A GEOGRAPHIC INFORMATION SYSTEM MODEL OF PREHISTORIC MOUND LOCATION IN IOWA

Chad Goings
Rolling Hills Consulting Services, LLC
1221 East 3rd Street
Washington, Iowa 52353
cagoings@aol.com

This paper presents an objective, statewide model of prehistoric mound locations in Iowa. It departs from others models through the dynamics of its geographic scope, its specificity to one type of archaeological site, and its methodology. The paper highlights the importance of considering the implications of the geographic nature of archaeological data and its effect on statistical modeling procedures. The model results were tested using 818 independent mound sites. Seventy-two percent were located in the ten percent of the area of the state that exhibits a high likelihood of containing mounds, a significant difference over chance models based on Kvamme's gain statistic.

Modeling Prehistoric Site Location

In 1928, Charles Keyes estimated that before Euro-American settlement as many as 10,000 mounds once existed in Iowa. Many mounds have been lost to development and heavy agricultural use of the land. However, enough have been recorded and studied to examine patterns in their distribution. As Clark Mallam (1976:19) stated with respect to effigy mounds in northeastern Iowa, "the environment… is of crucial importance in the assessing the cultural dynamics of the Effigy Mound manifestation." I would extend this statement to include that the environment is of crucial importance for distinguishing patterns in the modern distribution of all prehistoric mounds in Iowa and propose a model based on environmental factors to do so.

A model is an abstract conception of reality and the use of models to explain or predict prehistoric site location is not new to Iowa archaeology (e.g., Abbott 1982; Benn 1987; Chadderdon 2003; Goings 2003, 2005; Parrish 1998; Schermer 1982; Zimmerman 1977). A model never matches reality; however, some models come closer than others. Occam's razor tells us that the simplest explanation is usually the correct one.

Site modeling efforts in Iowa archaeology may be divided into two types: those developed before the availability of computer geographic information systems (GIS), and those that employ GIS software. GIS software combines maps with databases, has greatly advanced the ability to combine multiple variables in a statistical framework, and can express analytical results in map form. GIS started being used in Iowa archaeology in the early 1990s and its use has increased dramatically since the mid-1990s. Datasets representing the different variables used in a GIS project are often referred to as layers because they may be overlain together or on base maps such as topographic maps or aerial photographs to illustrate environmental characteristics and relationships.

Pre-GIS models in Iowa archaeology employed the site catchment approach and used the concept of environmental diversity (e.g., Abbott and McKay 1978; Schermer 1982; Schermer and Tiffany 1985; Tiffany and Abbott 1982). This approach used a circular territory around an archaeological site, often one or two kilometers in diameter, within which environmental features were mapped or described. To explain the location of a site, it was then argued that the environmental diversity within the catchment territory was greater than the environmental diversity in general. This was seen as a means of showing that prehistoric peoples considered the environmental diversity of their surroundings when determining settlement locations and satisfying subsistence needs. These pre-GIS models were generally more anthropological in nature, but lacked the technological ability to apply the model to a large geographic area at a high resolution.

GIS models in Iowa have characteristics similar to the site catchment models. They often involve a sample of known archaeological sites and examine several variables at those locations such as distance to the nearest water source or slope. The idea is that archaeological sites can be found on the modern landscape in specific environmental settings such as on level or gently sloping topography near major water sources. A GIS can combine such variables and produce a map depicting the possible locations of archaeological sites. These models have the technological ability to provide high-resolution maps over large areas, but generally lack in anthropological rigor and consist of obscure or esoteric statistical practices.

Two factors set this paper and its model apart from previous attempts. One is that there has not been an objective, statewide model specific to prehistoric mound locations in Iowa. The other is that many of the statistical models in Iowa archaeology, and in archaeology in general, do not take in to account the geographical nature of the data being used. This is problematic because the statistical procedures employed require that the inputted data be independent of one another. Spatial proximity can violate this requirement and bias the outcome. The purpose of this paper then is to demonstrate a statistical model of prehistoric mound locations in Iowa accounting for the geographical nature of the data itself in the statistical procedures.

Methodological Considerations

The first law of geography can be summarized in the statement: "things that are closer in space are more alike than things that are farther apart." In other words, two objects that are next to one another are generally not independent of one another. This is especially true if the objects are of the same type. For example, if one were to measure and record the distance to a tree, then take a step left or right and repeat the measurement, the distance will not be dramatically different. For statistical purposes, the values within a dataset should be independent of one another and should not have the ability to be predicted based on a determining factor.

