovel test units were dug in areas of less than 50% surface visibility. Shovel test units were placed at a 10-m interval within transects that were placed 20 m apart. All shovel test units were excavated into sterile subsoil and their contents screened through 1/4-inch mesh hardware cloth. This field testing strategy was employed in the same way within each of the 15 survey quadrants regardless of their modeled sensitivity. Due to field conditions and soil disturbance, additional quadrants were surveyed to substitute for those that could not be surveyed. At the conclusion of this survey a total of six upland sites and one lowland site were identified. Forty percent of the upland quadrants were found to contain at least one prehistoric archaeological site.

Model Classification, Efficacy, and Performance:

The model constructed to locate upland sites within the Crooked Creek drainage was tested through a survey of 15 300-m square quadrants. This represented a 1% sample of the 1,456 total upland quadrants within the Elderton, Pennsylvania, USGS quadrangle. In total, six upland archaeological sites were identified within the 15 quadrants that made up the 1% sample. Table 12 shows the total cells within each sensitivity ranking, the number of cells for the 1% sample, the number of sites identified within each ranking, and the K_g statistics. As seen in this table, the sensitivity ranking of the 15 quadrants in the 1% sample were representative of the sensitivity ranking distribution of the entire 1,456 quadrant study area.

Table 12 - Ranking of Quadrants, Sites Identified, and Gain Statistic (Neusius and Neusius 1989)

Ranking	Total cells	% of cells	Surveyed cells	% of survey	Sites Identified	% of sites	Kg
Sites Highly Probable	53	4%	1	7%	1	17%	0.600
Sites Probable	193	13%	2	13%	1	17%	0.200
Sites Possible	545	37%	6	40%	3	50%	0.200
Sites Improbable	532	36%	5	33%	1	17%	-1.000
Sites Highly Improbable	133	9%	1	7%	0	0%	n/a
Total	1456	100%	15	100%	6	100%	

The six sites identified within the survey were found in four of the five sensitivity rankings. The largest number of sites (n=3) was identified in the Sites Possible ranking, which was the largest ranking in terms of quadrants surveyed and quadrants overall. The gain statistics indicate that the Site Highly Probable ranking achieved a K_g of 0.600, with 17% of the sites in 7% of the area. This gain is relatively good, but based on a very small site sample. The Sites Probable and Sites Possible zones each achieved a gain of $K_g = 0.200$, which is relatively poor, with only a slightly higher percentage of sites than the area in that ranking. The Sites Improbable ranking gain of -1.000 is quite good given the intention of that zone. The negative gain of this ranking suggests that sites are less likely to be found here based on the model than by chance alone; this is a pretty good way to define improbable.

In an attempt to quantify their results, Neusius and Neusius collapsed the highest three sensitivity rankings into a site-likely category and the lowest two into a site-unlikely category. Table 13 shows the results of the collapsed rankings in terms of numbers and proportions of quadrants with sites and those without. Of the six quadrants with sites, five were correctly classified (83%) and one was incorrectly classified (17%). Of the remaining nine quadrants without sites, four were considered site-likely (44%) and five as site-unlikely (56%). The classification percentages are depicted within Table 14. The gain statistic for the collapsed rankings is $K_g = 0.280$. This gain in combination with the classification percentages tells an interesting story of these results. On the outset, the low gain does not inspire much confidence as the classifications show that this gain suffers because of a large percentage of false-positive classification errors. On its own, the 83% correct classification and 17% false-negative misclassification of sites does not make these necessarily bad results, but the large number of false-positives drive the gain statistic down. With the assumption that false-positives are much less costly than false-negatives, the balance of this model toward completeness over efficiency shines a more positive light on the low gain of 0.280.

Table 13 - Probabilities of Assignment for Neusius and Neusius (1989) Model

Probabilities of Assignment

		Model Prediction: Present (M)	Model Prediction: Absent (M')	Total	%
Site Observation	Present (S)	0.333 (5)	0.067 (1)	6	40.0%
	Absent (S')	0.267 (4)	0.333 (5)	9	60.0%
	Total	9	6	15	1
	%	60.0%	40.0%	1	

Table 14 - Classification Success and Error for Neusius and Neusius (1989)

Conditional Probabilities

		Model Prediction: Present (M)	Model Prediction: Absent (M')	Total
Site Observation	Present (S)	0.83	0.17	1.0
	Absent (S')	0.44	0.56	1.0

Further, referring to the gain statistic, Table 12 shows that the Sites Highly Probable class performed pretty well, but when collapsed its performance was moderated by the large areas included in the next two rankings. While a Fisher's Exact test performed on the collapsed results by Neusius and Neusius indicated that there was no statistical relationship between the two rankings, this test fails to incorporate the broader implications of the test results. Given that the results presented here represent a small number of sites and only a 1% sample survey, they are encouraging. The positive aspects of a relatively high correct classification and low false-negative classification are off-set by a rather high false-positive result. While the high false-positive result leads to a low K_g, but this does not lead to a poorly performing model.

With a goal of seeking to assist the BHP in its efforts to protect cultural resources in this drainage, it is somewhat difficult to say whether this model achieved its goal. In the end, it did produce a relatively decent performing model, albeit not the most efficient in regards to false-positive results, but how it may have assisted the BHP since its creation is impossible to know. Outside of the BHP, this model clearly influenced a debate about upland settlement and APM within the state; this is clear from the frequent citations of this report. Likely, this discussion did have some degree of influence on the way BHP understood upland settlement within the unglaciated portions of the Allegheny Plateau, as well as fostering the debate on the utility of APM. In this sense, this model achieved its goal.

Assessment:

Neusius and Neusius (1989:53-55) identified a series of issues, which fall within the categories of field testing and modeling, that may have negatively affected this model. First, in regard to field testing, the authors noted that the arbitrary quadrant size was an issue relative to landform size and site size. This was an extension to the Modifiable Unit Area Problem (MAUP) (Gehlke and Biehl 1934), which stated that the arbitrariness of the areal unit of aggregation can have a dramatic effect on the outcome of statistical tests of those units. Second, Neusius and Neusius identified surface visibility as a major factor in site detection during their field effort. These are both very valid concerns and are issues with many models, not inherent to this model type or application. With small sample sizes, the effects of detecting additional sites and the classes they fall in can have dramatic effects of the resulting statistics.

The other issues noted by Neusius and Neusius concerned the difficulties of obtaining accurate measurements from paper USGS maps, as well as concerns related to the nature of archaeological site types. Concerning the USGS maps, accurate measurements are an issue with any methods, GIS or otherwise, but this concern has been greatly reduced with today's availability of high resolution elevation and environmental data. Finally, Neusius and Neusius commented that this type of model is more applicable to locating habitation sites, as opposed to special purpose, extraction, and rockshelter sites. This is another valid concern that applies to most model types. One approach to circumvent this is to build models for different site types. However, the poor data quality of most identified site samples limits the ability to model by site type.

Along with the difficulties outlined above, the problems with this model include a small sample size and somewhat difficult to interpret performance results. The positive aspects of this model are a straightforward and well documented model method and performance results that when viewed from a different perspective become encouraging. The gain statistic that appears relatively poor at first glance must be viewed with respect to the classification percentages to understand why it appears so low. The larger of the two misclassifications are in the false-positive category, which are wasteful errors that are nonetheless generally preferable over gross errors. If the summed sensitivity distribution were reclassified in different rankings, the large area of false-positives in the Sites Possible ranking could be reduced, thereby making the model more effective. While the data to do so were not presented within this report, it is likely that the model results could be improved. Even with the results as presented, the outcome of this model is encouraging and appears to be a good start to representing the upland site distribution of the Crooked Creek drainage.

Whitley, Thomas G., and Keith R. Bastianini
1992 The Design and Testing of a Mathematical Archaeological Predictive Model for the APEC, DCQ, and Storage and Transport Project Areas, Pennsylvania. E.R. 1992-R001-042-A. Prepared for Texas Eastern Gas Pipeline Company, Houston, Texas. Center for Cultural Resource Research, Pittsburgh, Pennsylvania.

Region:

The APEC study area crosses almost the entire width of southern Pennsylvania. The DCQ project covers almost the entire width of central Pennsylvania, and the Storage and Transportation study is located within Clinton and Centre Counties, Pennsylvania.

Significance:

Whitley and Bastianini created a novel and interesting approach to incorporating environmental background correlation testing into model weighting. This approach allowed for the inclusion of all the variables under consideration, but added factor and individual weights equivalent to each variable's ability to discriminate sites from background values.

Model Type:

Correlative with testing

Variables:

The variables used in this model were all based on measures of the environment (Table 15). These variables were selected based on data availability and because they were easily measured for each archaeology site. These are commonly used variables within this APM report sample. Whitley and Bastianini expressed the desire to include additional variables into this model, such as soil pH and capacity, nearest lithic source, and number of frost-free growing days, but found the measurement of such variables unfeasible given their time constraints.

Table 15 - Variables Used in Whitley and Bastianini (1992)

Variable
Elevation
Slope orientation
Percent Slope
Topographic landform
Distance to water
Stream order
Stream intersection
Depth to bedrock

Model Methodology:

The model created by Whitley and Bastianini was an outgrowth of a number of Phase I archaeological surveys conducted by the Cultural Resource Management Program (CRMP), Department of Anthropology, University of Pittsburgh. Specifically, these projects included the DCQ Contract Adjustment Program, the Storage and Transport Project, and the APEC Phase I survey, all conducted within Pennsylvania. Based on the data gathered from these surveys, the authors sought to construct a standardized predictive model that could utilize the survey data, be applied to a new area without recalculating the entire model, contribute to our understanding of prehistoric settlement, and aid in the statewide management of cultural resources. As stated by Whitley and Bastianini:

> The problematic nature of aboriginal settlement patterning and the elucidation of a processual understanding of human/environment relationships can be addressed most adequately by a model of the following design…The end result of this synthesis is the generation of aboriginal site location hypotheses which, hopefully, will be of utility in planning future cultural resource reconnaissance efforts, allowing better management for the prehistoric cultural resources of Pennsylvania as a whole [Whitley and Bastianini 1992:1].

The model methodology developed to address this goal was in essence the weighting of classes within environmental variables based on the difference of proportion between sites and background values within that class. Further, the authors proposed an approach to relatively weight each variable based on the total difference between sites and background values within all classes of that variable. This approach differs from the correlative models without testing in that the variables chosen to represent the model were tested against background environmental values to prove or disprove a significant correlation between site locations and those values. This method also differs from other applications of the correlative model with testing, in that once a significant correlation is proven, most of the models of this type either stop testing or move on to use only those variables in other models such as regression. Whitley and Bastianini utilized the results from the correlation testing to further subdivide the variables into classes that were then weighted based on the result of the correlation test. In this way, the more common method of accepting or rejecting an entire variable (e.g., elevation) based on the correlation test was replaced with a method that added positive or negative weights within classes (e.g., 100-200 feet elevation) based on the strength of the correlation within that class. The overall correlation of a variable was then reconsidered in a second step to multiply the class weights by the overall strength of correlation of that variable.

At the beginning of the model building process, Whitley and Bastianini established the similarity of site locations and background environments between the three previously surveyed project locations (DCQ, Storage and Transport, and APEC) through physiographic, ecologic, and climactic background research. Following up on this, they used a method called Multi-

Dimensional Scaling (MDS) to compare the eight environmental variables (Table 15) for each of the 92 identified sites between the three project areas. The MDS method is an ordination technique that uses a similarity matrix in the computation of *N* dimensions (in this case three dimensions) that can be plotted to visualize the similarities between data points and data sets. Whitley and Bastianini produced three plots, each comparing two dimensions of the MDS results. Based on a qualitative visual interpretation of these plots, the authors felt confident that the site locations within each of the three project areas were sufficiently similar. Based on this similarity, they stated that predictive scores generated for the site sample within the APEC study area could be applied to and tested on sites within the DCQ and Storage and Transport project areas.

This finding was the basis from which the modeling methods were built. As stated, one of the goals of this model was to create a method that allowed for the predictive scores of known sites to be calculated for one area and then exported to other areas without the recalculation of the model. The assumption here was that if two areas are similar enough in their environmental character, then the sensitivity score of a site or area should be comparable relative to the original test area. In this way, a new project area, once established as similar to the test area, could be measured based on the eight variables and then weighted based on the pre-established weights from the test area. The resulting sensitivity index of the new area could then be compared to a scale generated in the test area to judge whether it is likely to contain sites. This approach requires the calculation of the eight variables for only the area within the new project, as opposed to building an entire new model for each new project. If successful, this method would theoretically allow for the establishment of a set of variables and weights that incorporated correlations and were specific to a region or watershed.

The preparation of the environmental variables began with the segmenting of the eight variables into classes. The segmentation was for the most part arbitrary, but did consider that range of measure that archaeological sites more often fall within. For example, elevation was broken down into seven classes from 0 to 2,000 feet, but then lumped into a single class for greater than 2,000 feet because sites are less frequently found at that elevation.

The first step at establishing correlations and weights began with a comparison of the percentage of sites located within classes of a variable to the percentage of that class representing the environmental background. Whitley and Bastianini completed all of the following correlations and weighting for all of the sites within their sample, as well as the sample broken down into the site ages of Archaic and Woodland, and the site types of Habitation, Isolated Finds, Small, and Medium-Large. Table 16 is an example of a table from Whitley and Bastianini comparing the percentage of sites and background for the classes of the elevation variable for all sites. These tables were created for each of the seven site groupings for each of the eight variables. Whitley and Bastianini calculated the chi-squared statistic (χ^2) for each of the tables. In the example below (Table 16), the authors reported a χ^2 of 42.733, with 7 degrees of freedom, and a p = < 0.001. Whitley and Bastianini used this significant p-value to reject the null hypothesis that the two samples had come from the same distribution and concluded that site locations can be differentiated from background values with the elevation variable. Repeating this test for each of the 56 combinations of site groups and variables, all but 11 combinations were found to be statistically significant.

Table 16 - Example of Test against Background Values from Whitley and Bastianini (1992)

Elevation by All Sites			
Subclass	**Sites (#)**	**Sites (%)**	**Total Area (%)**
0 -250 ft	0	0.00%	0.00%
251 - 500 ft	22	25.29%	21.17%
501 - 750 ft	44	50.57%	23.74%
751 - 1000 ft	13	14.94%	9.74%
1001 - 1250 ft	6	6.90%	8.29%
1251 - 1500 ft	1	1.15%	10.29%
1501 - 1750 ft	0	0.00%	5.09%
1751 - 2000 ft	0	0.00%	6.17%
2000 + ft	1	1.15%	15.51%
Total	87	100.00%	100.00%

From these tests, the sensitivity weights were calculated for each of the variable/site groupings that showed a significant difference. The weight, or predictive score (S), was derived by subtracting the proportion of the background area of a variable class (p^2) from the proportion of sites within that class (p^1), then multiplying that number by 10. Using the example data in Table 16 from the 501-750 feet of elevation class, the calculation was $S=10(p^1 - p^2)$ or S = 10(.5057 - .2374) equates to S = 2.68. The full calculation from the example data is shown in Table 17. The multiplication by a factor of 10 was done to make the numbers more readable. This calculation was completed for each of 308 combinations of variable classes and site groups for which the chi-squared test was significant. In the cases where the chi-squared test was not significant, S = 0. The predictive score is essentially the 10 times the proportion of difference between site locations and background values for that class. The more positive the S value, the higher the percentage of sites vs. a lower percentage of background values within a class. Positive values were interrupted to be positive correlations and capable of being used to distinguish site location from random locations. Conversely, negative values were interpreted as negative correlations, but were equally useful in distinguishing site location from random locations. Values of zero were assumed to be equal to random.

Table 17 - Example of Predictive Score (S) for Elevation from Whitley and Bastianini (1992)

S-score for Elevation by all sites			
Subclass	Sites % (p^1)	Total Area % (p^2)	Predictive score (S)
0 -250 ft	0.00%	0.00%	0.00
251 - 500 ft	25.29%	21.17%	0.41
501 - 750 ft	50.57%	23.74%	2.68
751 - 1000 ft	14.94%	9.74%	0.52
1001 - 1250 ft	6.90%	8.29%	-0.14
1251 - 1500 ft	1.15%	10.29%	-0.91
1501 - 1750 ft	0.00%	5.09%	-0.51
1751 - 2000 ft	0.00%	6.17%	-0.62
2000 + ft	1.15%	15.51%	-1.44
Total	100.00%	100.00%	

The predictive index (I) of a test location can be modeled as the sum of the predictive scores (S) for each of the selected variable classes at that test location: $I = (S^1 + S^2 + S^3 + \ldots S^{n)}$. This is the same summed weight method that was used within many of the models evaluated in the current study. The I value was viewed as a relative probability of a site occurring in the test location. A positive value of I indicated a greater likelihood of finding a site, a low I value was a less likely place to find a site, and a zero I value was equivalent to random chance. The authors noted that the range in I value and what may be considered high sensitivity, was relative to the number of variables summed into that value and the character of the background environment.

In the final step to creating this model, Whitley and Bastianini computed a relative weight (W) that was used as a factor weight to understand the overall importance of each variable. This weight was calculated for each variable by subtracting the largest S value (S^{max}) from the smallest S value (S^{min}) from each variable. Following the example from Table 17, the S^{max} was calculated as 2.68 for the 501-750 foot class, and the S^{min} equals -1.44 in the 2000+ foot class, therefore $W = (S^{max} – S^{min})$ or $W = 2.68 – (-1.44)$ equates to $W = 4.12$. This value applied to the elevation variable as a whole, not just those specific classes. In some other models evaluated in this report, this weight (W) would be used to multiply the class weights to give a boost to those variables that were more strongly correlated. However, Whitley and Bastianini utilized this W score as a way to understand the impact of each variable and the dynamics between variables. Given that each class within a variable, no matter how big or small the W value, can have a positive or negative S value, the resulting predictive index (I) would be skewed if the two weights were combined.

To sum up, Whitley and Bastianini created a model building method that 1) used a chi-squared test to accept or reject the null hypothesis that sites and background values were from the same distribution; 2) divided each variable into arbitrary classes; 3) calculated the predictive score (S) for each class by subtracting proportion of background area from the proportion of sites and multiplying by 10, $S=10(p^1 - p^2)$; 4) derived the predictive index (I) by summing all the values of S for each test quadrant or area of interest; and 5) computed the relative weight (W) by subtracting the smallest S value from the largest within each variable, $W = (S^{max} - S^{min})$. The final model was interpreted by viewing the highest I value as the most likely to contain sites, and the lowest as the least likely. In terms of other models in this study, the I value is the summed weight total sensitivity and the W is a factor weight used to help interpret the model.

Whitley and Bastianini calculated predictive score (S) and relative weights (W) from each of the variable classes for the 79 sites within the APEC project area. Based on the qualitative inspection of the MDS plots, the authors felt the environments of the APEC area to be similar enough to the DCQ and Storage and Transport project areas to make them a suitable test for the model. The 13 site locations within the DCQ/Storage and Transport project areas were divided into the same classes as those in the APEC area, and the corresponding S scores were summed to derive predictive indices (I) for each of the test sites. For the three historic sites in the test sample, the range in I values was -1.53 to 1.70. The 10 prehistoric sites had I values ranging from 1.66 to 13.49.

Model Classification, Efficacy, and Performance:

Unfortunately, the data presented within the Whitley and Bastianini report make it nearly impossible to assess the effectiveness and performance of this model. The only results offered were the total sensitivity index scores for the 13 known sites within the DCQ and Storage and Transport study areas. These values ranged from I = -1.53 to I = 13.49. The lowest two values, -1.53 and -1.16, were for historic foundation sites; these were the only negative values in the site sample. The third historic site in the sample had a sensitivity index of 1.70. The remaining 10 prehistoric sites had values ranging from 1.66 to 13.49. Without an indication of background values, non-site values, or random sample values, it is impossible to know how this model performed relative to chance or any other measure. From the numbers provided, there appears to be a wide range in the sensitivity index for the 10 prehistoric sites. The sample mean of these sensitivity indices was I = 6.755, with a range of 11.83, standard deviation of 4.64, and a variance of 21.51. These statistics represent a distribution with a pretty wide range. The 95% confidence interval was ±2.87, with a range of I = 3.88 to I = 9.63.

Attempting to gauge these results against a by-chance model, a background distribution of predictive indices was assumed to range from I = -14 to 14. This is a broad assumption, but lacking the true values, it is a starting point. The distribution of indices from the 10 sites in the

DCQ study area was compared to a run of 60,000 random samples of size 10 drawn with replacement from the assumed range of I = -14 to 14. A K-S test was applied to each of the samples to test whether the sensitivity indices from the site sample differed significantly from random. The p value from the K-S tests settled at $p \approx 0.18$, rejecting the null hypothesis that the 10 site sensitivity indices were drawn from a random population. As stated, this test is based on an assumption, but it does indicate that the results of the DCQ/Storage and Transport model application performed better than a by-chance model, but perhaps not substantially. Without additional information provided by the authors, it is impossible to assess this model further.

The authors (1992:28-29) stated that this model fulfilled its goals and could be exported to other areas once recalculated for the new site sample. It may be argued that the variable classes are already tested against a by-chance model because they are calculated relative to the background values. Therefore, a positive I value is in itself a successful test against randomness. This argument may be the case, but lacking the data to support it, the results cannot be fully assessed.

In addition to the results of the DCQ/Storage and Transport model test, a note of caution is required regarding the use of the chi-squared test in establishment of positive and negative correlations. Whitley and Bastianini's use of percentages, as opposed to counts, within the chi-squared test is inconsistent with standard practice. Furthermore, the use of cells with a value less than one and a test with greater than 20% of the cells containing a value less than five violates one of the principal assumptions of the chi-squared test (Yates et al. 1999). These two issues may lead to spurious results in the correlation of variable classes and site locations. This is a significant issue, because the method presented here used the results of the chi-squared test as the basis for the remainder of the calculations. If these tests are incorrect, the results would bias all steps that followed.

A possible alternative to the chi-squared test in this situation would be a Mann-Whitney U Test. This is a non-parametric two sample test with a null hypothesis that both samples are from the same population. Conducting this test on the sample data from Table 16 yields a result of p = 0.40 (two-tailed, U = 30.5). This result is not significant and the null hypothesis that the site sample and the background sample came from the same population cannot be rejected. The application of a K-S test yields comparable results to the U test. This result is counter to the significant result of the chi-squared test reported by Whitley and Bastianini. Based on this, the model's validity, effectiveness to achieve its goals, and performance cannot be adequately assessed.

Assessment:

The model presented by Whitley and Bastianini (1992) is very creative in construction and appears to streamline the incorporation of correlation and weights into a single measure, the S score. Further, the construction of a modeling method that would allow for the assessment of environmental background value similarity and application of a model by simply calculating the variables for a given area of interest is a compelling idea. While the same could be done for any judgmentally weighted or other correlative model, the packaging of correlation and weight into a single score would simplify the process.

On the other hand, there were a number of assumptions built into this model that may or may not apply well across broad areas. The effect of dissimilarities in environmental background values would need to be thoroughly explored before the full implications of this method could be understood. Further, the building of calculations upon each other without internal checks may lead to errors being compounded throughout the process. Unfortunately, the small sample of test sites and limited test data presented in the report do not allow for the performance of this method to be assessed. Combined with the questionable use of chi-squared as a significance test, these deficits undermine the authors' qualified acceptance of this model as having satisfied its goals. Revising the correlation tests, additional applications to environmentally similar areas, and additional information of the range of background values would be very helpful in assessing this promising method.

Hart, John

1994 Development of Predictive Models of Prehistoric Archaeological Site Location, for the Lake Erie Plain and Glacial Escarpment in the Erie East Side Access Project Area Erie County, Pennsylvania. E.R. 1992-0858-049-E. Prepared for; Pennsylvania Department of Transportation, Harrisburg, Pennsylvania. GAI Consultants, Monroeville, Pennsylvania.

Region:

Lake Erie Plain and Glacial Escarpment, Erie County, Pennsylvania.

Significance:

This model represents one of the few and the best executed examples of a regression model within the sample of reports studied in this project. Hart used incorporated K-S tests for correlation testing with a stepwise logistic regression to produce this model.

Model Type:

Stepwise Logistic Regression

Variables:

The list of variables in Table 18 are environmental measures that Hart chose because of the availability of the data, their potential correlation to site location, and their use in previous regional surveys and models.

Table 18 - Variables used by Hart (1994)

Variable
Distance to nearest stream
Elevation above nearest stream
Order of nearest stream
Distance to nearest stream confluence
Elevation above nearest stream confluence
Slope (percent grade)
Topographic Relief (1,000 m neighborhood)
Elevation above Lake Erie
Soil Texture
Soil drainage class
Depth to seasonally high water table
Woodland suitability class

Model Methodology:

The goal of this project was to create an APM to be incorporated into the development of the Environmental Impact Statement (EIS) for the Erie East Side Access study. From the outcome of this model, archaeologists and engineers could gain a better understanding of the potential impacts from specific projects and project alternatives within the East Side Access study area. This goal is very much in-line with the practice of CRM and does not attempt to incorporate research as a primary goal.

Hart's study area includes 25-square miles within Erie County, Pennsylvania, and incorporates the Lake Erie Plain and Glacial Escarpment physiographic sections. Separate models were created for each of the two physiographic sections. The first step of the modeling process was data collection. Over 600 archaeological site locations were collected from five counties in three states: New York, Pennsylvania, and Ohio. This data collection was based on transcribing site locations by hand onto 7.5-minute series USGS quadrangle maps. First, sites beyond the Lake Erie Plain and Glacial Escarpment were eliminated; secondly, sites were eliminated if they: 1) lacked verifiable location data; 2) were located based on informant interviews and not field verified; 3) the recorder was uncertain; or 4) there were differences between the plotted location and UTM coordinates that could not be resolved. Finally, the site sample was further trimmed in an attempt to reduce survey bias. This was accomplished through the use of a 1,000 x 1,000-m square grid that was placed over the entire project area. For any cell that contained more than one archaeology site, all but one of the sites was eliminated from the sample. Hart's manner for choosing the site to retain from each grid cell is not discussed. This procedure was done for sites within both physiographic sections, with the resulting sample sizes of 103 prehistoric sites in the Lake Erie Plain and 32 prehistoric sites in the Glacial Escarpment. Later, Hart revises the number of sites within the Lake Erie Plain to 94 due to measurement errors.

The 1,000-m grid was also used to derive the background values from the two physiographic sections. Within the Lake Erie Plain, background values were derived from every other cell in a checkerboard fashion. Cells were excluded from the background sample if they contained recorded archaeological sites or were heavily disturbed or developed. This resulted in a total of 404 background measure for the Lake Erie Plain. A similar method was followed for the Glacial Escarpment, but only half of the background values were collected to bring the total of 118 cells more in line with the smaller sample of known sites, as compared to the Lake Erie Plain. As is the case with sites, Hart revises the number of background cells within the Lake Erie Plain to 346 due to measurement errors.

The variables listed in Table 18 were manually measured for site and background cells from USGS 7.5-minute series quadrangle maps and USDA soils maps. Distance variables were measured from the center of each cell, presence/absence was taken from anywhere within the cell, and soil variables were assigned to the class that covered the largest portion of each cell. All of the variable measurement and spatial UTM coordinates from site and background cells were recorded and then entered into a FileMaker Pro database for statistical analysis within the SYSTAT program.

The statistical analysis of site and background variable measures began with generating descriptive statistics including mean, median, standard deviation, and coefficients of variation for each variable for sites and background values. Secondly, the Kolmogorov-Smirnov (K-S) test was used to compare the Cumulative Distribution Function (CDF) of each variable for sites against background values. The null hypothesis of this test is that each CDF was drawn from the same population. To reject this null hypothesis is to suggest that distribution of site cells can be discriminated from background cells for the particular variable. To accept the null hypothesis is to suggest that the variable being tested cannot be used to distinguish site locations from the background. For the Lake Erie Plain, the K-S test resulted in the variables of distance to nearest stream confluence ($p = 0.621$) and woodland suitability ($p = 0.060$) being rejected at the $a = 0.050$ level. For the Glacial Escarpment, the K-S test resulted in six of the variables being rejected at $a = 0.050$, the variables were, elevation above the nearest stream ($p = 0.639$), elevation above nearest stream confluence ($p = 0.313$), topographical relief ($p = 0.621$), slope ($p = 0.313$), depth to water ($p = 0.066$), and woodland suitability ($p = 0.350$).

Using the variables that demonstrated a significant difference from background values, a stepwise logistic regression was conducted for both the Lake Erie Plain and Glacial Escarpment study areas. Within this method, the stepwise procedure begins the regression with all of the variables proved to it and then prunes variables until it finds the combination of variables that create the best fit model. In the case of the Lake Erie Plain study area, the five variables listed in Table 19 offered the best fit. The parameter estimates listed in this table show the net change expected in site sensitivity with a unit change in that variable. The positive estimates show variables that increase site sensitivity with increases in the variable; the negative variables decrease site sensitivity as they increase. For example, site sensitivity decreases as the distance from water increases.

In concert, these estimates show that higher sensitivity is found on areas that are relatively level, well-drained, and near a stream or confluence, but at a higher elevation above the stream or confluence. The model significance is demonstrated by p values of <0.05 for each of the variables, and an overall chi-squared result of $p < 0.001$. The null hypothesis of the chi-squared test states that there is no effect of the independent variables, taken together, on the dependent variable; this hypothesis is rejected. Table 20 shows the variables selected by the stepwise procedure for the Glacial Escarpment study area model. The parameter estimates describe a high sensitivity of areas located near low-order streams and confluence, on relatively level ground, and within well-drained soils. Each of the variables is significant at $p < 0.001$, and the overall model significance is $p < 0.001$.

Table 19 - Regression Results from Lake Erie Plain Model, Hart (1994)

Variable	Estimate	Standard Error	T-Statistic
Constant	-4.44733 (α)	0.8972	-4.9566
Distance to nearest stream (x_{1i})	-0.00355 (β_1)	0.00828	-4.29
Elevation above nearest stream (x_{2i})	0.01133 (β_2)	0.00571	1.9831
Elevation above nearest stream confluence (x_{3i})	0.00996 (β_3)	0.00437	2.2794
Slope (percent grade) (x_{4i})	-0.1423 (β_4)	0.00473	-3.0084
Soil drainage class (x_{5i})	0.94386 (β_5)	0.14792	6.4809

Model Significance: $\chi^2 = 115.726$, $df = 5$, $p < 0.001$
All variables are significant at the a= 0.05 level

Table 20 - Regression Results from Glacial Escarpment model, Hart (1994)

Variable	Estimate	Standard Error	T-Statistic
Constant	0.69389 (α)	1.21901	0.56923
Order of nearest stream (x_{1i})	-0.61439 (β_1)	0.2497	-2.4605
Distance to nearest stream confluence (x_{2i})	-0.00505 (β_2)	0.00124	-4.0799
Slope (percent grade) (x_{3i})	-0.21611 (β_3)	0.08083	-2.6738
Soil drainage class (x_{4i})	0.90722 (β_4)	0.32221	2.8157

Model Significance: $\chi^2 = 58.355$, $df = 4$, $p < 0.001$
All variables are significant at the a = 0.05 level

Prior to the application of this model, a GIS was used to grid the study areas into 100-m square cells, resulting in 8,000 cells. The process of applying the regression outcomes to each of the 8,000 cells began with the measurement of each variable within each cell. This was accomplished with the use of the GIS. With each of these measurements, the logistic regression formula was applied by substituting each parameter estimate and variable measurement, as symbolized in the tables above, into the logistic regression equation:

$$(y = 1/(1 + Exp(\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}))$$

From this equation and the data in Table 19 and Table 20, and a recreation of the variables within a GIS, Hart's models could be rerun to achieve very similar, if not identical, results.

Alternatively, new models could be run on a different mix of variables and compared to the results above to judge model fit.

The range of sensitivity probabilities that result from this process will be contained within a range of 0 to 1.0. The lowest end of the spectrum is where the model predicts a very low probability of that cell to contain the mix of variables that best predict site presence. The highest end of the scale is where the model predicts a high probability for the presence of the environmental features that often signify site presence. The shape of the distribution of sensitivity values between zero and one will depend on the environmental, site sample, variables, and fit of the model. The process of segmenting this sensitivity distribution is one where the modeler has the ability to balance the model towards either efficiency or completeness. The resulting gain statistic is directly affected by the choice of cut-points for the site-likely and site-unlikely classes. The modeler may chose at this point to lower the gain statistic in trade for a more accurate model; conversely, they may choose to maximize the gain in favor of a more precise model. This choice depends on the model maker and the model goals.

Hart chose to use the cross-over method to choose the cut-point between site-likely and site-unlikely classes. This method plots the distribution of sensitivity probabilities (0 to 1) on the x-axis and the percent correct predictions (0 to 1) on the y-axis. These two curves generally mirror each other and can be used to understand the percentage of correct predictions that can be achieved at a given cut-point. Hart follows the convention of Warren (1990:105) by selecting the point at which the two opposing plot lines cross as representing the optimum point where efficiency and completeness are balanced. Choosing a cut-point to the right of the cross-over point balances the model towards completeness, as increasingly site-likely area is incorporated to gain additional sites. Choosing a cut-point to the left of the cross-over balances the model toward efficiency, with a decrease in site-likely area retaining correct predictions for the higher probability areas. This generalization depends on the shape of the distributions and the choice of balance depends on the intention of the model.

Based on the cross-over plot, Hart chose a cut-off point of 0.25 for both the Lake Erie Plain and Glacial Escarpment models. At this cut-off point, the Lake Erie Plain model correctly classified 76% of the site cells and the Glacial Escarpment model correctly classified 81% of the site cells. To segment the model into high, moderate, and low sensitivity zones, Hart used the 0.25 cut-point as the boundary between the high and moderate classes. A cut-point of 0.11 was arbitrarily chosen to be the cut-point between moderate and low sensitivity. Finally, Hart produced a map from the GIS that contained the East Side Access project alternative routes overlain on the study area classified into 100-m square cells classified into high, moderate, and low sensitivity. This map was used in the planning of the East Side Access project.

Model Classification, Efficacy, and Performance:

Overall, Hart's models for the sensitivity of site locations within the Lake Erie Plain and Glacial Escarpment physiographic sections of the East Side Access project area perform rather well in terms of classification and efficiency. However, these models were only tested internally with the site locations that were used to train the model. Using the same sites that created the model to test the model will invariably lead to biased and likely overly optimistic assessments of model performance. Ideally, these models would be tested with site locations independent of the model. However, without such a test, this assessment is based on the interval results Hart provided (1994:12).

At the 0.25 cut-off point, the Lake Erie Plain model correctly classifies 69 of the 94 site present cells for a success of 73% (Table 21 and Table 22). The false-positive classification error of 23% is relatively balanced with a false-negative of 27%. However, a 27% false negative is on the higher side of a decent model. The types and location of sites within the false-negative cells would need to be better understood to know if this error was acceptable or not. The 0.25 cut-point of the model correctly classifies 73% of the site cells within 34% of the model area for a gain of K_g = 0.536. The even balance of classification results and the overall gain suggest a pretty efficient model.

Table 21 - Model Class Assignments from Lake Erie Plain Model, Hart (1994)

		Probabilities of Assignment Model Prediction Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.157 (69)	0.057 (25)	94	21.4%
	Absent (S')	0.184 (81)	0.602 (265)	346	78.6%
	Total	150	290	440	1
	%	34.1%	65.9%	1	

Table 22 - Model Classification Results from Lake Erie Plain Model, Hart (1994)

Conditional Probabilities

		Model Prediction: Present (M)	Model Prediction: Absent (M')	Total
Site Observation	Present (S)	0.73	0.27	1.0
	Absent (S')	0.23	0.77	1.0

At a cut-off point of 0.25, the Glacial Escarpment model correctly classifies 26 of the 32 site present cells for a success of 81% (Table 23 and Table 24). The false-positive and false-negative rates are both at 19%, attesting to the even balance of this model. With 81% of the sites being correctly classified within 32% of the site-likely cells, the model achieves a gain of $K_g = 0.606$. This efficient model performs better than the Lake Eire Plain model, both in classification error and overall gain.

Table 23 - Model Class Assignments from Glacial Escarpment Model, Hart (1994)

Probabilities of Assignment

		Model Prediction: Present (M)	Model Prediction: Absent (M')	Total	%
Site Observation	Present (S)	0.173 (26)	0.04 (6)	32	21.3%
	Absent (S')	0.147 (22)	0.64 (96)	118	78.7%
	Total	48	102	150	1
	%	32.0%	68.0%	1	

Table 24 - Model Classification Results from Glacial Escarpment Model, Hart (1994)

Conditional Probabilities

		Model Prediction: Present (M)	Model Prediction: Absent (M')	Total
Site Observation	Present (S)	0.81	0.19	1.0
	Absent (S')	0.19	0.81	1.0

In terms of this model's efficacy, the results of the classifications and gains presented above suggest that these models would fulfill their goals as a planning tool. As part of an EIS and alternatives planning, these models would very likely be able to give planners and engineers a good idea of portions of the project area that are more likely to contain prehistoric archaeological sites, relative to other portions of the project area. Clearly these models did not correctly classify all of the known sites, as no model will, but they did classify a large portion of the sites without compromising with an overly large survey area or an overly high false-negative rate. Combined with the clear presentation of pertinent data to allow for this model to be understood and repeated, these qualities illustrate well performing and successful models.

Assessment:

While the positive assessment of the models classification, efficacy, and performance are based on the results of internal testing alone, it is not unlikely that these models would also test well with an independent data set. Of course, only those tests would confirm their ability to identify areas likely to contain archaeological sites. Given the clear and concise documentation of this short report, these models would be recreated and tested with new data gathered in the nearly 20 years since their creation.