This phenomenon can be measured in statistics through what is called spatial autocorrelation. The higher the spatial autocorrelation, the more alike the neighboring value and vice versa. This information is important to note in statistical models, because it creates two problems. The first is that most statistical procedures, even simple ones like finding an average value or viewing a histogram, require that the data in the sample are independent of one another. The second is that redundancy is undesirable and including neighboring values generally biases the sample towards a geographical location.

GIS models always involve geographic data; hence, spatial autocorrelation is always an issue. The archaeological data in a GIS layer can be stored as points, lines, or polygons. Since it is difficult to place discrete boundaries around an archaeological site, as the people who once utilized that location would not have restricted themselves in this way, lines and polygons outlining site areas have limited utility in my analysis. The first law of geography tells us that the immediate area at an archaeological site will have similar characteristics. To use the distance example given earlier, if one were to measure the distance to the nearest major water source from several locations within an archaeological site, the values would be similar. Therefore, to reduce redundancy and eliminate spatial autocorrelation issues, one point in the center of the archaeological site would be a close representation of the environmental factors at that particular site. Using all of the values within a site polygon would again create redundancy, would violate statistical assumptions, and would bias the results.

Methods

Statistical Samples

Ideally, one would want to randomly survey tracts across Iowa before European settlement to create a statistical sample of mound locations to use in a study such as this. Since we cannot go back in time, I have no choice but to use the mound location data at hand, which is the product of non-systematic surveying in an area where agricultural activity and urban development have greatly influenced the distribution of known prehistoric mounds. With this in mind, a random sample of known prehistoric mounds was used.

Location information on Iowa's prehistoric mounds was obtained from the Burials Program at the University of Iowa-Office of the State Archaeologist (UI-OSA). The information was received in a Microsoft Access database, containing attributes associated with the mounds including the Universal Transverse Mercator (UTM) coordinates and mound type including conical, effigy, linear, and unspecified types. The UTM coordinates were used to create a point layer in the GIS containing 1,153 mound locations.

In an effort to further minimize spatial autocorrelation problems, a one-kilometer buffer was created around each mound point location. Any recorded mound points that occurred within the buffer of another mound were removed from the sample. This resulted in 335 mound points that were used to study the environmental conditions. This set of points was called the "training" sample, and in a sense was used to train the model to identify environmental conditions in which prehistoric mounds are found. The remaining known mounds in the database were left out to independently test the model.

To obtain the background or "non-mound" environment, a systematic sample of point locations was generated across the state of Iowa at a fifteen-kilometer interval resulting in a set of 640 points to use against the 335 mound points. These non-mound points are sufficiently distant from each other to capture the background environment and keep the values