Hart's methods and presentation are very firmly based on the papers published in Judge and Sebastian (1988). Hart followed many of the conventions that were made popular during the initial development of computer-based statistical APMs in the United States. Further, where many authors paid lip service to these theories and methods of these foundational publications, few were able to obtain the resources necessary to implement these methods according to best-practices of the time. Hart countered this trend by utilizing GIS and statistical analysis software to perform tests and present the appropriate data. However, for Hart's adherence to the best-practices of the time, it is interesting that he did not attempt to test his models with an independent site sample, nor did he report his results in terms of the Kvamme gain statistic. Both of these techniques and their benefits would most likely have been quite familiar to him, given his understanding of the Kvamme and Warren's publications. Perhaps the site sample or project funding did not allow for these steps to be completed, but it would have been a beneficial use of time and added some confidence to these seemingly well-executed and well-performing models.

Duncan, Richard B., Thomas C. East, and Kristen A. Beckman
1996 Allegheny and Washington Counties Mon/ Fayette Transportation Project Interstate 70 to Route 51. Evaluation of Crooked Creek Predictive Model. E.R. 1987-1002-042-A02 & A03. Prepared for; Pennsylvania Turnpike Commission, Harrisburg, Pennsylvania. Skelly and Loy, Inc., Monroeville, Pennsylvania.

Duncan, Richard B., and Brian F. Schilling
1999a Fayette and Washington Counties Mon/Fayette Expressway Project Uniontown to Brownsville, Archaeological Predictive Model Development. E.R. 1987-1002-042-B03. Prepared for; Pennsylvania Turnpike Commission, Harrisburg, Pennsylvania. Skelly and Loy, Inc., Monroeville, Pennsylvania.

Duncan, Richard B., and Brian F. Schilling
1999b Northumberland, Snyder and Union Counties. Central Susquehanna Valley Transportation Project. S.R. 0015, Section 088. Archaeological Predictive Model. E.R.1997-0475-042-Q. Prepared for; Pennsylvania Department of Transportation, Engineering District 3-0. Harrisburg, Pennsylvania. Skelly and Loy, Inc., Monroeville, Pennsylvania.

Region:

Allegheny, Washington, Fayette, Northumberland, Snyder, and Union Counties, Pennsylvania.

Significance:

This evaluation includes three models Duncan created. These three reports chronicle the evolution of a methodology that Duncan created for the Mon/Fayette and Central Susquehanna Valley transportation (CSVT) projects. These models are significant principally due to the wide area of application; taken together, they are the most ambitious modeling effort thus far in the Commonwealth. The secondary significance lies in Duncan's creation and consideration of a very broad list of environmental variables. The method employed in these models is not particularly novel or rigorous, but is noteworthy for the scale at which it is applied.

Model Type:

Correlative with background testing and elements of regression

Variables:

The models Duncan created considered a very large list of environmental variables. These variables were for the most part based on the primary data sets of elevation, hydrology, soils, and geology. Duncan goes to lengths to calculate many different secondary variables based on these and then calculates a number of different permutations for each secondary variable.

For the Mon/Fayette models, Duncan chose to incorporate a total of 20 variables (Table 25). These variables were selected from a list totaling 72 potential variables. As stated, the 72 variables are comprised of a number of permutations for each variable and not 72 unique variables. For the CSVT model, Duncan selected a list of 17 variables from a total body of 59 variables (Table 26). The difference between the body of 59 and 72 variables from the two projects is minor. Many of the differences are expanding upon and slightly changing existing variables and permutations. The two lists of variables below are also quite similar. Roughly three-quarters of each list are the same variables. The difference includes more hydrologic variables and a proximity to chert sources in the CSVT model, and more topographic and upland focused variables in the Mon/Fayette model.

Duncan selected the variables for each model through a variety of means. Variables were selected based on expert judgment and regional studies, and then dropped or reweighted based on internal testing, correlation testing, and the results from a logistic regression. The reports describing these methods are not entirely clear on what specific variables were selected for each method, but do provide the below tables documenting the variables considered for the model.

Table 25 - Variables in Mon/Fayette Models, Duncan et al. (1996), Duncan and Schilling (1999a)

Variable (Mon/Fayette)
Agricultural soil capability ranking
Cost distance to all streams
Cost distance to confluence along perennial streams, major tributaries, and the rivers
Cost distance to confluence along the river
Cost distance to drainage divides
Cost distance to historically documented trails
Cost distance to major tributary
Cost distance to nearest river body
Cost distance to topographic saddle
Cost distance to localized peaks
Cost distance to vantage points
Distance to springs
High potential soils
Openland wildlife soil suitability
Soil depth to bedrock rank
Soil drainage character
Soil fertility: corn productivity
Solar insolation gain
Topographic relief within 900m neighborhood
Topographic slope

Table 26 - Variables in CSVT Model, Duncan and Schilling (1999b)

Variable (Central Susquehanna)
Agricultural soil capability ranking
Cost distance to all streams
Cost distance to chert resources
Cost distance to confluence along perennial streams, major tributaries, and the rivers
Cost distance to headwater confluence
Cost distance to historically documented trails
Cost distance to major tributary
Cost distance to nearest river body
Cost distance to topographic saddle
Cost distance to wetlands
Flood frequency and hydrologic group
High potential soils
Soil depth to bedrock rank
Soil drainage character
Soil fertility: corn productivity
Solar insolation gain
Topographic slope

Model Methodology:

The goal of the models constructed for these studies each were designed to assist in the design, alternative selections, and testing of transportation projects. As stated by Duncan and Schilling:

> The results of the predictive model for prehistoric archaeological site potential will be utilized within the proposed highway alternatives selection process to minimize the impacts of the project on significant archaeological resources and to reduce the cost and work effort required for subsequent archaeological testing and/or mitigation within the selected preferred alternatives [Duncan and Schilling 1999a].

As with other models evaluated in this study, these models were created primarily for a planning purpose, then often reused for a scoping and field methods purpose. Duncan is perhaps the first author in this evaluation to impart the added goal of cost-savings benefit. This benefit is somewhat implied in most models, as they are rarely of pure research interest; however, few—if any—other authors admit this as an explicit goal. As stated by Duncan and Schilling:

> The chief reasons to employ predictive modeling in archaeology today are cost effectiveness and planning utility. While environmental and preservation regulations mandate that state and federal agencies locate and preserve cultural resources, sufficient funding is not available to completely inventory all such resources in all mandated areas. In order to be cost effective, predictive models must potentially be able to project likely

> cultural resource distributions across an area based on a sample of that region or on fundamental notions of human behavior (Kohler and Parker 1986) [Duncan and Schilling 1999b:21].

The methodology of Duncan's models evolves across these reports, but is summed up most clearly by Duncan and Schilling's (1999b) CSVT model. The following discussion of methods will primarily refer to the CSVT project, but will reference the Mon/Fayette model where the two approaches depart. The overall model type employed here is a judgmentally weighted sum model with variable selection and informed through site and background correlation testing. The inclusion of a stepwise logistic regression step appears to be done to help inform the weighting scheme, but its specific contribution is unclear. The general steps used in these models are as follows: 1) collect and vet PASS site data; 2) collect primary variable data, create secondary variable raster layers, and transform variables into common scale; 3) use qualitative and statistical means to test the correlation of sites and background data for each variable; 4) create an initial weighted model based on variables with strong correlation; 5) conduct stepwise logistic regression to inform weighting scheme; 6) revise variable selection and weighting scheme based on internal testing; and 7) produce revised model and test with external or independent data. This evaluation will summarize this process.

The first step in these models was the collection of archaeological site locations. Prehistoric sites from within the study area were collected from the PASS files, as well as CRM reports, published resources, and collector interviews. Paper and map based site information was digitized into a GIS and sites with conflicting information or lacking accurate spatial locations were excluded. Additionally, the site types of rock shelters, petroglyphs, and isolated finds were dropped from consideration. Information such as a site's slope, distance to water, soils, and other environmental information recorded in the PASS files was omitted in favor of recalculating these measures from the GIS. Duncan then tabulated the site data by type, location, temporal affiliation, and by some environmental variables in order to better understand the variability within the data. The total site sample was split into 70% to 30% portions for the exclusive purposes of model creation and internal model validation, respectively.

Following this data collection, the primary environmental data sets are created. These include digital elevation models (DEM), hydrology, roads, soils, and landuse. Where not available digitally, these variables were digitized into the GIS. From these, the 59 variables for the CSVT model and 72 variables of the Mon/Fayette model were generated. The raster GIS layers that represent these variables had a resolution of 30 x 30 m, commensurate with the original DEM data. The many different permutations of "cost distance" variables utilized by Duncan appear to be all cost allocated by slope. However, this detail is not explained within the reports.

At this point, the models are split into two separate models to represent upland and lowland areas. This is done due to the significant difference in environments and site locations between the uplands and lowlands. From this point on, all of the steps are repeated for each of the two modeled areas.

The statistical analysis of the potential variables began with applying the Kolmogorov-Smirnov two-sample (K-S) test to compare the distributions of environmental measures for the site sample versus a non-site sample. The site sample is represented by the 70% of the total site sample randomly selected and intended only for model creation. The non-site sample is a random sample of background grid cells, of roughly the same area as the site sample, that represent areas where no known sites exist. The intention of the non-sites is to represent the background environment and provide a basis for testing which variables are more strongly correlated to site locations. This follows the logic that settlement locations are non-random and can be differentiated from the background environmental data based on the model variables. If the distribution of measurement for a variable for site locations is significantly different from the distribution of the same measure for non-sites, the K-S test will result in a large distance value (D) coupled with a small p-value. In this case, the variable is considered able to discriminate site location.

Duncan conducted the K-S test for each of the variables and sorted the list of variables by the D value to gain a relative view of variables scaled from the most preferential to the least. In order to confirm these results and gain a better understanding of the internal variation within a variable, Duncan created histograms of both site and non-site values for each variable to visually compare the two. This allows for the understanding of which classes within each variable are more discriminant than others. This is an important piece of information when employing the weighted sum method. From these results, Duncan compared the significant variables to the findings of previous regional surveys and settlement analyses. The final list of variables used in each model was selected by a combination of the statistical tests and expert judgment.

From the selected variables, an initial model was constructed. The first step in creating this initial model was the transformation of the variable measures into a standardized scale ranging from 0 to 100. This involved rescaling or inverting variables so that the values most associated with site locations were transformed to 100, and the least associated values to zero. This seemingly very important step in the process is not well detailed by Duncan, but summed up in the statement, "The transformation process was carefully performed using simple mathematical formulas, and the results were inspected for aberrant or unintentional changes in data relations" (Duncan and Schilling 1999b:44). From these new values, the variables were classified and weighted based on the relative significance (D values), site distribution within classes, expert judgment, and the aid of a logistic regression.

The stepwise logistic regression analysis was undertaken to support or refute the selection and validity of variables. The logic of this step is that if the blend of variables selected by the stepwise logistic regression and the significance of the variables and model overall are similar to the understanding gained from the previous steps, then the model is supported. On the other hand, if the results of the regression show that the selected model or significance of certain variables differed from the previous understanding, the model may require revision. While this step is not well documented in any of Duncan's reports, he states in Duncan and Schilling (1999b:45) that:

> Although the results of the logistic regression analysis supported the use of the variables selected for the model, the weights created by the regression analysis result from a complicated interaction of the variables within the analysis and could not be directly applied to the algebraic model formula. The weights were used as a guide to adjust the correlative model…

From this it is assumed that the results of the regression acted as a guide, but did not contribute directly to the resulting model.

Following the regression and any further revisions to the weighting scheme, the initial upland and lowland models were finalized by summing the individual weights for each variable within each 30 x 30-m cell in the project area GIS raster layer. Higher sensitivity is signified by a higher value of the summed weights. Duncan does not disclose the weights of variables or whether the weights were only for individual classes or products of relative and factor weights for any of the models. For model testing and application, the upland and lowland models were combined into a single continuous sensitivity layer.

The initial model was validated internally with the site sample used to create it. Based on the results of this test, the model was refined to create the final model. The final model was then tested against the reserved test site sample to gauge its performance on independent data. With satisfactory performance, this model was then applied to the study area and utilized in the planning process.

Generally, the approach Duncan documented is similar to many of the judgmentally weighted and correlative models evaluated thus far. Where these models differ are in the complete implementation within GIS, which allowed for the creation of many environmental variables, multiple rounds of testing, and application across wide areas. Methodologically, the use of statistical testing of correlation backed up by visual comparison of histograms and model validation through stepwise logistic regression is a wide ranging compilation of techniques that are more often employed individually. Duncan's use of internal validation, revision, external testing with a set-aside sample, further revision, and the production of a final model follows best practices. Adherence to this cycle of testing and revision is all but absent in the remaining models of this evaluation study.

Model Classification, Efficacy, and Performance:

The three models evaluated here—Duncan et al. 1996 and Duncan and Schilling 1999a for the Mon/Fayette models, and Duncan and Schilling 1999b for the CSVT model—will be discussed separately, but are all the result of the same basic methods described above. All in all, the model results appear to be good and consistent with the better performing models evaluated here.

Reported in Duncan et al. (1996), the model for the Allegheny to Washington Counties section of the Mon/Fayette project, was the first of the three models to utilize these methods. Duncan used the cross-over plot method to pick the optimal cut-point for which to establish balanced site-likely and site-unlikely areas. In this model, a cut-point of 245 was used. The results of the internal model are depicted in Table 27 and Table 28. This model achieved a gain of $K_g = 0.588$, with a correct classification of 82% of the site cells within 33.7% of the study area. The erroneous classifications account for 18% of site cells and 32% of the non-site cells. The results of the initial model depict a good correct classification and a good balance of misclassifications.

Table 29 and Table 30 depict the results of the revised model from Duncan et al. 1996 as tested on independent data. The gain of this model is $K_g = 0.593$, with a correct classification of 83% of the site cells within 33.8% of the study area. The erroneous classifications account for 17% of site cells and 33% of the non-site cells. The cut-point for this model was set at 220, slightly below the initial model. This adjustment in cut-point may account for the slight difference between these two models. The 1% increase in correct site classification is offset by a 1% increase in false-positive classification. All the change between the two models is incremental; the end result being an efficient model that would serve well as a planning tool.

The second model using this method is for the Fayette and Washington Counties section of the Mon/Fayette project (Duncan and Schilling 1999a). In this model, a cut-point of 245 was used. The results of the internal model are depicted in Table 31 and Table 32. This model achieved a gain of $K_g = 0.532$, with a correct classification of 75% of the site cells within 34.9% of the study area. The erroneous classifications account for 25% of site cells and 27% of the non-site cells. This initial model has a moderate correct classification of site cells, but suffers from a relatively poor false-negative percent. The lower gain reflects the broader site-likely area covered by this model.

The revised model (Table 33 and Table 34) tested with external data had a cut-point of 199. This model achieved a gain of $K_g = 0.631$, with a correct classification of 82% of the site cells within 30.1% of the study area. The erroneous classifications account for 18% of site cells and 19% of

the non-site cells. This model is a rather significant improvement over the initial model. Overall, this model has very favorable results and would be seemingly quite effective as a planning tool.
The final model evaluated under this methodology is the CSVT model report in Duncan and Schilling (1999b). In this model, a cut-point of 120 was used. The results of the internal model are depicted in Table 35 and Table 36. This model achieved a gain of $K_g = 0.308$, with a correct classification of 82% of the site cells within 56.8% of the study area. The erroneous classifications account for 18% of site cells and 22% of the non-site cells. The low gain statistic is resultant from a large site-likely area; however, the low false-positive suggests that distribution of site and non-site cells may be affecting these results in a different way than the previous models. More information would be required to investigate this.

The revised model (Table 37 and Table 38) tested with external data had a cut-point of 122. This model achieved a gain of $K_g = 0.382$, with a correct classification of 82% of the site cells within 50.7% of the study area. The erroneous classifications account for 18% of site cells and 20% of the non-site cells. This model is an improvement over the initial model. The revised model was able to reduce the site-likely area without reducing the number correct classification of site cells or adding to the misclassifications. The revised model is an improvement over the initial model, but still includes a majority of the study area as likely to contain sites; the low gain is a result of this. With nearly the same methodology between the Mon/Fayette and CSVT models, the difference in results may be attributed to different site distributions, varying environments, and additional variables that contribute to site selection. While the CSVT model results are slanted heavily towards completeness with little efficiency, a model with an 82% correct classification and 18% false-negative is still likely a serviceable model for planning purposes.

Table 27 - Internal Model Class Assignments for Duncan et al. (1996)

		Probabilities of Assignment			
		Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.035 (72)	0.008 (16)	88	4.2%
	Absent (S')	0.303 (630)	0.655 (1364)	1994	95.8%
	Total	702	1380	2082	1
	%	33.7%	66.3%	1	

Table 28 - Internal Model Classification Results for Duncan et al. (1996)

		Conditional Probabilities Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.82	0.18	1.0
	Absent (S')	0.32	0.68	1.0

Table 29 - External Model Class Assignments from Duncan et al. (1996)

		Probabilities of Assignment Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.008 (44)	0.002 (9)	53	1.0%
	Absent (S')	0.33 (1735)	0.66 (3474)	5209	99.0%
	Total	1779	3483	5262	1
	%	33.8%	66.2%	1	

Table 30 - External Model Classification Results from Duncan et al. (1996)

		Conditional Probabilities Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.83	0.17	1.0
	Absent (S')	0.33	0.67	1.0

Table 31 - Internal Model Class Assignments for Duncan and Schilling (1999a)

Probabilities of Assignment

		Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.131 (65)	0.044 (22)	87	17.5%
	Absent (S')	0.219 (109)	0.606 (302)	411	82.5%
	Total	174	324	498	1
	%	34.9%	65.1%	1	

Table 32 - Internal Model Classification Results from Duncan and Schilling (1999a)

Conditional Probabilities

		Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.75	0.25	1.0
	Absent (S')	0.27	0.73	1.0

Table 33 - External Model Class Assignments from Duncan and Schilling (1999a)

Probabilities of Assignment

		Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.143 (71)	0.032 (16)	87	17.5%
	Absent (S')	0.159 (79)	0.667 (332)	411	82.5%
	Total	150	348	498	1
	%	30.1%	69.9%	1	

Table 34 - External Model Classification Results from Duncan and Schilling (1999a)

		Conditional Probabilities Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.82	0.18	1.0
	Absent (S')	0.19	0.81	1.0

Table 35 - Internal Model Class Assignments from Duncan and Schilling (1999b)

		Probabilities of Assignment Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.477 (4603)	0.104 (1005)	5608	58.2%
	Absent (S')	0.091 (874)	0.328 (3160)	4034	41.8%
	Total	5477	4165	9642	1
	%	56.8%	43.2%	1	

Table 36 - Internal Model Classification Results from Duncan and Schilling (1999b)

		Conditional Probabilities Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.82	0.18	1.0
	Absent (S')	0.22	0.78	1.0

Table 37 - External Model Class Assignments from Duncan and Schilling (1999b)

		Probabilities of Assignment			
		Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.406 (3239)	0.088 (706)	3945	49.4%
	Absent (S')	0.101 (808)	0.404 (3226)	4034	50.6%
	Total	4047	3932	7979	1
	%	50.7%	49.3%	1	

Table 38 - External Model Classification Results from Duncan and Schilling (1999b)

		Conditional Probabilities		
		Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.82	0.18	1.0
	Absent (S')	0.20	0.80	1.0

Assessment:

The model building process Duncan conducted for the Mon/Fayette and CSVT projects is an interesting mix of qualitative and quantitative methods that appear to have effective results. The only downside to the approach outlined in these reports is that very little information is offered in terms of judgmental decisions for weighting and class selection. Further, many of the more technical steps are glossed over and not described in any detail. The methodological approach described above, in some cases, required teasing of details from all three reports to understand the approach. By the successful results reported for most of these models, the lack of technical detail does not seem to represent a lack of understanding in the techniques.