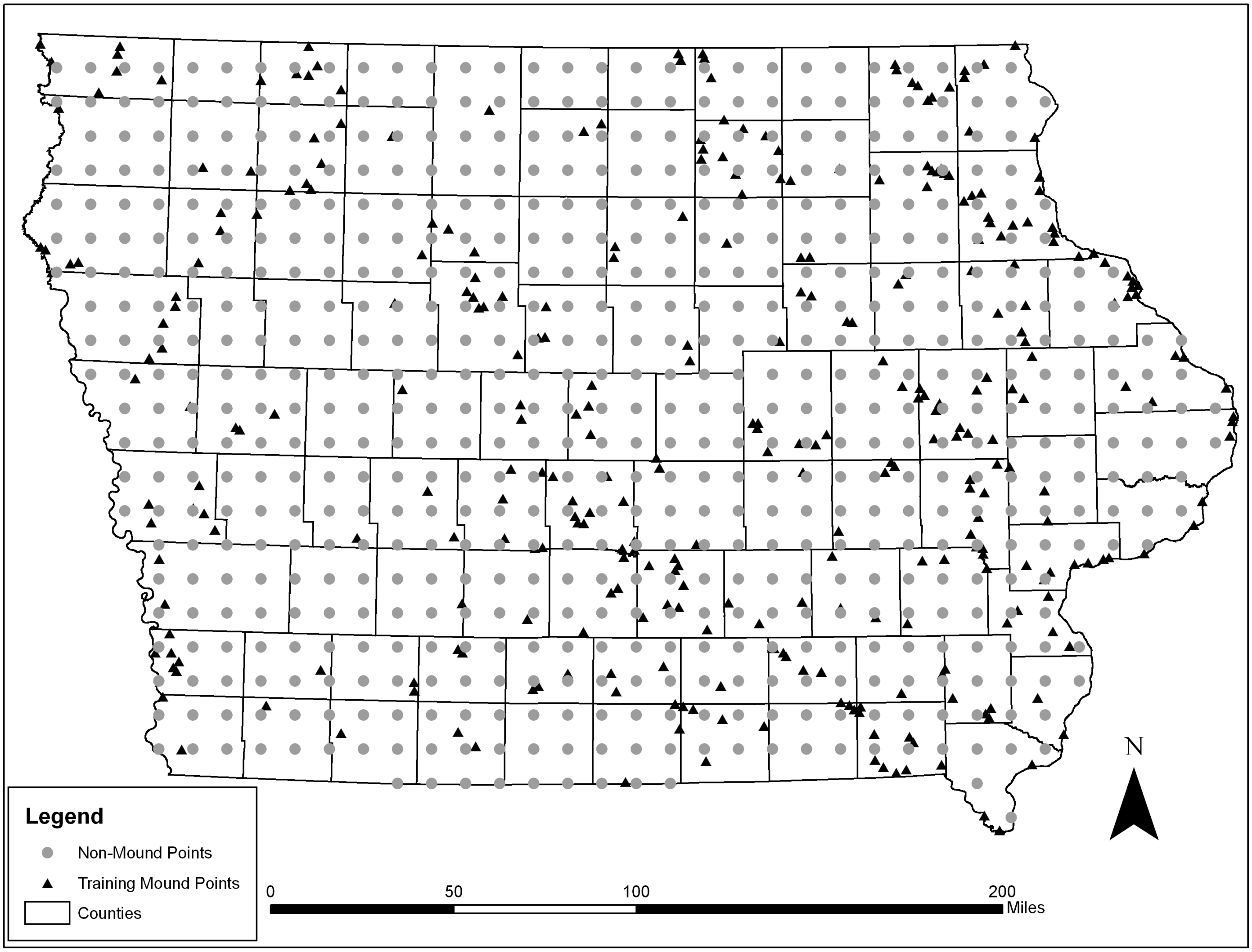

Figure 1. Map showing the distribution of training mound points and non-mound points.

independent. It is assumed that prehistoric mounds were placed in specific environments and that the probability of choosing a mound location by chance using this procedure is extremely low. Population ecology categorizes plant and animal species distributions as random, systematic, or clustered. Most populations and associated behavior, including human, are naturally clustered. Given the low likelihood of selecting a mound site by chance, these systematically-generated points can safely be considered "non-mound" points (Kvamme 1992:28). The non-mound points will be used to contrast the environmental differences with the training mound points. In other words, the question asked is: what is different about the environmental conditions at mound locations compared to the background or environmental conditions in general? Figure 1 shows the spatial distribution of the 335 training mound points and the 640 non-mound points.

Independent Variables

The independent environmental variables used in my model were absolute elevation, slope, local relief, and distance to a major water source (Figure 2). An elevation grid for the state of Iowa was downloaded from the Natural Resources Geographic Information Systems (NRGIS) Library maintained by the Iowa Department of Natural Resources-Iowa Geological and Water Survey (IDNR-IGWS). The grid originated with the EROS Data Center at the United States Geological Survey (USGS), and was clipped to the state boundaries by the IDNR-IGWS. The cell size of the elevation grid is 30 x 30 m and the grid is referenced in UTM coordinates using the North American Datum of 1983. Therefore, all datasets used in my model are at a similar resolution and spatial reference.