As stated in the results evaluation, these models provide good examples of how the gain and associated statistics are very useful in comparing different iterations of a model with the same data. This is particularly true for the weighted sum model types. For regression models, other metrics generated through the estimation of parameter values provide a basis for model comparison. The steady increase in gain between Duncan's internal testing and revised externally tested models is a very positive sign, especially considering that the internal testing is likely to bias towards positive results. Further, these models are a good example of how the same methods applied to a different area and data set can give quite different results. The consistent results of the Mon/Fayette models are contrasted against the much less precise results of the CSVT model. Many factors outside of the model methods may have led to this divergence. These may include site density, physiography, and site types, as well as simple differences in the random selection of site test samples and non-site sample. Without recycling a model through numerous random selections of sites and non-sites, the volatility of the model is unknown. Poor results may be more an effect of a non-representative sample or site sample inhomogeneity than variable selection, weighting, or methods. This effect could also lead to a very good model run based on non-representative data. Testing the volatility of a model lends a greater degree of confidence in its results.

While the documentation of the methodology Duncan utilized is somewhat unclear and devoid of the details necessary to fully understand the testing results or to recreate the models, the models appear valid. In terms of testing, the models improved each time after being tested with internal data, revised, and again tested with external data. Certainly for the Mon/Fayette models, and to a lesser degree for the CSVT model, these results suggest a high degree of utility in the planning and route alterative selection process. For field scoping and survey, the Mon/Fayette models would be very applicable, but the CSVT would suffer from such a large site-likely area. Considering that the goals of these projects were geared more towards planning, this methodology produces effective models that perform well.

Katz, Gregory M., John P. Branigan, Paul W. Schopp, and Steven J. Biondo
2002 S.R. 0228, Section 290 Cranberry, Adams, and Middlesex Townships, Butler County, Marshall, Pine, and Richland Townships, Allegheny County, Pennsylvania. Volume 1, E.R. 1999-6127-019-H. Prepared for; Pennsylvania Department of Transportation, Harrisburg, Pennsylvania. A. D. Marble & Company, Conshohocken, Pennsylvania.

Region:

Butler and Armstrong Counties, Pennsylvania; Pittsburgh Low Plateau Section of the Appalachian Plateau Physiographic Province

Significance:

The significance of this model is not in in the model itself or the methods used to achieve it, but instead in the fact that Katz et al. reapply earlier APM models to their study area. These older models offer Katz et al. a baseline for understanding performance and test out potential variables. Additionally, these older models can be revaluated based on the results of this reapplication on independent data.

Model Type:

Predominantly judgmental, with a non-statistical testing of site versus background correlation

Variables:

The variables selected for this model were based on the variables utilized in the Stewart and Kratzer (1989) and the Cowin (1980) model. These two models were applied to this project's study area and their results were evaluated. Additionally, a visual comparison of histograms for site and background locations was performed for each variable. The results of the previous models and histogram comparison informed Katz et al. and resulted in the selection of these variables (Table 39).

Table 39 - Variables from Katz et al. (2002)

Variable
Landforms (lowland flats, lowland slopes, upland flats, and upland slopes)
Cost distance to water (cost by slope)
Soil drainage
Slope
Soil Capacity
Solar Aspect

Model Methodology:

The purpose of this model was to aid in the preliminary design and constraints analysis for the PennDOT-funded S.R. 0228, Section 290 project in Butler and Allegheny Counties, Pennsylvania. This project area covered a linear distance of approximately 10.6 miles and a width of 6.0 miles, for a total area of approximately 23,912 acres.

As stated in Katz et al. (2002:51) the, "precontact-era site probability was calculated in a generalized approach that relied on the role of the environment in determining probable site locations." In this case, the "probabilities" were calculated through the summing of arbitrarily assigned weights to variable classes and factor weights to the overall variables. This methodology is relatively simple and repeated numerous times within the use of APM in Pennsylvania. In and of itself, it is not a new approach.

The steps Katz et al. took before constructing their model are of the most interest here. Two previously published settlement analyses—Cowin's (1980) settlement analysis entitled *Archaeological Survey in West Central Pennsylvania, Region VII*, and Stewart and Kratzer's (1989) *Prehistoric Site Locations on the Unglaciated Appalachian Plateau*, published in *Pennsylvania Archaeologist*—were formalized by Katz el al. and applied through a GIS to this study area. The variables (Table 7), methods, and results of the Stewart and Kratzer model are discussed above, and Table 40 below shows how Katz et al. adapted them. Table 41 lists the variable from the Cowin (1980) settlement analysis as adapted by Katz et al.

The factor weights assigned to each variable were arbitrarily assigned, but based in part on the model builders' understanding of the correlation of these variables to sites and background values. This understanding was gained though a visual comparison of the distribution and histograms for the variables: landform type, soil drainage, cost-distance to water, solar aspect, slope percent, and soil capability. For these comparisons, a random sample of 59 (60%) of the 98 known sites within the study area selected and compared to background values. Katz et al. segmented these variables into classes and assigned relative weights to each class, but the method to do so or actual weights are not documented in their report.

Table 40 - Variables from Stewart and Kratzer (1989) Model Formalized by Katz et al. (2002).

Variable (Stewart and Kratzer, 1989)	Factor Weight
Cost distance to water (cost by slope)	27.5%
Slope	27.5%
Proximity to upland areas	dropped from model
Saddles (presence/absence)	15.0%
Soil Drainage	15.0%
Drainage heads (presence/absence)	dropped from model

Table 41 - Variables from Cowin (1980) Model Formalized by Katz et al. (2002)

Variable (Cowin, 1980)	Factor Weight
Cost distance to water (cost by slope)	25%
Landforms (lowland flats, lowland slopes, upland flats, and upland slopes)	25%
Cost distance to stream junctions (cost by slope)	dropped from model
Cost distance to Native American trails (cost by slope)	15%
Benches and saddles (presence/absence)	15%
Soil Drainage	10%
Stream density - as a measure of biodiversity	10%

Following the formalization of the Cowin and Stewart and Kratzer settlement analyses, Katz et al. classified and weighted the models. These models were put into a GIS and applied by Katz to his study area. The results of the Cowin and Stewart and Kratzer models were evaluated based on their ability to correctly classify known sites within the study area. It is not disclosed whether these tests were conducted with all of the 98 known sites, or only with the 60% random sample. Katz et al. use a score of "Model Efficiency" (M.E.) that is simply the percentage of sites located in the site-likely area divided by the percentage of total study area classified as site-likely; or ($P_{m|s}$ / %Total M) in the notation used here. As will be discussed in the following section, this measure is similar to the K_g, but more difficult to interpret. Based on the reapplication of these models, Katz et al. assigned a M.E. score of 2.0 to the Cowin model and a M.E. of 1.7 for the Stewart and Kratzer model. From these scores and an interpretation of the variables used to construct them, Katz el al. developed a list of variables and weighting scheme for a new model (Table 42).

Table 42 - Variables from Katz et al. (2002) Model

Variable	Factor Weight
Landforms (lowland flats, lowland slopes, upland flats, and upland slopes)	36%
Cost distance to water (cost by slope)	18%
Soil drainage	18%
Slope	9%
Soil Capacity	9%
Solar Aspect	9%

Katz et al. applied this model to their study area in the same manner as the previous models. The variables were classified and weighted judgmentally. The weights of the classes within each variable are not disclosed by the authors. Within the GIS, these weights were summed for each grid cell to produce a total sensitivity value. From here, Katz et al. went beyond where many models stop and combined the prehistoric sensitivity model values with the historic period sensitivity values derived from a separate model, resulting in a combined sensitivity. An additional raster layer was created from aerial photographs and USDA soil maps in order to quantify ground disturbance. Ground disturbance values were then subtracted from the combined archaeological sensitivity values to derive a model that predicts locations with a heightened sensitivity for intact archaeological sites.

In the final step, Katz et al. tested the prehistoric sensitivity model (not the combined or total sensitivity models) against the remaining 40% sample (n=39) of known site locations in their study area. This sample was set aside and not considered during the model construction or evaluation of variable correlations. The model was divided into five classes consisting of high, moderately high, moderate, moderately low, and low sensitivity. The total range of sensitivity values or the manner in which they were classified into these sensitivity strata was not discussed. According to Katz et al., 85% of test sites were found within the combined high, moderately high, and moderate classes which occupied 72% of the project area. The moderately low and low classes contained 16% of the test sites in 28% of the study area. That the site sample percentages add to equal 101% is inherent in the Katz et al. report. The authors state that future refinements to the variables within the model and the weighting scheme may lead to improving the model efficiency.

At the end of their report, Katz et al. describe the results of a field reconnaissance of two areas of high archaeological sensitivity. Archaeologists identified a single prehistoric archaeological site in the form of a possible pre-contact mound.

Model Classification, Efficacy, and Performance:

Overall, there was a pretty significant difference between the performance of the two initial models based on prior settlement analysis and the final model Katz et al. created. The first two models performed reasonably well, albeit very biased toward completeness and with a large percent of false-positive classifications. On the other hand, the final model Katz et al. created suffered from a very high false-positive classification without the benefit of a very high true-positive classification. In order to compare the models based on the information provided in the report, two assumptions are made: that each site equals one cell in the model, and that the Cowin and Stewart and Kratzer models were tested with the same sample of 39 points used in the final model test.

Table 43 and Table 44 show the class assignments and classification of the model based on Cowin's (1980) settlement analysis of west central Pennsylvania. Katz et al. calculated a Model Efficiency score of M.E. = 2.0 (more precisely, 1.94); from the information presented in the report, a gain of $K_g = 0.484$. The 13% false-negative classification error is relatively low, while the more acceptable false-positive classification error is moderate at 45%. The moderate gain of K_g = 0.484 is being dragged down by a large site-likely model area. Overall, this model is balanced toward completeness, but for a planning tool would work pretty well

Table 43 - Model Class Assignments from Cowin (1980) Model Formalized by Katz et al. (2002)

Probabilities of Assignment

		Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.001 (34)	0 (5)	39	0.2%
	Absent (S')	0.449 (10724)	0.55 (13145)	23869	99.8%
	Total	10758	13150	23908	1
	%	45.0%	55.0%	1	

Table 44 - Model Classification Results from Cowin (1980) Model Formalized by Katz et al. (2002)

Conditional Probabilities

		Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.87	0.13	1.0
	Absent (S')	0.45	0.55	1.0

Table 45 and Table 46 show the class assignments and classification of the model based on Stewart and Kratzer's (1989) settlement analysis of the Unglaciated Appalachian Plateau physiographic section. Katz et al. calculated a Model Efficiency score of M.E. = 1.7 (more precisely, 1.67); from the information presented in the report, a gain of K_g = 0.402. A 3% false-negative classification percentage is quite low. Balanced against a 58% false-positive classification, this model is very well balanced toward completeness. While the overall area of the site-likely model is high at 58.3%, the total correct site classification (97%) is very high. Compared to the Cowin model, these results are very similar. Essentially, a 10% increase in the site-likely equated to 10% more sites and reduced the false-negative by the same amount. Overall, like the Cowin model, this model is balanced toward completeness, but for a planning tool would work well.

Table 45 - Model Class Assignments from Stewart and Kratzer (1989) Model Formalized by Katz et al. (2002)

Probabilities of Assignment

		Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.002 (38)	0 (1)	39	0.2%
	Absent (S')	0.581 (13897)	0.417 (9972)	23869	99.8%
	Total	13935	9973	23908	1
	%	58.3%	41.7%	1	

Table 46 - Model Classification Results from Stewart and Kratzer (1989) Model Formalized by Katz et al. (2002)

Conditional Probabilities

		Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.97	0.03	1.0
	Absent (S')	0.58	0.42	1.0

Table 47 and Table 48 show the class assignments and classification of the model Katz et al. developed based on the previous results. Katz et al. calculated a Model Efficiency score for the previous two models, but did not report this score for the final model. This model achieves a score of M.E. = 1.18 and a gain of K_g = 0.150. While the classification false-negative is moderately low at 15%, the very high false-positive percentage impairs this model's ability to predict the likely location of archaeological sites.

Compared to the Cowin and Stewart and Kratzer models, these results are quite poor. The Cowin and Stewart and Kratzer models had nearly equal or much better correct site classifications, but within site-likely areas of almost half to three-quarters the size of the Katz et al. site likely area, respectively. It is difficult to discern where the final model departed drastically from the previous models, but it would seem somewhat unlikely that it was because of the choice of variables. With many variables in common between the models, it may have been the weighting or balance of cut-points for the model classification that sent them in different directions. A more remote but potential source of difference may be in the way Katz et al. reported their findings and the way they are interpreted here. The assumptions posed earlier were necessary to squeeze this assessment out of their report, but should not have led to such differences in the results. Overall, it appears that the final model Katz et al. produced performed much worse than the sample models initially created.

Table 47 - Model Class Assignments from Katz et al. (2002)

		Probabilities of Assignment Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.001 (33)	0 (6)	39	0.2%
	Absent (S')	0.718 (17170)	0.28 (6699)	23869	99.8%
	Total	17203	6705	23908	1
	%	72.0%	28.0%	1	

Table 48 - Model Classification Results from Katz et al. (2002)

		Conditional Probabilities Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	0.85	0.15	1.0
	Absent (S')	0.72	0.28	1.0

Assessment:

The three models Katz et al. created offer a very interesting opportunity to evaluate different models built from and tested with the same data. Further, this study allowed for the retesting of two models created for different data sets and environments. What we learn from this latter point is that the successful implementation of the Cowin and Stewart and Kratzer models by Katz et al. suggests that the variables and weights may be broadly applicable throughout western Pennsylvania. The high correct classification of these models, with accompanying larger site-likely areas, may signal that these models hit on some of the basic variables that influence settlement throughout upland regions. The large false-positive classifications make it clear that there are additional variables that need consideration, but the basic variables seem to be present. The utility of simple judgmental weighting of basic variables is verified in these results.

Aside from the poor performance of the final model created by Katz et al., there are additional topics within this study that can be reviewed. First, the Model Efficiency (M.E.) measure of a model's performance is not a very accurate or easily interpreted metric. This measure, referred to by Verhagen (2009) as the indicative value, can range from 0 to 100. However, the issue with interpreting this result is that the difference scale is very much non-linear. For example, if 70% of the sites were found in 1% of the area, the M.E. is 70.00; for 10% of the area, the M.E. is 7.00; and for 20% of the area, the M.E. drops to 3.50. This creates a distribution that approximates a power law. With such a dramatic difference in the M.E. resulting from relatively small changes in the percent of sites or site-likely areas, this measure is volatile. Interpreting the different between two M.E. values can be difficult based on this volatility. On the other hand, the Kvamme gain statistic (K_g) is basically linear (R^2 = 0.995 across the 70% of sites distribution). Therefore, the difference between two measures is constant and more easily comparable.

THE REMAINING MODELS

The models referenced above were chosen for evaluation because they each exhibited an innovative or original methodology, published enough data to evaluate their outcomes, or served as reference points for models that followed. Within the full sample of APM reports collected for this study, most models did not qualify for detailed evaluation due to repetition of methods already evaluated or, most commonly, absence of adequate methodological detail or testing data to assess their performance. For the goals of this study, a model without any testing or results is of no utility. What this report seeks to understand is which methods, variables, and techniques produced the best results and more accurately predicted locations with a high sensitivity for the presence of prehistoric archaeological material. Below, the reports that were not able to be fully evaluated are briefly discussed in order to understand the full breadth of this APM sample. This discussion will be organized by model type and reflect the potential within each model.

Qualitative Models

Two models fit this category: Becher et al. (1997) and Polglase (1997). Both of these models were constructed for the early planning stages of pipeline projects in western Pennsylvania. Similarly, each model utilized the same small list of basic variables including slope, distance to water, and landform type. Additionally, Becher et al. looked at the types of water sources, but this variable had no bearing on the final associations. The models that resulted from these projects associated very similar qualities with high archaeological sensitivity. These qualities focused the high sensitivity toward areas near water, on level ground, and preferably on floodplains, terraces, benches, or near saddles and gaps.

With no published internal or external test results, evaluating the utility of these models is not possible. The method of these models was quite simple, but likely served their purposes. The selection of variables was equally basic, but very much in line with the findings of regional surveys and settlement pattern analyses of that time (e.g., Cowin 1980). Variable combinations similar to these were tested by Stewart and Kratzer (1989) with limited success, and again with greater success by Katz et al. (2002) while retesting Stewart and Kratzer's model.

Associative and Composite Models

As a very common model type, there are a number of reports of this type that could not be fully evaluated. The majority of these included models created for planning purposes that did not include any internal testing or field checking. The methods utilized within this group ranged from simple ad hoc models of limited variables to more complex attempts at geomorphological analysis and ecological community modeling.

The simplest models of this group included pipeline studies by McIntyre (2009) and Reinbold (2010) and Means's (1998) model for the Meyersdale Bypass project. All of these projects created simple models that associated high sensitivity with low slope on or near landforms such as floodplains, benches, and saddles. McIntyre (2009) and Reinbold (2010) added the variable of wetlands and soils. Neither report provided results that can be evaluated. Blades et al. (2007) created a model for the Deer Creek watershed that incorporated many of these same variables and used the associative method. However, Blades et al. incorporated a few attempts at statistical testing to gauge the effects of model variables. While it is arguable whether their use of the Analysis of Variance was the appropriate statistical test with this data set, Blades et al. did test the ability of cost distance variables to "explain" the variability in settlement location relative to different prehistoric time periods. This is an interesting approach and may have good utility. Two additional models of very similar methods and variables were created by A.D. Marble (2003) and Duncan (2002). The model from A.D. Marble did not provide test results. The model by Duncan was field tested, but no sites were identified.

Two models documented in Baublitz et al. (2003) and Baublitz and Shaffer (2004) applied within Centre, Clearfield, and Jefferson Counties, Pennsylvania, were completed with a very similar approach. The authors utilized the variables of slope, distance to third order streams, soil drainage, soil type, distance to confluences, distance to wetlands, distance to quarries, and previous disturbance. As with other associative models, Baublitz and his colleagues ranked the variables based on judgment and previous studies. The more interesting aspect of this methodology is the use of negative weights for the variables that were considered to indicate the highest sensitivity (slope, distance to third order streams, and soil drainage). If any quadrant of the model exhibited a low value for any one of these variables, the entire quadrant was considered low sensitivity regardless of the other variables. In the same vein, if the variable of previous disturbance was high, all other variables were disregarded and the sensitivity was set to low.

Perazio (1995) created an interesting and potentially well performing model for a Pocono area study for the Bushkill Road School Complex. Perazio used a similar group of variables as the previous models, including slope, distance to water, soil drainage, and aspect. Perazio also used elevation above nearest water and soil, and plant and animal habitat as variables. A binary weighting scheme of high sensitivity (1) and low sensitivity (2) was employed and led to a range of sensitivity scores from 7 to 13. Unlike many of the reports here, Perazio did provide data from a field survey (Mooney et al. 2003) that allowed for some assessment of performance. However, the sample of sites was small and the initial sampling strategy was very biased toward high sensitivity quadrants. Perazio revised the strategy and resampled with less bias, but the site sample was still too small to make quantitative judgments.

The final two models in this group utilized methodologies quite different from the previous models (and each other), but still based the assessment of sensitivity on the association of other factors. For the Susquehanna Beltway project, Lawrence et al. (2002) created what is considered the most "deductive" model in this study. The intention of this model was to side-step the bias inherent in known site samples and follow a method that modeled plant and animal communities assumed to be important to prehistoric life and composite these resources into an overall model of "maximum habitat overlap." Lawrence et al. utilized the work of Versaggi to classify resource availability into units called MSR (Multiple Resource Seasons). As a basic description, each MSR was modeled as a collection of sub-models such as medicinal plant species distribution, white-tailed deer yards, beaver habitat, wind protection, soil fertility, and presence of lithic resources. Each MSR contained a unique mix of variables pertinent to that season and the final model was a composite of the MSR. The lack of reliable PASS site information for the study area, a prime reason for the creative model methodology, did not allow for testing of the model results. However, Lawrence et al. pointed out that a known prehistoric site that contained hearth features and a second site identified in Phase I testing were located in an area modeled as high sensitivity for season 1 macroband base camps. While the methods were very well researched and interesting, repetition of this model in a more well-surveyed area would help confirm or refute the results.

The final model of this type was a small model created by Yamin et al. (2010) for a one-block area within Old City, Philadelphia. With a lack of detailed site location data and an undeniable history of disturbance in these environs, a different approach to modeling sensitivity was needed. The basis of this model was a geomorphological recreation of the Dock Creek stream valley based on comparison to other coastal streams, historic documentation of stream conditions and alterations, borehole data, and an early nineteenth-century topographic survey. The theory of this model was to recreate the shape of the surrounding land prior to historic development and then trace the possible alterations, both cutting and filling, to this landscape throughout history to assess the potential for intact ground surfaces. The stratigraphic model was then put into the context of known prehistoric sites and contact-period documentation to better understand site potential. While no testing was done at this location, the original approach and use of a wide variety of data sources is quite interesting.

Correlative with and without Background Testing Models

The final grouping of the remaining models is those that included attempts to measure the correlation of site locations to specific variables within the study area. Most of these models were interesting and well done, but did not provide survey results or the necessary information to interpret model performance.

The earliest of these models was Johnson et al. (1989), with a model for a Monongahela cultural resources inventory. This relatively simple model was part of a larger study of the Monongahela. Essentially, this model attempted to formalize many of the environmental variables often cited in regional studies of the Monongahela. The initial set of variables investigated were topographic landforms, drainage divides, number of frost-free days, proximity to Native American trails, and soil type. Johnson et al. quantified and tabulated the relationship of known Monongahela sites to these variables and produced a model that was much more of a settlement analysis than an actual sensitivity assessment. However, the interesting aspect to this study was the authors' recognition that drainage divides and Native American trails were highly correlated to each other. What Johnson et al. describe is the property of multicollinearity. Noting that these two variables were strongly correlated to each other has implications for the introduction of bias and the overall validity of testing using those variables. This is a very important concept within any form of spatial analysis.

A settlement analysis created for the 202 Bypass in eastern Pennsylvania created by Diamanti et al. (1993) utilized correlation of variables to site locations. Diamanti et al.'s study did not result in an actual predictive model or sensitivity assessment, but it did develop a well-constructed associative framework that considered site types and variables. Believing that PASS site locations were too biased to serve as a reliable data source, Diamanti et al. utilized information only from sites identified through systematic survey. Along with these sites and an extensive review of regional literature and settlement analyses, the authors created a framework of site type expectations relative to variables such as topographic landforms, slope, soil drainage, soil productivity, and distance to water. Tables were then made that correlated site types and variables. The framework was modified based on the empirical evidence. In and of itself, this framework stands alone as a very useful piece of research combining deductive theory and empirical observations. However, when applied to a 10% field survey of the 202 Bypass project, the results were not sufficient to evaluate this model. With a total of two identified prehistoric sites, both within moderate sensitivity areas, the test data set was too small to draw conclusions.

The final correlative without testing model was documented in Coppock and Heberling (2001) and Coppock et al. (2003) for the U.S. 219 projects in Somerset County, Pennsylvania. The model from 2001 utilized a small list of variables including slope, distance to water, and site type. These information classes were taken from information within the PASS database. An initial model was created based on landform sensitivity assessments from Dr. Frank Vento and combined with information gained from correlating sites to distance to water and slope. Following a Phase I field survey, this model was tested with 50 previously undocumented site locations. The results of this test indicated that the classification of both slope and distance to water should be adjusted for a better fit. Once these variables were adjusted, the resulting revised model correctly classified 86% of the site sample in high sensitivity. Unfortunately, the actual size of the high sensitivity area was not disclosed, therefore the gain statistic and classification errors could not be calculated.

The model by Coppock et al. (2003) was an expansion of the previous model that used the variables of cost distance to water, slope, distance to Native American trails, soil drainage, landform type, bedrock type, site type, and disturbances. While the performance of this model cannot be assessed without results, two interesting methodological outcomes should be noted. First, whereas most models in this study consciously ignored the potential for rockshelters, Coppock et al. made a parallel model that utilized the variables of bedrock type, slope, and cost distance to water. This model was made specifically to assess the sensitivity for the presence of rockshelters. Secondly, Coppock et al. used the correlation of sites to variables to create both relative weights within variable classes and factor weights for each variable. In descending order, Coppock et al. ranked cost distance to water, percent slope, soil drainage, distance to Native American trails, and the presence of a floodplain or terrace as the variables with the largest contribution to overall sensitivity. The model was created by multiplying the intra-variable class weights by the overall class weight to get a total weighted value for each class. This method was used a number of times throughout the reports in this study.

Last, unlike the previous correlative type models, the final report of this type used the correlative technique and incorporated background data testing. A model by Glenn (2010) for the Erie National Wildlife Refuge, Crawford County, Pennsylvania, considered a number of variables at the onset. These variables included elevation, slope, aspect, solar insolation, distance to streams, distance to confluences, distance to prime farmland, cost distance to streams, confluences, prime farmland, hydric soils, and distance to Native American trails. The method of correlating these variables to site locations was simply the use of counts of known sites per variable class. A visual inspection of the tables was used to interpret which variables had the best ability to differentiate site locations. From this, Glen retained five variables for the model: slope, cost distance to confluence, cost distance to streams, cost distance to prime farmland, and hydric soils. Glenn then used a proportional weighting scheme that assigned sensitivity weights to each variable class proportional to the percent of known sites counted within that class. The final model was a sum of the weights.

4. RESULTS OF MODEL EVALUATIONS

The evaluations presented above cover a wide range of model types, methodologies, and outcomes. Through this, a number of observations concerning good and not-so-good APM practices become apparent. This section will synthesize the results and findings relative to what made for successful models and what led to poorly performing models. Three areas of the modeling process will be discussed: model reporting, variable selection, and methodology.

REPORTING

Second only to models not being validated or tested, the lack in reporting of key quantities such as quadrants of all sensitivity classes or percent of area surveyed, was the most common barrier to evaluating model results. In many cases methodology, goals, and theory were clearly stated and elaborated on well enough to understand the approach and expectations. However, the results of tests and validations, if conducted, were not as frequently reported with such clarity. That being said, a number of reports did an exemplary job of presenting all the data necessary to fully interpret the results. The reports by Duncan, for example, each included succinct tables documenting model classifications for each test and revision.

Also problematic is that many of the reports that utilized the weighted methods (associative, judgmental, and correlative) did not adequately define the way in which variables were classified, the class breaks, or the weights of each class. Without this information, models cannot be repeated, and interpreting the contribution of each variable is very difficult. However, reports such as Bailey and Dekin (1980), Neusius and Neusius (1989), and Coppock et al. (2003) provided ample information for anyone to recreate their models.

Finally, a number of the models in this study were bare in their documentation of model creation methodology. In the simpler weighted sum models, the lack of details generally concerned the methods by which variables were classified and weighted. The end result generated through summing these weights did not require much explanation. In the models with more involved methods, correlation statistics, or regression techniques, knowing the methods is critical to understanding the assumptions on which the model was built. Without understanding these assumptions, the model results become removed from context and cannot be correctly interpreted. Additionally, without clearly defined methods, a model cannot be recreated for validation or use in new areas.

VARIABLES

Within the APM models evaluated here, the variables chosen to contribute to archaeological site sensitivity or aid in the explanation of variance within known settlement locations were decidedly environmental. This is to be expected given the wide availability of environmental data through USGS quadrangle maps and publications and USDA soil surveys. Attempts to explain the variation in site location through cultural variables were very few. The most common attempt to use cultural input was in the form of historical Native American trails as documented by Wallace (1965). This variable was used in a number of studies. Although it has not been shown to strongly correlate with site locations, Johnson et al. (1989) have suggested that it is highly correlated with drainage divides and other environmental variables. Lawrence et al.'s (2002) used of "deductive" variables such as deer yards, beaver habitat, plant species distribution, and ecological variables was an attempt to sidestep the most common environmental variables for those that may be more able to explain habitation location choices. While still environmental in character, these variables were intended to seek a new dimension of cultural choice in favor of the more typical environmental constraints.

Table 49 is a generalized list of the most commonly used variables within the APM reports reviewed here. Most every model included two variables pertinent to topographic slope and some measure of access to water. Following these most basic attributes were topographic landform, often categorized as floodplain, terrace, bench, saddle, etc., and some measure of soil characteristics. Soils were measured by drainage, depth, texture, type, or suitability. Together, these four variables—slope, water, landform, and soil—were the most commonly used within the models studied here. These four are by no coincidence the simplest to generate from the base data of USGS maps, digital elevation models (DEM), and stream coverage. Following these in usage were a series of distance and cost distance variables measuring various aspects of hydrology, bedrock, and proximity to Native American trails. These reflect some of the many permutations of measurements of hydrology, the use of cost distance with slope as the cost, and additional data sources such as the USGS and Wallace (1965). Less frequently used are the variables of topographic aspect, hydrology network variables such as confluences and stream rank, solar insolation, and topographic relief. Aspect, relief, and solar insolation are all based on the DEM and slope and can be strongly correlated to each other. Finally, the list in Table 49 contains a few attempts to use a variety of cost distance measures and more variations of hydrology along with absolute elevation.

Table 49 - Relative Percentage of Variables Used within Reports in this Evaluation

Variable Class	Relative percent of use in report sample
Soil Characteristics (Drainage, Texture, Type)	75%
Topographic Slope	75%
Topographic Landform	70%
Elevation above Water	40%
Cost Distance to Confluence	30%
Cost Distance to Streams and Water Bodies	30%
Distance to Streams and Water Bodies	30%
Surface Geology	30%
Cost Distance to Native American Trails	20%
Depth to Bedrock	20%
Distance to Native American Trails	20%
Stream Order	20%
Topographic Aspect	20%
Wildlife Suitability	20%
Cost Distance to Topographic Saddles	15%
Distance to Drainage Divides	15%
Distance to Steam Confluence	15%
Number of Frost Free Days	15%
Solar Insolation	15%
Local Topographic Relief	15%
Cost Distance to Drainage Divides	5%
Cost Distance to Headwaters	5%
Cost Distance to Lithic Sources	5%
Cost Distance to Peaks	5%
Cost Distance to Vantage Point	5%
Cost Distance to Wetlands	5%
Depth to Ground Water	5%
Elevation	5%
Flood Frequency	5%
Local Stream Density	5%

Determining which of these variables are the most successful at predicting archaeological site locations in these studies is quite difficult. The reason for this is that the model types of qualitative, associated, judgmental, and correlative without testing do not have a mechanism for determining how background data influence the variables or how each variable contributed to the overall success or failure of a model. Unfortunately, these model types make up the majority of the models in this study. The only evidence for which variables were the most successful in assessing sensitivity is based on the model builders' assignments of weights. However, these assignments are generally always based on judgment. While these judgmental weights may be useful, without testing against chance alone, we can only assume that they reinforce the circularity of the bias in know site locations. Further, the very frequent use of the summed weights methodology does demonstrate which variables, or perhaps more importantly the relationship between variables, that contribute to each rank of sensitivity. This is because the weights of each variable are summed into a single measure that can be attained through a number of different variable permutations.

Only studies that utilized some form of statistical test for sites and background data, such as the K-S test, or used a regression methodology, contain the mechanisms to determine which variables contributed to the model results. Table 50 lists the environmental variables that show a statically significant difference in distributions at archaeological site locations as compared to background locations. Some studies in this table are included because they incorporated this step, but did not publish documentation of their findings. Other studies tested within the classes of variables, while others only tested the variables themselves. Finally, some used the K-S test while others used the Pearson's Chi-Squared test and the results were variously reported as test statistics or p-values. While some interpretation of the results was necessary, Table 50 indicates the most likely variables that have been found to be statistically significant in discriminating archaeological site locations from background values. These results are only relevant within the physiographic setting within which the study was applied. As is expected, the variables pertaining to slope, soil characteristics, availability of water, and landform are the most common.

Table 50 - Significant Variables from Studies with Background Testing

Citation	Model Type	Significant Variables	Significant Variable Classes
Miller (2002) and Miller and Kodlick (2006)	Correlative with background testing	Soil types	Poorly drained floodplains
			Slopes/ridges, moderate fertility
			Water/frequently flooded soils
		Aspect	Southeast aspect
			West aspect
		Slope	0 - 3% slope
			3 - 8% slope
		Distance to water	300 - 375 meters
			375 - 450 meters
			450 - 525 meters
Whitley and	Correlative	Elevation	501 - 750 feet

Citation	Model Type	Significant Variables	Significant Variable Classes
Bastianini (1992)	with background testing	Slope	0 - 5%
		Landform	Floodplain
		Distance to water	0 - 100 meters
		Distance to confluence	0 - 100 meters
Hart (1994)	Logistic Regression	Order of Nearest Stream	N/A
		Distance to nearest stream	N/A
		Elevation above nearest stream	N/A
		Elevation above nearest confluence	N/A
		Slope	N/A
		Soil Drainage	N/A
		Soil texture	N/A
		Depth to water	N/A
Duncan et al. (1996)	Correlative with background testing	Slope	N/A
		Cost distance to nearest trail	N/A
		Cost distance to river confluence	N/A
		Cost distance to river	N/A
		Local relief	N/A
		Cost distance to major tributary	N/A
		Distance to nearest spring	N/A
		Cost distance to nearest vantage point	N/A
		Cost distance to drainage divide	N/A
		Cost distance to localized peak	N/A
		Cost distance to nearest saddle	N/A
		Solar insolation	N/A
		Soil fertility	N/A
		Soil agricultural capability	N/A
		Soil suitability of wildlife	N/A
		Soil water capability	N/A
		Soil drainage	N/A
		Bedrock formation	N/A
Duncan and Schilling (1999a)	Correlative with background testing	Significance measures not published	N/A

Citation	Model Type	Significant Variables	Significant Variable Classes
Duncan and Schilling (1999b)	Correlative with background testing	Significance measures not published	N/A
Duncan et al. (1999)	Correlative with background testing	Significance measures not published	N/A

METHODS

The overall trends noted within this body of APM reports could be summarized by saying that most models are based on the summation of judgmental weights applied to a common group of untested environmental variables. The majority of models derived the weights, which were ultimately either multiplied by a factor or simply summed, through judgmental means. Further, these models were not frequently validated against internal or independent data and were reported with various degrees of methodological detail. Alternative methods included the assigning of weights proportional to site percentages in a class (Glenn 2010) or weights derived from the difference between site and background percentages (Whitley and Bastianini 1992) by regression coefficients (Hart 1994) or through associative and composite models (Lawrence et al. 2002; Yamin et al. 2010). However, this generalized statement masks the many interesting techniques, variables, and discussions that were contained within these reports.

Interestingly, it does not appear that any one modeling method is more successful than any other. Each different methodological approach is shown to be capable of producing a model that was efficient and performed well relative to its goals. Both simple models with few variables and more complex models with many variables had model performance results that seem to be about equal. Given the difference in how many times each method was practiced, differences in validation techniques, and variety of reporting standards, the different effect of each method cannot be teased out of the performance results detailed above. However, there are other methodological issues that set these model types apart.

Qualitative and associative models, despite their simplicity, are capable of producing useful results. To their advantage, these models are intuitive, broadly applicable, and rely on research as opposed to site samples. Their negative qualities are that the research from which they are built may be biased and may not be applicable to all the variables of a study area. Further, it is difficult to assess what variables contributed to the pattern or if the results are meaningful.

The two composite models of this evaluation did not contain any validation of their results and their success is unknown at this time. The methods employed in the models used very different data sources and sought to understand very different dynamics, but both approached their results by creating explanatory models as the source of sensitivity. The advantage of this is that the confidence of the model can be judged based on the quality of the research from which it is based and the inclusion of an explanatory mechanism. However, the negative for such an approach is the availability of the data from which the explanatory models are built and the fact that the variance of each explanatory model is carried over as they are composited into the final model. This more "deductive" approach may carry a much greater list of assumptions than other types of models.

Judgmental model methods produced many of the successful models in this study. This likely has much to do with the large number of studies that used this technique. Like qualitative and associative models, the judgmental models are generally simple, easy to understand, weighted based on research, and do not require known sites in the study area. Also like qualitative and judgmental models, it is difficult to assess why a model works or why it fails. Revising judgmental models is done mainly through adjusting weights with no real feedback until the model is tested again. In this sense, this method searches for the best solution without guidance from the results. Therefore this method can be quite inefficient if the modeler is seeking to find the best fit.

The correlative methods are only slightly more complex than the judgmental methods with the added requirement of creating histograms or statistical testing. The correlative model without background testing builds on the previous models by understanding which variables are more strongly associated with site locations, as opposed to relying on experience and research. However, without background testing, it is difficult to know how the results compare to a model by chance. The element of background testing greatly improves the ability of a model to identify those variables that can discriminate site locations from background values. In theory, this should remove much of the background "noise" from the pattern and allow for an understanding of which variables contribute most significantly to the variation.