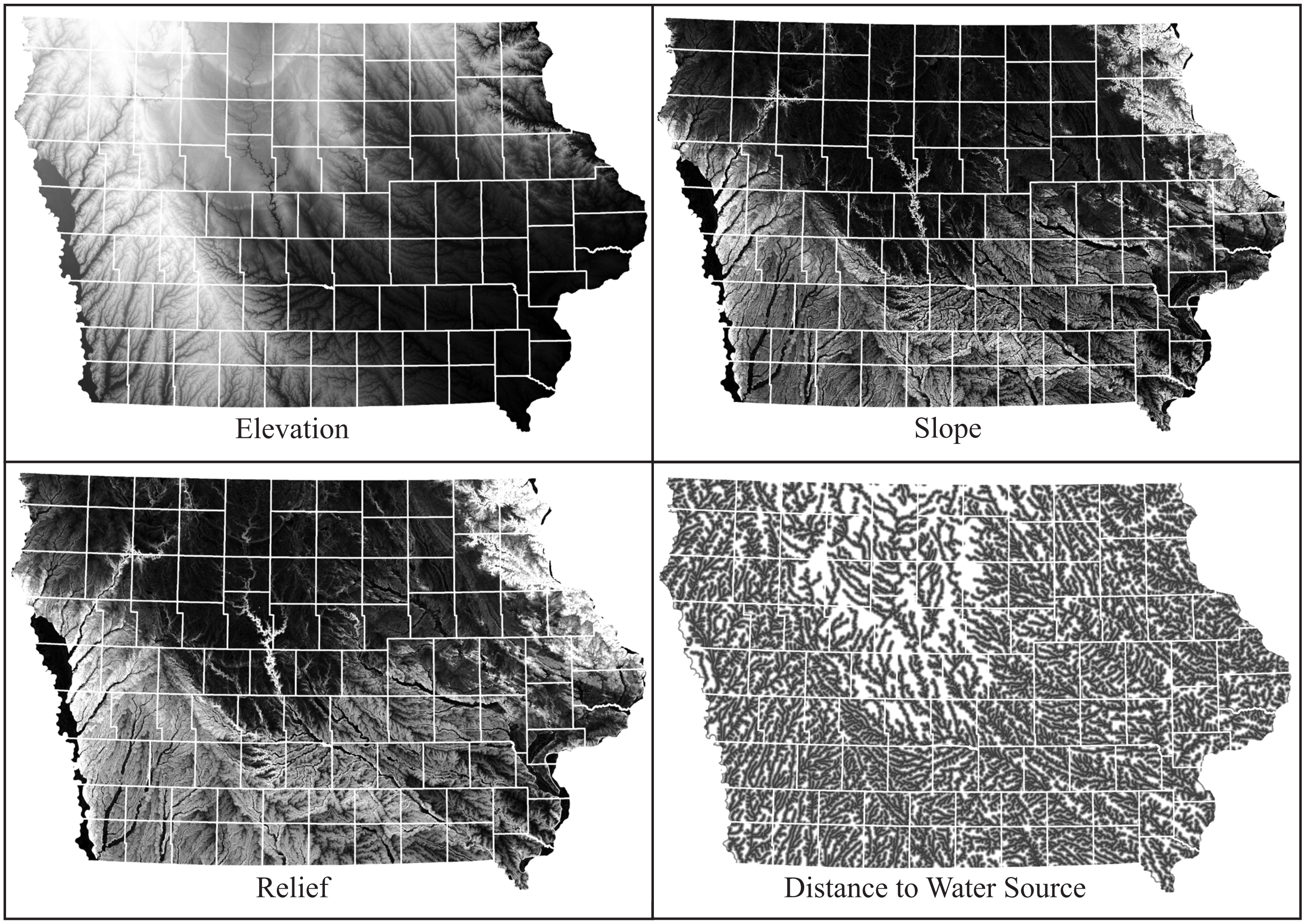

Figure 2. Independent environmental variables where dark colors are low values and light colors are high values.

A layer consisting of slope by degree data was generated by using the elevation grid as an input layer, and the local relief at training mound and non-mound points was calculated using a circular neighborhood statistic on the elevation layer. This statistic finds the range of elevations within a circle of a given diameter. The average distance between a summit landscape position and its adjacent channel is an appropriate width for the diameter of this circle (Gallant and Wilson 2000:74). The reciprocal of the drainage density statistic is an estimated distance between any two neighboring stream channels within a drainage basin. Half of this distance would be the estimated length from a summit or ridge to the stream channel (Leopold et al. 1964:146). Drainage basin statistics for the state of Iowa have been generated and compiled by the IDNR-IGWS, revealing that the average distance from summit to stream channel across the state is 0.387 kilometers. This distance was used to calculate local relief.