Finally, the regression approach was used to success by Hart (1994) and incorporated into the methods of Duncan and Schilling (1999a, 1999b). As opposed to the previous model types, which are mostly mathematical models in that variable weights are summed to create an index of sensitivity, the regression approach is a statistical model that uses site presence/absence to estimate probability. This difference makes the regression model somewhat more difficult to create, much more reliant on existing data, and saddled with a different set of assumptions. The negatives include the requirement of software (and the knowledge to use it) and a data set that does not drastically violate the assumptions of the regression method. The benefits of this approach are that the relationship between variables can be modeled; the test results contribute to the understanding of the strength and effect of each variable on the outcome; predictions are based more on empirical evidence and less on judgment; and each model can be compared and adjusted based on a number of metrics. As with all of these models types, however, in some situations the strengths are indeed weakness and vice versa. The available data, environmental character, and goal of each modeling situation will determine which model is best suited.

5. RECOMMENDATIONS AND CONCLUSIONS

The review of 32 archaeological predictive model studies from Pennsylvania has created a context within which to understand the utility of many different modeling approaches, techniques, and variables. This body of reports contains examples of success and failure, simple to complex model methods, rigorous methods to ad hoc applications, and documentation ranging from the detailed to the abstract. The challenge in this review was finding a core contribution from each study, despite any given report exhibiting a number of the opposing characteristics listed above. No one single study stood out as the best (or worst) example of how to create an archaeological predictive model. While some were certainly better than others, each had positive and negative aspects and were created within a number of constraints including scope, budget, site data quality, background data quality, computing power, software, map resolution, and model approach.

Four very clear conclusions can be reached based on this study: 1) each study area and situation has constraints that call for flexibility, creativity, and explicit assumptions when choosing a modeling method; 2) model variables may range from a simple few to numerous permutations, but they must be tested against background values and preferably assessed for multicollinearity and spatial autocorrelation and evaluated through a stepwise method; 3) models must be tested at a minimum against internal data, preferably against independent data, and preferably against numerous data sets to understand the model's stability—otherwise the success of a model cannot be evaluated; and 4) the reporting of any model must be done thoroughly and with as much detail as possible, including variables, weights, calculations, and most importantly, assumptions and test results. The guidelines and recommendations below detail how these conclusions can be implemented in this statewide predictive model set.

GUIDELINES FOR REPORTING

The reporting of model goals, assumptions, methods, and results should be clear and concise. In order for a model to be understood or reapplied to a new study area a reader must be able to understand why the model was created in a particular way and how the independent variables were combined to create the resulting sensitivity. As was the case with many models in this study, the lack of key details led to the inability to interpret model results and judge effectiveness. At a minimum, specific documentation should include the following: model goals; theoretical orientation and justification; variables selected for evaluation; variables accepted for the model and how they were tested; modeling steps and details such as weighting schemes or regression metrics depending on the model type; the percentages of the study area covered by each sensitivity class and the classification of site and non-site areas into each class; the evaluation of the findings; and the assumptions and limitations that guide the implementation of the final model. These guidelines will be implemented in this study through the consistent and thorough presentation of methodology, validations, and results. The inclusion of these data will allow the models from various regions to be compared, evaluated, and recreated with new data.

GUIDELINES FOR MODEL VARIABLES

As shown by Duncan and Schilling (1999a, 1999b), the use of variables does not have to be limited to the basics of slope, access to water, and landforms. While these more basic variables may produce the best model in some circumstances, there may be other more important variables in other circumstances. Duncan's approach of creating many permutations and alternate variables is time and data intensive, but beneficial when using a correlation with testing or regression approach. In these approaches, many variables can be screened to assess which best differentiates the site location pattern from the environmental background. Having numerous variables to test gives flexibility to find aspects of the pattern that may not be obvious or previously considered. With regression approaches, a stepwise method can be used to produce a model with the best fit from a selection of numerous variables, or other model metrics can help determine which variables contribute the most. Additionally, variable relationships can be modeled in numerous ways. Possible complications to this approach are that using too many variables may add noise to the model and that certain variables may be strongly correlated to each other (multicollinearity) and affect the model outcome.

On the other hand, simple models created for expedient scoping or in areas of limited data availability are likely to be best served by limiting the number of variables. In these instances, an understanding of the environment is key. Often, access to water, slope, landform, and soil type will serve as a solid foundation for a judgmental model, but if the environment appears to have more exaggerated distributions of these basic variables an alternative may be necessary. For example, attempting to model site sensitivity in an area covered in wetlands will greatly limit the usefulness of access to water as a variable. Research may show that sites are often found on small, dry, raised landforms, in which case a variable such as local topographic relief in conjunction with soil type would be the most useful. Without testing variables against the background, it is best to create simple models and then introduce variables one at a time and compare the results against the previous model using the same independent test sample.

To implement these guidelines in this project, the number and complexity of variables will be predicated on tested correlations and the type of model to be employed in each area. A reference database of variables that cover the entire state will be created, from which variables appropriate to a given model type and area can be selected. The database will include primary variables such as elevation, hydrology, and geology, as well as secondary variables such as cost-distance and measures of topography (e.g., roughness and topographic position index). The secondary variables will be computed for various neighborhood sizes, such as 30 m, 100 m, 1 km, etc. The first step in variable selection for any given area will be the use of statistical tests, such as the K-S test, to define each variable's usefulness in discriminating site locations from background values. Those variables that do not discriminate better than chance will be excluded from the model. Those variables that do successfully discriminate locations will be considered further in the model building and testing phases.

For models in areas of low data quality, a simple model may be employed in which a bottom-up approach will be taken. In this approach, a few basic model variables will be utilized for an initial model, then additional variables will be added and tested to assess multicollinearity. If a variable is not strongly correlated to another and increases the ability of the model to predict site locations, it will be retained. Conversely, in areas where data quality is good, more complex models stemming from the regression family of methods may be used. In these models, a top-down approach to variable selection may be employed. In this case, many variables will be utilized, and the final model will be selected by the combination of variables that achieve the best fit and balance of efficiency and completeness.

GUIDELINES FOR MODELING METHODS

There is no one best modeling methodology or model type. What this study shows is that each situation has different constraints that help inform which model type(s) may work and which may not. That being said, some of these model types are better able to give the feedback necessary to understand if they are an appropriate approach, while other methods do not.

Qualitative and associative model types were popular earlier in the history of APM due in part to the lack of digital environmental data, computers, and software. These models were also popular because a number of regional survey programs in the preceding years had generated a large body of synthetic and site/settlement data. These models were often the first attempts to formalize the survey findings into models that could be used to focus field efforts and planning. However, these models were very much based on survey data that may or may not have been collected systematically, and the model builders did not have many mechanisms for assessing the meaning of the survey results. Situations that may be conducive to the application of these models include areas with few to no known sites, little if any research, or situations where the goal of the study is to develop a framework of sensitive environmental associations and site expectations that can be tested and refined with further survey. In most situations where the outcome is intended to be a continuous surface of quantified site location sensitivity, this model type is not likely to be the most appropriate choice.

Judgmentally weighted models are the most common within this APM report study. That is likely because they are intuitively easy to understand, are based on research, and require less data and less statistical understanding than correlation or regression models. This model type can be applied quickly and modified to suit research hypotheses and different environments. However, these models are difficult to separate from the random environmental component because the attributes are not tested against chance. This model type is very useful in a study area that has very few or no known archaeological site locations or when the goal is to construct weighting schemes for a research purpose. In situations where a sample of site locations are known, it is recommended that some form of correlation and preferably testing against the background become part of the modeling methodology.

Correlative models are used throughout the chronology of reports studied here. These models offer a deeper understating of how archaeological sites are distributed relative to the environmental background and can be used to support or dispute hypotheses regarding site distribution relative to the environment. In particular, the correlative model type with testing allows for not only the characterization of archaeological sites relative to other sites within a variable, but also relative to the environmental background. This characterization is critical if the goal of the study is to attempt to isolate portions of the landscape that are similar to the environmental pattern observed at known site locations. The correlative method is appropriate in any situation where a sizable sample of site locations is known or in an environment of the same background character as a well-surveyed area where settlement patterns are likely to be comparable. Any time the correlative method is used, it should always include some form of statistical testing of variables at site and background locations. If that is not possible, at a minimum a visual comparison of histograms from sites and background should be used.

Finally, regression methods, specifically logistic regression, was not used frequently within this study, but has been researched by numerous authors throughout the history of APM (see Judge and Sebastian 1988; Warren 1990). Helping to explain both of these observations is the fact that the family of regression methods requires an elevated understanding of statistics, is technically more difficult to apply, may provide more accurate models in some cases, and carries a certain cachet based on its complexity compared to the more mundane non-statistical methods. In spite of these qualities, these methods also carry numerous assumptions and data requirements, can be difficult to implement and understand, and may not produce better results than any of the previously mentioned model types. A successful situation for the use of the regression model type would include well surveyed regions with ample and representative known site locations, a well-defined pattern of landform usage relative to the background data, numerous variables that may be used to seek a reduction in model variance, and an understanding of the assumptions on which the particular model is built.

To paraphrase Kvamme (1988:327), the two most basic assumptions of archaeological predictive modeling are: 1) that site location decisions made by Native Americans were not made randomly relative to environmental variables; and 2) that this pattern is represented in known sites and can therefore be extrapolated to large areas. The methods that we choose to employ are framed largely by the degree to which we are able recognize this pattern from a known site sample and the ability of our chosen variables to serve as proxy for those variables that Native Americans considered in their settlement decisions. The studies discussed here have shown a wide range of ability to successfully approximate this pattern and extrapolate it throughout a study area. Some failures were due to methodological choices, others were due to our lack of understanding of settlement systems, and yet others to the lack of a strong settlement pattern relative to the environment. These are only three of the innumerable unknowns that impinge upon our attempts to predict the cumulative effects of thousands of years of individuals and groups making an intractable constellation of decisions. However, the results from this study inform us that it is possible to predict site locations better than chance alone, that the methods used here and others beyond this sample are capable given different situations, and that these methods can contribute to the goals of efficient project planning and resource management.

For this statewide predictive model set, numerous modeling methods will be employed in order to get the best-fit model. This review has demonstrated that there is no one type of model that works in every region. Further, it shows us that both simple and complex models can be equally successful (or less than successful) given different data qualities and rigor of implementation. Physically, Pennsylvania has very diverse physiographic terrains, hydrologic conditions, geology, and preservation. Archaeologically, the knowledge of site locations and attributes from region to region varies greatly, from very well surveyed to completely unsurveyed. Finally, the underlying cultural and individual decisions that created the archaeological record are the most complex variables of all. This project seeks to utilize the many different forms of APM to tailor each model to the physical and data realities of each region. This approach will provide the best-fit model given available data as opposed to a consistent, statewide modeling method regardless of how well the data are suited to it.

At a minimum, the most basic models of this study will use the correlative method with background testing for environmental variables. This method, used by Whitley and Bastianini (1992) and Duncan and Shilling (1999a and 1999b), utilizes weighting schemes that are summed to represent the overall sensitivity of different combinations of variables. The pros and cons of this method are detailed in earlier chapters, but overall this method can produce successful models that are easily understood. This type of model will be used in areas with limited data quality. In areas with moderate to high data quality, models from the regression family will be attempted. These models include the logistic regression type as used by Hart (1994) and developed by Kvamme (1998). Additionally, more complex regression methods, such as multivariate adaptive recursive splines (or “MARS”) will be attempted in areas of high data quality. Models of this type have not been attempted previously in Pennsylvania and are only currently documented in studies within the ecological field (Friedman 1991; Munoz and Felicisimo 2004). They do, however, offer an alternative method that overcomes some of the issues with other regression techniques. In the end, numerous approaches will be attempted to best describe and model the pattern of known archaeological site locations that is unique to each and every watershed in the state.

6. REFERENCES CITED

A.D. Marble & Company
2003 S.R. 1056, Section 001 Athens Bridge Replacement Project Athens Township, Bradford County, Pennsylvania. Volume 1. E.R. 2000-8029-015-R. Prepared for Dewberry-Goodkind, Inc., Carlisle, Pennsylvania, and the Pennsylvania Department of Transportation, Engineering District 3-0, Montoursville, Pennsylvania. A. D. Marble & Company, Conshohocken, Pennsylvania.

Altschul, Jeffery
1988 Models and the Modeling Process. In *Quantifying the Present and Predicting the Past*, edited by W. Judge and L. Sebastian, pp. 325-428. U.S. Government Printing Office, Washington, D.C.

Bailey, Douglas L., and Albert A. Dekin
1980 A Survey of Archaeology, Bi Story and Cultural Resources in the Upper Delaware National Scenic and Recreational River, Pennsylvania and New York States. E.R. 1981-0311-42-A. Prepared for United States Department of Interior National Park Service, Mid-Atlantic Region, Philadelphia, Pennsylvania. Public Archaeology Facility, Department of Anthropology, State University of New York, Binghamton, New York.

Baublitz, Richard T., and Barbara J. Shaffer
2004 Archeological Predictive Model South Central Centre County Transportation Study Centre County, Pennsylvania. E.R. 2000-8003-027-N. Prepared for Pennsylvania Department of Transportation, Engineering District 2-0, Clearfield, Pennsylvania. McCormick, Taylor & Associates, Inc., State College, Pennsylvania.

Baublitz, Richard T., Charles A. Richmond, and Barbara J. Shaffer
2003 Archaeological Predictive Model, S.R. 0830, Section 590, DuBois-Jefferson County Airport Access Project, Jefferson and Clearfield Counties, Pennsylvania. E.R. 1993-0231-065-M. Prepared for the Pennsylvania Department of Transportation, Engineering District 10-0, Indiana, Pennsylvania. McCormick, Taylor & Associates, Inc., Pittsburgh, Pennsylvania.

Bayard, Donn T.
1969 Science, Theory, and Reality in the "New Archaeology." *American Antiquity* 34(4):376-384.

Becher, Matthew E., Carol S. Weed, Mark S. Warner, and Rita G. Walsh
1997 Phase I Cultural Resources Investigations of Columbia Gas Transmission Corporation's Proposed Market Expansion Project: Artemas Storage A and B Line 29520 Loop in Mann, Southampton, Monroe Townships, and Bedford Counties, Pennsylvania. Volume 8. E.R. 1996-2683-009-B. Prepared for Columbia Gas Transmission Corporation, Charleston, West Virginia. Gray & Pape Inc., Cultural Resource Consultants, Cincinnati, Ohio.

Bettinger, Robert L.
1980 Explanatory/Predictive Models of Hunter-Gatherer Adaptation. *Advances in Archaeological Method and Theory* 3:189-255.

Blades, Brooke, Frank Vento, and David Brett
2007 Pennsylvania Archaeological Data Synthesis: Deer Creek Watershed (Watershed A of the Lower Allegheny River Sub Bbasin 18) Allegheny River Bridge Replacement, Pennsylvania Turnpike. Harmar Township, Allegheny County, Pennsylvania. E.R. 2004-0897-003-K. Prepared for Pennsylvania Turnpike Commission, Harrisburg, Pennsylvania. A.D. Marble and Company, Conshohocken, Pennsylvania.

Borillo, M.
1974 Construction of a Deductive Model by Simulation of a Traditional Archaeological Study. *American Antiquity* 39:243-252.

Chiarulli, Beverly A., Douglas C. Kellogg, Robert G. Kingsley, William J. Meyer, Jr., Patricia E. Miller, Philip A. Perazio, and Peter E. Siegel
2001 Prehistoric Settlement Patterns in Upland Settings: An Analysis of Site Data in a Sample of Exempted Watersheds. E.R 2001-R001-042-A. Prepared for Pennsylvania Historical and Museum Commission Bureau for Historic Preservation Grants Program, Harrisburg, Pennsylvania. KCI Technologies, Inc., Mechanicsburg, Pennsylvania.

Coppock, Gary F.
2009 Phase I Archaeological Survey, US 219 Improvement Project, Meyersdale to I-68, Somerset County, Pennsylvania and Garrett County, Maryland. E.R. 2002-8042-111-Q. Prepared for Pennsylvania Department of Transportation. Heberling Associates, Alexandria, Pennsylvania.

Coppock, Gary F., and Scott D. Heberling
2001 Predictive Model for Archaeological Resources, US 219 Improvements Project S.R. 6219, Section 020, Somerset County Pennsylvania. E.R. 2001-8012-111-C. Prepared for

the Pennsylvania Department of Transportation, District 9-0 and the Federal Highway Administration. Heberling Associates, Inc., Huntingdon, Pennsylvania

Coppock, Gary F., Scott D. Heberling, David A. Krilov, and Ronan A. Carthy
2003 Phase I Archaeological Survey U.S. 219 Improvement Project Meyersdale to I-68 Somerset County, Pennsylvania and Garrett County, Maryland. E.R. 2001-8012-111-C Prepared for the Pennsylvania Department of Transportation, District 9-0, The Maryland State Highway Administration, and The Federal Highway Administration. Heberling Associates, Inc., Alexandria, Pennsylvania and McCormick, Taylor and Associates, Inc., Philadelphia, Pennsylvania.

Corrie, Jean
1984 Predictive Archaeological Model Study, Third Street to Ferry Street Redevelopment Parcel, Easton, Northampton County, Pennsylvania. E.R. 1984-1641-095-B. Prepared for Urban Design Archaeological Grant. Lafayette College, Easton, Pennsylvania.

Cowin, Verna L.
1980 Archaeological Survey in West Central Pennsylvania, Region VII. Prepared for the Pennsylvania Historical and Museum Commission, Harrisburg, Pennsylvania.

Davis, Christine
1989 Archaeological Land Use History of the Pittsburgh Technology Center Site, Pittsburgh, Pennsylvania. E.R.1989-1053-003-A. Prepared for Urban Redevelopment Authority of Pittsburgh, Pittsburgh, Pennsylvania. Christine Davis and Associates, Inc., Verona, Pennsylvania.

Diamanti, Melissa
2006 Addendum to Phase l a Sampling Design for Urban Archaeological Resources Mon/Fayette Transportation Project S.R. 51 TO I-376 Section Allegheny County, Pennsylvania. E.R. 1987-1002-042-B86. Prepared for Pennsylvania Turnpike Commission, Harrisburg, Pennsylvania. Archaeological and Historical Consultants, Inc. Centre Hall, Pennsylvania.

Diamanti, Melissa, Patricia Miller, D. Dinsmore, and Conran A. Hay
1993 Predictive Model for Archaeological Resources, U.S. Route 202, Section 700, Bucks and Montgomery Counties, Pennsylvania. E.R. 1991-1019-042-KK. Prepared for the Pennsylvania Department of Transportation, Engineering District 6-0, St. Davids, Pennsylvania.

Duncan, Richard

2002 Centre and Clearfield Counties, Pennsylvania. S.R. 0322, Section 802 Corridor Project. Phase IA, Archaeological Investigation and Predictive Model Summary. E.R. 1999-2755-033-M. Prepared for the Pennsylvania Department of Transportation, Engineering District 2-0. Harrisburg, Pennsylvania. Skelly and Loy, Inc., Monroeville, Pennsylvania.

Duncan, Richard B., and Brian F. Schilling

1999a Fayette and Washington Counties Mon/Fayette Expressway Project Uniontown to Brownsville, Archaeological Predictive Model Development. E.R. 1987-1002-042-B03. Prepared for Pennsylvania Turnpike Commission, Harrisburg, Pennsylvania. Skelly and Loy, Inc., Monroeville, Pennsylvania.

1999b Northumberland, Snyder and Union Counties. Central Susquehanna Valley Transportation Project. S.R. 0015, Section 088. Archaeological Predictive Model. E.R.1997-0475-042-Q. Prepared for the Pennsylvania Department of Transportation, Engineering District 3-0. Harrisburg, Pennsylvania. Skelly and Loy, Inc., Monroeville, Pennsylvania.

Duncan, Richard B, Thomas C. East, and Brian F. Schilling

1999 U.S. Route 15 Improvement Project, Tioga County, Pennsylvania. S.R. 6015, Sections G20 and G22, Steuben County, New York. E.R. 1997-2018-117-H. Prepared for the Pennsylvania Department of Transportation and New York State Department of Transportation, Harrisburg, Pennsylvania. Skelly and Loy Inc., Monroeville, Pennsylvania.

Duncan, Richard B., Thomas C. East, and Kristen A. Beckman

1996 Allegheny and Washington Counties Mon/Fayette Transportation Project Interstate 70 to Route 51. Evaluation of Crooked Creek Predictive Model. E.R. 1987-1002-042-A02 & A03. Prepared for Pennsylvania Turnpike Commission, Harrisburg, Pennsylvania. Skelly and Loy, Inc., Monroeville, Pennsylvania.

Engelbrecht, William E.

1974 *The Iroquois: Archaeological Patterning on the Tribal Level*. Routledge & Kegan Paul, London.

Friedman, John H.

1991 Multivariate Adaptive Regression Splines. *Annals of Statistics* 19(1):1-67.

GAI Consultants

1998 Abbreviated Technical Report Phase I Cultural Resources Survey Proposed Knowledge Parkway Project Harborcreek Township, Erie County, Pennsylvania. E.R. 1992-0329-018. Prepared for Pennsylvania Department of Transportation, Engineering District 1-0, Franklin County, Pennsylvania. GAI Consultants, Inc., Monroeville, Pennsylvania.

Gehlke, C.. and Biehl, H.

1934 Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. *Journal of the American Statistical Association Supplement* 29:169–170.

Glenn, Jonathon

2010 Archaeological Overview and Sensitivity Models Erie National Wildlife Refuge Crawford County, Pennsylvania. E.R. 2012-1218-042-A. Prepared for U.S. Fish and Wildlife Service, Region 5 Hadley, Massachusetts. GAI Consultants, Inc., Homestead, Pennsylvania.

Hart, John

1994 Development of Predictive Models of Prehistoric Archaeological Site Location for the Lake Erie Plain and Glacial Escarpment in the Erie East Side Access Project Area, Erie County, Pennsylvania. E.R. 1992-0858-049-E. Prepared for the Pennsylvania Department of Transportation, Harrisburg, Pennsylvania. GAI Consultants, Monroeville, Pennsylvania.

Heberling and Associates

1995 Phase 1 Archeological Investigation, West Fairview Borough Park, Cumberland County, Pennsylvania. E.R 1985-1323-041-B. Prepared for The Redevelopment Authority of the County of Cumberland and The Borough of West Fairview, Cumberland County, Pennsylvania. Heberling Associates, Huntingdon County, Pennsylvania.

Jochim, Michael A.

1976 *Hunter-Gatherer Subsistence and Settlement: A Predictive Model.* Academic Press, New York.

Johnson, William C., William P. Athens, Martin T. Fuess, Luis G. Jaramillo, and Elizabeth Ramos

1989 Late Prehistoric Period Monongahela Culture Site and Cultural Resource Inventory. E.R. 1989-R015-042. Prepared for Pennsylvania Historical and Museum Commission, Harrisburg Pennsylvania. The Cultural Resource Management Program, Department of Anthropology, University of Pittsburgh, Pittsburgh, Pennsylvania.

Judge, James W.
1973 *Paleoindian Occupation of the Central Rio Grande Valley, New Mexico*. University of New Mexico Press, Albuquerque.

Judge, James W., and L. Sebastian (editors)
1988 *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, Denver, Colorado.

Katz, Gregory M., John P. Branigan, Paul W. Schopp, and Steven J. Biondo
2002 S.R. 0228, Section 290 Cranberry, Adams, and Middlesex Townships, Butler County, Marshall, Pine, and Richland Townships, Allegheny County, Pennsylvania. Volume 1. E.R. 1999-6127-019-H. Prepared for the Pennsylvania Department of Transportation, Harrisburg, Pennsylvania. A. D. Marble & Company, Conshohocken, Pennsylvania.

Kohler, Timothy A., and George G. Gumerman
2000 *Dynamics in Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes*. Santa Fe Institute Studies in the Sciences of Complexity. Oxford University Press, New York.

Kohler, Timothy A., and Sandra C. Parker
1986 Predictive Models for Archaeological Resource Location. In *Advances in Archaeological Method and Theory 9*, edited by M. B. Schiffer, pp. 397-452. Academic Press, Orlando, Florida.

Kvamme, Kenneth L.
1983 A Manual for Predictive Site Location Models: Examples from the Grand Junction District, Colorado. Bureau of Land Management, Grand Junction District.

1984 Models of Prehistoric Site Location near Pinyon Canyon, Colorado. In *Papers of the Philmont Conference on the Archaeology of Northeastern New Mexico*, edited by C.J. Condie. Proceedings of the New Mexico Archaeological Council 6(1), Albuquerque.

1988 Development and Testing of Quantitative Models. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling,* edited by W.J. Judge and L. Sebastian, pp. 325-428. U.S. Department of the Interior, Bureau of Land Management, Denver, Colorado.

1990 The Fundamental Principles and Practices of Predictive Archaeological Modeling. In *Mathematics and Information Science in Archaeology: A Flexible Framework*, edited by A. Voorrips, pp. 297-305. Studies in Modern Archaeology, vol. 3. Holos-Verlag, Bonn, Germany.

Lawrence, John W., Robert Herbstritt, John Branigan, and Paul W. Schopp
2002 Susquehanna Beltway Project S.R. 0220, Section 077 Woodward, Piatt, and Porter Townships and Jersey Shore, Lycoming County, Pennsylvania. Volume 1. E.R. 2002-8006-081-K. Prepared for the Pennsylvania Department of Transportation, Engineering District 3-0, Montoursville, Pennsylvania. A.D. Marble & Company, Conshohocken, Pennsylvania.

Lawrence, John W., David L. Weinberg, and Daniel R. Hayes
2003 Alternative Mitigation to the Interstate Fairgrounds Site (36BR210). S.R. 1056, Section 001, Athens Bridge Replacement Project, Athens Township, Bradford County, Pennsylvania. E.R. 2000-8029-015-R. Prepared for Dewberry-Goodkind, Inc. Carlisle, Pennsylvania. A.D. Marble & Company, Conshohocken, Pennsylvania.

MacDonald, Douglas
2006 Pennsylvania Archaeological Data Synthesis Subbasin 9: The Central West Branch Susquehanna River Watersheds A (Pine Creek), B (Kettle Creek) & C (Bald Eagle Creek) with a Focus on Great Island, Clinton County, Pennsylvania. E.R. 2004-1413-035-H. Prepared for PPL Gas Utilities, Quarrysville, Pennsylvania. GAI Consultants, Pittsburgh Office, Homestead, Pennsylvania.

MacDonald, Douglas, Jonathan C. Lothrop, and David L. Cremeens
2003 Pennsylvania Archaeological Data Synthesis: The Raccoon Creek Watershed (Watersheds D, Subbasin 20) Bridge Replacement Project T-319, Beaver County Bridge No. 36 (Link Bridge), Independence Township, Beaver County, Pennsylvania. E.R. 1996-8232-007-G. Prepared for Pennsylvania Department of Transportation. GAI Consulting, Inc., Monroeville, Pennsylvania.

MacDonald, Douglas H., Kenneth W. Mahoney, and Lisa Dugas
2003 Pennsylvania Archaeological Data Synthesis: The Upper Juniata River Sub-Basin 11 (Watersheds A-D) Walter Industrial Park: Mitigation of Adverse Effects, Grennfield Township, Blair County, Pennsylvania. E.R. 2000-2888-013-P. Prepared for Keller Engineers, Inc. Hollidaysburg, Pennsylvania. GAI Consultants, Monroeville, Pennsylvania.

McIntyre, J.

2009 East Resources Inc. Troy Pipeline Project, Lycoming and Bradford Counties, Pennsylvania. E.R. 2009-0922-042-B. Prepared for Entech Engineering, Inc., Reading, Pennsylvania, Utility Line Services, Inc., and Pennsylvania Historical and Museum Commission, Harrisburg, Pennsylvania. Pan Cultural Associated, Inc., Pittston, Pennsylvania.

Means, Bernard K.

1998 Phase 1 and Phase 2 Archaeological Investigations, U.S. 219 Meyersdale Bypass Project S.R. 6219, Section B08, Somerset County, PA. Volume 1., E.R. 1992-0237-111-A19. Prepared for the Pennsylvania Department of Transportation, District 9-0, Blair County, Pennsylvania. Greenhorne & O'Mara Inc., Mechanicsburg, Pennsylvania.

Miller, Patricia E.

2002 Archaeological Predictive Model Report and Recommendations, PA 23 EIS Project, SR 0023, Section EIS, Lancaster County, Pennsylvania. E.R. 2003-8015-071-G. Prepared for the Pennsylvania Department of Transportation Engineering District 8-0, Harrisburg, Pennsylvania. KCI Technologies Inc., Mechanicsburg, Pennsylvania.

Miller, Patricia E., and Marcia M. Kodlick

2006 Archaeological Predictive Model Field Test Results, PA 23 EIS Project SR 0023, Section EIS, Lancaster County, Pennsylvania. E.R. 2003-8015-071-G. Prepared for the Pennsylvania Department of Transportation, Engineering District 8-0, Harrisburg, Pennsylvania. KCI Technologies, Inc., Mechanicsburg, Pennsylvania.

Mooney, Douglas B., Rose L. Moore, Philip A. Perazio, Niels R. Rinehart, and James P. Davis

2003 Phase I Cultural Resource Investigations of the Planned Bushkill Road Schools Complex, Project Area, Lehman Township, Pike County, Pennsylvania. E.R. 1995-0370-103-H. Prepared for F .X. Browne, Inc., Lansdale, Pennsylvania. Kittatinny Archaeological Research, Inc., Stroudsburg, Pennsylvania.

Munoz, Jesus, and Angle M. Felicisimo

2004 Comparison of Statistical Methods Commonly used in Predictive Modeling. *Journal of Vegetative Science* 15:285-292.

Nass, John P., Jr., John Roger Wright, Lori Frye, and Rory Krupp

1992 Phase I Historic Properties Investigations, Youghiogheny River Lake Project, Fayette and Somerset Counties, Pennsylvania, and Garrett County, Maryland. E.R. 1981-0150-042-P.

Prepared for U.S. Army Engineer District, Pittsburgh, Pennsylvania. Archaeological Services Consultants Inc. Columbus, Ohio.

Neusius, Sarah W., and Phillip D. Neusius
1989 A Predictive Model for Prehistoric Settlement in the Crooked Creek Drainage. E.R. 1989-R016-042-A. Prepared for Pennsylvania Historic and Museum Commission, Harrisburg, Pennsylvania. Archaeology Program, Department of Sociology and Anthropology, Indiana University of Pennsylvania, Indiana, Pennsylvania.

Parker, Sandra.
1985 Predictive Modeling of Site Settlement System Using Multivariate Logistics. In *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique and Theory*, edited by Christopher Carr, pp. 173-205. Waveland Press Inc., Illinois.

Perazio, Phillip A.
1995 East Stroudsburg Area School District, Bushkill Road School Complex Project, Cultural Resources Sensitivity Study. ER 1995-0370-103-C. Prepared for F.X. Browne, Inc., Lansdale, Pennsylvania. Kittatinny Archaeological Research, Inc. Stroudsburg, Pennsylvania.

Polglase, Christopher
1997 Letter Report. Archeological Predictive Model for the ANR Independence Pipeline Project. E.R. 1984-1506-042-G. Prepared for Pennsylvania Historical and Museum Commission, Harrisburg, Pennsylvania. Christopher Goodwin and Associates, Fredrick, Maryland.

Reinbold, M.
2010 Talisman Energy USA Pipelines Ostrander to Longenecker Pipeline Located in Jackson Township, Tioga County, and Wells Township, Bradford County, Pennsylvania, and Yurkanin to Boor Pipeline, Columbia Township, Bradford County, Pennsylvania. Phase I A, Archaeological Survey and Predictive Model. E.R. 2010-1506-042-B. Tailsman Energy USA Inc., Horseheads, New York. Pan Cultural Associates, Inc., Pittston, Pennsylvania.

Salmon, Merrilee H.
1976 "Deductive" versus "Inductive" Archaeology. *American Antiquity* 41(3):376-381.

Smith, Bruce D.
1974 Middle Mississippi Exploitation of Animal Populations: A Predictive Model. *American Antiquity* 79(387):575-583.

Stewart, R. Michael, and Judson L. Kratzer
1989 Prehistoric Site Locations on the Unglaciated Appalachian Plateau. *Pennsylvania Archaeologist* 59(1):19-36.

VandenBosch, Jon, Edward J. Siemon III, and William C. Johnson
2000 Phase 1 Archaeological Survey of the Proposed East Side Access Highway, wintergreen George Bridge Project Area. SR4034-A91, Harborcreek Township, Erie County, Pennsylvania. E.R. 1992-0858-049-F Prepared for; Pennsylvania Department of Transportation Engineering District 1-0, Venango, Pennsylvania. Michael Baker Jr., Inc., Coraopolis, Pennsylvania.

Vento, Frank
1994 Volume IA, Genetic Stratigraphy: The Model for Site Burial and Alluvial Sequences in Pennsylvania. E.R. 1994-R001-042-A. Prepared for Pennsylvania Bureau For Historic Preservation, William Penn Museum and Historical Commission, Harrisburg, Pennsylvania. Department of Geology and Geography, Clarion University of Pennsylvania, Clarion, Pennsylvania.

Verhagen, Philip
2009 Testing Archaeological Predictive Models: A Rough Guide. In *Archaeological Prediction and Risk Management. Alternatives to Current Practice*, edited by H. Kamermans, M. van Leusen, and Ph. Verhagen (eds), pp. 63-70. Archaeological Studies Leiden University 17. Leiden University Press, The Netherlands.

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* 19(1):49-100.

Versaggi, N.M.
1979a 1979 Highway Program PIN 6108.05 Horseheads—Montour Corridor Pre-Reconnaissance Survey. Public Archaeology Facility, State University of New York, Binghamton.

1979b Routes 5 and 2902 Cultural Resources Reconnaissance Survey. Public Archaeology Facility, State University of New York, Binghamton.

Wall, Robert, Tim Sara, Eric Schmidt, and Andrew Ross

2008 Phase 1 Survey for the Armenia Mountain Wind Energy Project, Tioga and Bradford Counties, Pennsylvania. E.R. 2007-1478-042-D. Prepared for; AES Armenia Mountain Wind, LLC, Arlington, Virginia. TRC, Ellicott City, Maryland.

Wallace, Paul A.

1965 *Indian Paths of Pennsylvania.* Commonwealth of Pennsylvania, The Pennsylvania Historical and Museum Commission, Harrisburg, Pennsylvania.

Warren, Robert E.

1990 Predictive Modeling in Archaeology: A Primer. In *Interpreting Space: GIS and Archaeology*, edited by K.M.S. Allen, S. W. Green, and E.B.W. Zubrow, pp. 90-111. Taylor and Francis, London.

Weed, Carol

2002 Prehistoric Context Study (Chapter Three) in Support of Data Recovery at Site 36AL480, Leetsdale, Allegheny County, Pennsylvania. E.R. 1999-2661-003-T. Prepared for U.S. Army Corps of Engineers, Pittsburgh District, Pittsburgh, Pennsylvania. David Miller & Associates, Inc., Arlington, Virginia.

Whitley, Thomas G.

2005 A Brief Outline of Causality-Based Cognitive Archaeological Probabilistic Modeling. In *Predictive Modelling for Archaeological Heritage Management: A Research Agenda*, edited by Martijn van Leusen and Hans Kamermans, pp. 125-139. Nederlandse Archeologische Rapporten 29, Amersfoort.

Whitley, Thomas G., and Keith R. Bastianini

1992 The Design and Testing of a Mathematical Archaeological Predictive Model for the APEC, DCQ, and Storage and Transport Project Areas, Pennsylvania. E.R. 1992-R001-042-A. Prepared for Texas Eastern Gas Pipeline Company, Houston, Texas. Center for Cultural Resource Research, Pittsburgh, Pennsylvania.

Willey, G. R.

1953 *Prehistoric Settlement Patterns in the Virú Valley, Peru*. Bureau of American Ethnology Bulletin 155. Washington, D.C.

Yamin, Rebecca, Matthew D. Harris, Douglas C. McVarish, and Grace H. Ziesing
2010 Independence National Historical Park Archaeological Sensitivity Study (Phase IA Archeological Assessment, Independent Living History Center, North Lot). Prepared for Independence National Historical Park, Philadelphia, Pennsylvania. John Milner Associates, Philadelphia, Pennsylvania.

Yates, D., D. Moore, and G. McCabe
1999 *The Practice of Statistics* (1st Ed.). W.H. Freeman, New York, New York.

APPENDIX A

DERIVING THE KVAMME GAIN STATISTIC

DERIVING THE KVAMME GAIN STATISTIC

A measure of a model's performance needs to consider two very important aspects: 1) classification error; and 2) success in achieving the model's goals. The first measure of performance, classification error, is the calculation of the presence/absence of archaeological sites versus the predicted presence/absence generated from the model. This calculation begins with the creation of a 2 x 2 matrix of predicted and observed presence/absence as quantified by raster cells or site/non-site points (Table A-1).

Table A-1 - Schematic of 2 x 2 Table of Model Outcomes

		Probabilities of Assignment Model Prediction Present (M)	 Absent (M')	Total	%
Site Observation	Present (S)	M|S	M'|S	Total S	P_s
	Absent (S')	M|S'	M'|S'	Total S'	$P_{s'}$
	Total	Total M	Total M'	N	
	%	P_m	$P_{m'}$		

Kvamme (1990) established this method as a way to assess classification errors, probabilities, and a model's classification relative to chance. In this presentation, M and M' are the modeled presence and absence of sites, and S and S' are the observed presence and absence of sites, respectively (Figure A-1). The values of M|S, M'|S, M|S', and M'|S' are the count of cells (or site/non-site points) within the space assigned to each of those four classes. Table A-2 is an example of the results of a model that contained 200 raster cells (or 200 site/non-site points) and the calculation of the proportion of cells assigned to each of the four classes.