The IDNR-IGWS has created a GIS layer of streams in Iowa that have been ordered using Strahler's stream ordering system, in which the larger the stream, the higher its order. This layer was queried to show only streams that are greater than or equal to a second-order stream. These streams were exported and used as the major streams in Iowa. Then a straight-line distance grid was calculated from the major streams to create a stream-distance grid. A layer also showing major lakes and wetlands was obtained from the IDNR-IGWS for the Des Moines Lobe landform region to be used in this analysis of distance to major water source. Since mound locations dramatically decrease as one moves away from a major water source, the distance to nearest water factor is not a linear expression. A first-order polynomial was therefore used on the straight-line distance grid to better capture the close proximity of mounds to major water sources. This is based on a similar idea in Lensink's (1984) model of foraging habits on the Des Moines Lobe in relation to wetlands.

Univariate and Multivariate Statistical Analysis

Univariate and multivariate analysis was conducted using Insightful Corporation's S-Plus software. Values for the independent variables were extracted at the mound and non-mound point locations. Differences in the frequency distribution of these values were examined using histograms (Figure 3). Ideally, the non-mound points will match the values of the state of Iowa as whole on each variable and the training mound points should show a difference.

Histograms permit visual comparison of two or more distributions. However, a more objective analysis of the differences suggested by a histogram can be conducted through a two-sample Kolmogorov-Smirnov test. These were generated for each variable. Each variable was also examined for non-linear and interaction effects by noting any significant improvements in the t-value when polynomials or products were used, respectfully.

For multivariate analysis, a logistic regression was used. A logistic regression is similar to a standard regression in statistics except that the dependent variable is binomial. That is, it can either be in one of two categories. In this case, we want to know if a 30-x-30-meter parcel on the ground contains a mound or does not contain a mound based on the independent environmental variables and the samples used.

A stepwise logistic regression was used to generate the most efficient model. This method systematically removes and adds independent variables from the model to see their significance in explaining the environmental variance between mound and non-mound locations. Then the coefficients were jackknifed[1] by removing each point from the analysis and seeing what influence each point had on the regression coefficient. The mound or non-mound points determined to have a big influence on the logistic regression coefficients were located and investigated. The average coefficients resulting from the jackknife procedure were used to map the model in the GIS software.

Table 1. Results of Kolmogorov-Smirnov (KS) Tests on Each Independent Variable.

Variable	KS	p-Value
Elevation	0.2485	<0.005
Slope	0.2746	<0.005
Relief	0.3982	<0.005
Distance to Major Water Source	0.3384	<0.005

Results

Histograms (Figure 3) and Kolmogorov-Smirnov tests (Table 1) show a statistically significant difference for values at mound and non-mound locations for each of the independent variables. Relief and the distance to major streams have the highest significance. It appears that the training mounds are located in lower absolute elevations, in steeply sloping areas with greater relief, and near major water sources compared to non-mound locations.

There were no non-linear or interaction effects[2] detected for the independent variables based on decreases in the associated t-values when polynomials or products were used. The stepwise method demonstrated that all four independent variables contribute to the explanation of the environmental variance between mound and non-mound locations. Therefore, all four variables were kept in the jackknife procedure. Table 2 demonstrates that the mean coefficient values from the jackknife procedure were similar to the observed values, meaning that the distribution of mound and non-mound points influence was generally Gaussian[3] and the sample size was sufficient to minimize the bias of outliers.

The model results consisted of a 30-x-30-meter continuous grid of Iowa with each grid cell being assigned a value between zero and one. A value near zero means that a 30-x-30-meter piece of land is less suitable for mound locations, and a value near one means that a 30-x-30-meter parcel is more suitable for mound locations based on the environmental variables and the samples used.

There are a number of objective methods for determining proper cut-off values for continuous models and creating categories of low, moderate, or high suitability. The method used here was to break the model grid up in to three natural breaks categories. A map (Figure 4) was created to show the model probabilities across the state of Iowa and the model probabilities at all recorded mound locations in Iowa. Using this classification, a majority of the state (57 percent) has a low probability for containing mounds. Of

Table 2. Summary Statistics for Logistic Regression Parameters.