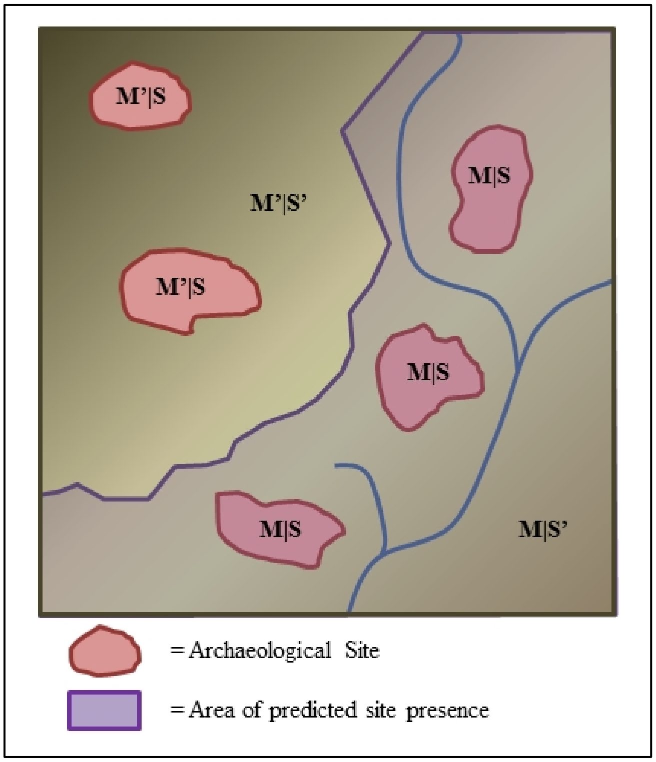

Figure A-1 - Schematic of APM results and components of classification.

Table A-2 - Example of Calculating Probabilities of Model Assignment

Probabilities of Assignment

		Model Prediction			
		Present (M)	Absent (M')	Total	%
Site Observation	Present (S)	0.085 (17)	0.015 (3)	20	0.10
	Absent (S')	0.18 (36)	0.72 (144)	180	0.90
	Total	53	147	200	1
	%	0.265	0.735	1	

From Table A-2, these numbers can be understood as the percent of the model space classified into each of the four classes. The cell for M|S (0.085) shows that 8.5% of the model space was correctly classified as an archaeological site location by the model. The cell for M'|S' (0.72) shows that 72% of the model space was correctly classified as not containing sites; the cells of M|S' and M'|S follow suit. However, to calculate the classification error these percentages must be divided by the percent total of each row, that is, the percent of site-present cells ("Total S") and the percent of site-absent cells ("Total S'") (Table A-3).

Table A-3 - Schematic of Calculations of Classification Success and Error

Conditional Probabilities

		Model Prediction		
		Present (M)	Absent (M')	Total
Site Observation	Present (S)	$P_{m \mid s}$	$P_{m' \mid s}$	Total
	Absent (S')	$P_{m \mid s'}$	$P_{m' \mid s'}$	Total

$P_{m|s}$ = (M|S) / %Total S

$P_{m'|s}$ = (M'|S) / %Total S

$P_{m|s'}$ = (M|S') / %Total S'

$P_{m'|s'}$ = (M'|S') / %Total S'

The normalization of each percentage from Table A-2 by the total percent of the observed site presence/absence cells results in the correct classification percent ($P_{m|s}$ and $P_{m'|s'}$) and the classification error percentage ($P_{m'|s}$ and $P_{m|s'}$) (Table A-4). Using the same example as above, the model correctly classified sites 85% (true-positive) of the time and non-sites 80% (true-negative) of the time. Alternately, the model had a classification error of 15% for site absence ($P_{m'|s}$) and a 20% classification error for site presence ($P_{m|s'}$). The error of site presence ($P_{m|s'}$) indicates a false-positive (Type I) error (considered a “wasteful error” because it requires additional survey that may turn up no evidence of archaeological sites). On the other hand, the classification error for site absence ($P_{m'|s}$) indicates a false-negative (Type II) error (considered a “gross error” because it may lead to the destruction of archaeological sites in regions thought to be of low sensitivity).

Table A-4 - Example of Calculating Classification Success and Error

		Conditional Probabilities Model Prediction Present (M)	 Absent (M')	Total
Site Observation	Present (S)	0.85	0.15	1
	Absent (S')	0.20	0.80	1

$P_{m|s}$ = 17/20
$P_{m'|s}$ = 3/20
$P_{m|s'}$ = 36/180
$P_{m'|s'}$ = 144/180

Condensing the measures of classification success into a single number that can generally describe a model's classification efficiency is an important step in comparing models to one another. The standard measure of model efficiency used within the APM literature is the Kvamme Gain Statistic (K_g) (Kvamme 1988). The K_g is a measure of a model's percent area predicted for site presence divided by the percent of all sites within that area.

$$K_g = 1 - \left(\frac{\%\ predicted\ site\ present\ area}{\%\ correct\ site\ prediction}\right)$$

Using the terms of our example, K_g = 1 - (%Total M / $P_{m|s}$). Substituting the values of our example, K_g = 1- (0.265/0.85), this equation divides the 26.5% of the study area predicted to contain sites (from Table A-2, column one, total percent) by the 85% true-positive rate (from Table A-4, $P_{m|s}$ result). The resulting gain is K_g = 0.688. This measure of efficiency is useful because it allows for variations of a single model or completely separate models to be compared on the same basis of correct site classification. However, the gain statistic has drawbacks. Most importantly, the K_g does not distinguish between model completeness and efficiency. A model biased toward completeness may encompass all of the known archaeological sites (a high $P_{m|s}$) but do so only because the site-likely area covers a large region (a high Total M). On the other hand, a model biased toward efficiency may minimize the region of site-likely (lower Total M), but also correctly classify fewer known sites (lower $P_{m|s}$). The fact that you can have models with the same K_g but different classification errors demonstrates that the K_g is most useful when considered in tandem with the false-negative and false-positive classification error rates.

The tradeoff between model completeness and efficiency is the tradeoff between identifying a larger area needed for survey and identifying a smaller number of known sites within the site-likely model (Figure A-2). In terms of the K_g statistic, the 0.688 gain of the example model can also be achieved with a %Total M as low as 0.05 and a $P_{m|s}$ of 0.16, or a %Total M as high as 0.30 and a $P_{m|s}$ of 0.95. In the first instance, you have a very precise model that claims that only 5% of the study area is site-likely, but it only correctly classifies 16% of the sites (misclassifies 84%). In the second instance, the model includes 30% of the study area as site-likely, but correctly classifies 95% of known sites (misclassifies 5%). In each of these situations, the model's K_g = 0.688, but the balance between efficiency and completeness are very different. In one model, the area suggested for survey is quite small, but arguably, the potential for a construction project to impact an unidentified site is quite high, given that 84% of the known sites were misclassified. The other model has a very good chance of identifying most archaeological sites (5% misclassification), but at the expense of a much larger survey area. This leads directly to the point of clearly defining a model's purpose.

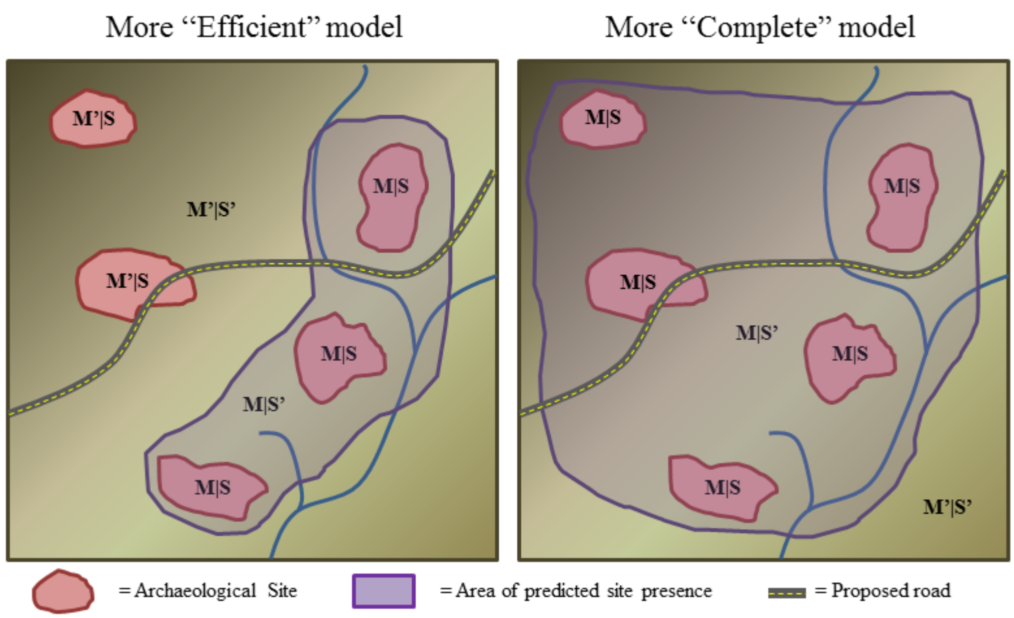

Figure A-2 - Schematic of model results biased toward efficiency and completeness.

In most CRM situations, it is assumed that both resource managers and construction managers prefer completeness over efficiency. This is because presumably a manager would prefer to spend resources upfront on additional survey to avoid spending potentially greater resources during construction when an unanticipated archaeological site is impacted. As a general rule, the goal of APMs in a CRM context is to bias toward completeness in order to decrease the chance of false-negatives and avoid costly project interruption. In the planning phase, a model that achieves a K_g of 0.688 and recommends survey of only 5% of the study area may sound very good at the outset. However, once balanced with the high misclassification percent and potential for site impacts, it is likely not as appealing. This model, albeit good by the numbers presented to the client, would not likely achieve the goals of the resource manager or modeler. This model would have low performance based on missing its goals and being ineffective at best and disastrous to unidentified archaeological sites at worst. The other example above and the original example from Table A-2, which have a lower misclassification percent but a larger recommended survey region, are much better performers and more ideally balanced toward the mix of completeness and efficiency necessary to give a resource manager confidence when implementing the models. In practice, many models that are considered relatively successful achieve a K_g of between 0.60 and 0.80. Lower values are likely the result of too few correct classifications and too large of a site-likely area. Higher K_g value is certainly possible, and desired, in a successful model, but is often the result of overly precise model bias that can contain high rates of misclassification errors.

In assessing the performance of the models reviewed in this report the K_g statistic, classification errors, and the efficacy in achieving a model's goals will be considered together. The K_g statistic will give an overall sense of a model's efficiency in correctly classifying site locations relative to the size of the region it considers likely to contain sites. The classification successes and errors will give a more refined picture of the model's balance between completeness and efficiency. And finally, a model's effectiveness in achieving its goals will be combined with the previous two measures to judge the model's overall performance.

APPENDIX B

VARIABLES FROM DUNCAN ET AL. (1996) AND DUNCAN AND SCHILLING (1999A, 1999B)

Table B-1 - Full List of Variables Considered by Duncan et al. (1996)

Var#	Name	Field Type	Description
1	abvstrm	continuous	elevation above stream
2	macdems_sm1	continuous	elevation of cell in feet
3	macslp	continuous	slope of the cell in percent
4	macslp_inv	continuous	inverted, transformed slope
5	macasp	continuous	compass heading for aspect
6	asp0_180	continuous	aspect as degrees from North
7	solar1	continuous	insolation value in morning
8	relief10	continuous	local relief of the cell
9	relief30	continuous	wider local relief
10	terruf10	continuous	local terrain roughness
11	trlscst	continuous	cost-distance to nearest trail
12	trlsdist	continuous	distance to nearest trail
13	sprngcst	continuous	cost-distance of nearest spring
14	sprngdist	continuous	distance to nearest spring
15	strmcst1-6	continuous	cost-distance to nearest stream of any type
16	stream 1-3	continuous	cost-distance to nearest intermittent stream
17	strmcst4-6	continuous	cost-distance to nearest perennial stream/river
18	strmcst4	continuous	cost-distance to minor tributary
19	strmcst5	continuous	cost-distance to major tributary
20	strmcst6	continuous	cost-distance to nearest river
21	solar2	continuous	insolation at noon
22	solar3	continuous	insolation in afternoon
23	solarsum	continuous	weighted average of solar 1-3
24	solarwt	continuous	insolation adjusted for slope
25	solarwt2	continuous	insolation adjusted for 4th root of slope
26	maccur	cont. (-1 to +1)	a curvature difference measure
27	macplancrv	cont. (-1 to +1)	convexity measure
28	macprocrv	cont. (-1 to +1)	concavity measure
29	teruf210	continuous	standard deviation of change in slope
30	rimindx2	continuous	relationship to topographic rim
31	vantgcst	continuous	cost-distance to vantage point
32	peakcst	continuous	cost-distance to vantage point (peak)
33	ridgcst	continuous	cost-distance to ridgetop
34	divcst	continuous	cost-distance to nearest drainage divide
35	divdist	continuous	Euclidean distance to drainage divide
36	wtrdist	continuous	Euclidean distance to nearest perennial water source
37	cnflcst	continuous	cost-distance to nearest stream confluence
38	cnfl4cst	continuous	cost-distance to stream confluence along tributaries

Var#	Name	Field Type	Description
39	cnfl5cst	continuous	cost-distance to confluence along major tributaries
40	cnfl6cst	continuous	cost-distance to confluence along river
41	abvstrm2	continuous	elevation difference from nearest stream cell
42	sad1cst	continuous	cost-distance to nearest saddle > 0
43	sad2cst	continuous	cost-distance to nearest saddle > 1
44	sad3cst	continuous	cost-distance to nearest saddle > 2
45	sad4cst	continuous	cost-distance to nearest saddle > 3
46	sad5cst	continuous	cost-distance to nearest saddle > 4
47	soil#	integer (1 to 24)	SCS soil code
48	rock#	categorical	geologic bedrock formation code
49	soilcap_r	rank value (0 to 4)	agricultural capability class
50	opnland_r	rank value (0 to 4)	soil suitability for open land wildlife
51	wetland_r	rank value (0 to 4)	soil suitability for wetland wildlife
52	wdland_r	rank value (0 to 4)	soil suitability for woodland wildlife
53	hrdwd_r	rank value (0 to 4)	soil suitability for hardwood growth
54	d_bdrk_r	rank value (0 to 3)	code ranking the depth to bedrock
55	drain_r	rank value (0 to 3)	ran of soil drainage character
56	wtrcap_r	rank value (1 to 4)	rank of soil water capacity
57	dssnhiw_r	rank value (0 to 3)	rank of depth to seasonal high water table
58	text_r	rank value (0 to 3)	rank of soil texture category
59	solidstrb	rank value (0 to 10)	rank of soil disturbance potential
60	dist_pres2	rank value (0 to 10)	combined road, water, and soils disturbance code
61	strmmk	rank value (1 to 6)	stream rank of nearest stream
62	corn_fert	ratio/continuous	corn bushels per acre
63	foot_slope	binary (0 or 1)	soil in foot slope environment
64	hill_side	binary (0 or 1)	soils in hillside environment
65	ridge_top	binary (0 or 1)	soils in ridgetop environment
66	stream_terr	binary (0 or 1)	soils in stream terrace environment
67	flood_plain	binary (0 or 1)	soil in floodplain environment
68	saddle	binary (0 or 1)	soil in saddle environment
69	trisbf5g	binary (0 or 1)	cell located around historic Indian trail
70	divgf5g	binary (0 or 1)	cell located around drainage divide
71	rddstrbbf	binary (0 or 1)	road disturbance factor
72	wtrdstrb	binary (0 or 1)	water disturbance factor

Table B-2 - Full List of Variables Considered by Duncan and Schilling 1999a and 1999b

Var#	Name	Field Type	Description
1	abvstrm	continuous	elevation difference from nearest stream cell
2	asp0_180	continuous	aspect as degrees from North
3	asp90	continuous	aspect as degrees for East-West
4	aspect	continuous	compass heading for aspect
5	chrt_cst	continuous	cost distance to geologic bedrock formation with lithic resource
6	cnflcst1	continuous	cost distance to nearest intermittent stream confluence
7	cnflcst1_2	continuous	cost distance to nearest inter. or perennial stream confluence
8	cnflcst2	continuous	cost distance to nearest inter. to perennial stream confluence
9	cnflcst3	continuous	cost distance to confluence, inter. to major tributary or river
10	cnflcst4	continuous	cost distance to confluence of perennial streams
11	cnflcst4_7	continuous	cost distance to confluence: per. Streams, tributaries and rivers
12	cnflcst5	continuous	cost distance to confluence of per. Streams and river or tributary
13	cnflcst5_7	continuous	cost distance to confluence along major tributaries and rivers
14	cnflcst6	continuous	cost distance to confluence of major tributary and river
15	cnflcst7	continuous	cost distance to confluence of river
16	csvtdem	continuous	elevation of cell in feet
17	euc_strm	continuous	Euclidean distance to nearest perennial water source
18	flatsum	continuous	degree of surrounding flat terrain
19	peakcst	continuous	cost distance to vantage point (peak)
20	relief10	continuous	local relief of the cell
21	relief30	continuous	wider local relief
22	ridgcst	continuous	cost distance to ridgetop
23	rimindx2	continuous	relationship to topographic rim
24	rvrcst	continuous	cost distance to nearest river
25	sadcst0	continuous	cost distance to nearest saddle (all)
26	sadcst1	continuous	cost distance to nearest saddle (moderate)
27	sadcst2	continuous	cost distance to nearest saddle (major)
28	slp_g	continuous	slope of cell in percent
29	slp_inv	continuous	inverted, transformed slope
30	sola1rwt	continuous	morning insolation adjusted for slope
31	solarwt	continuous	full day insolation adjusted for slope
32	strm_rvrcst	continuous	cost distance to major tributary an driver
33	strmcst1	continuous	cost distance to nearest intermittent stream
34	strmcst1-2	continuous	cost distance to nearest stream of any type
35	strmcst2	continuous	cost distance to nearest perennial stream
36	strmcst3	continuous	cost distance to lakes and ponds
37	strmcst4	continuous	cost distance to major river tributary
38	terruf10	continuous	local terrain roughness

Var#	Name	Field Type	Description
39	teruf210	continuous	standard deviation of change in slope
40	trlscst	continuous	cost distance to nearest historic Indian trail
41	vantgcst	continuous	cost distance to vantage point
42	wetcst	continuous	cost distance to nearest mapped wetland
43	soil#	integer (1-100)	SCS soil code
44	corn_fert	rank value	corn bushels per acre
45	d_bdrk_r	rank value	code ranking the depth to bedrock
46	drain_r	rank value	rank of soil drainage character
47	dssnhiw_r	rank value	rank of depth to seasonal high water table
48	flood_r	rank value	rank of flood frequency
49	hydro_r	rank value	rank of hydrologic character
50	opnlnd_r	rank value	soil suitability for open land wildlife
51	soilcap_r	rank value	agricultural capability class
52	text_r	rank value	rand of soil texture category
53	wdlnd_r	rank value	soil suitability for woodland wildlife
54	wetland_r	rank value	soil suitability for wetland wildlife
55	distrb3	rank value	combined road, building, water, and soils disturbance
56	soildstrb	rank value	rank of soil disturbance potential
57	rddstrbbf	rank value	road disturbance factor
58	bldgdstrb	rank value	building disturbance factor
59	wtrdstrb	rank value	water disturbance factor