Factor	Observed	Mean	Bias	Standard Error
(Intercept)	0.14	0.20	0.07	0.41
Elevation	-0.001	-0.001	0.00	0.00
Slope	0.07	0.06	0.00	0.03
Relief	0.01	0.01	0.00	0.00
Distance to Major Water Source	-0.0008	0.00	0.00	0.00

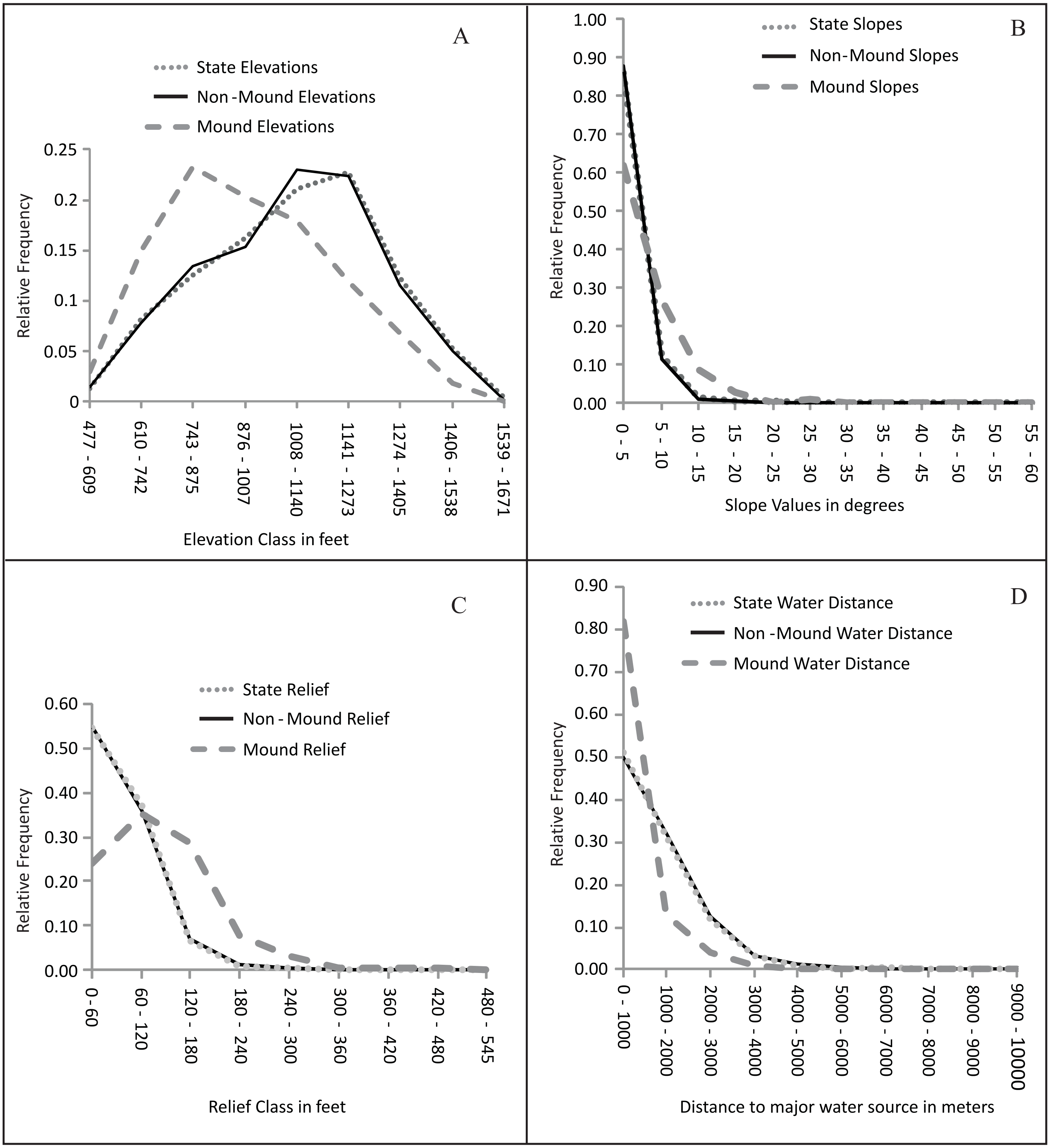

Figure 3. Histograms showing the frequency of values for (a) elevation, (b) slope, (c) relief, and (d) distance to water source.

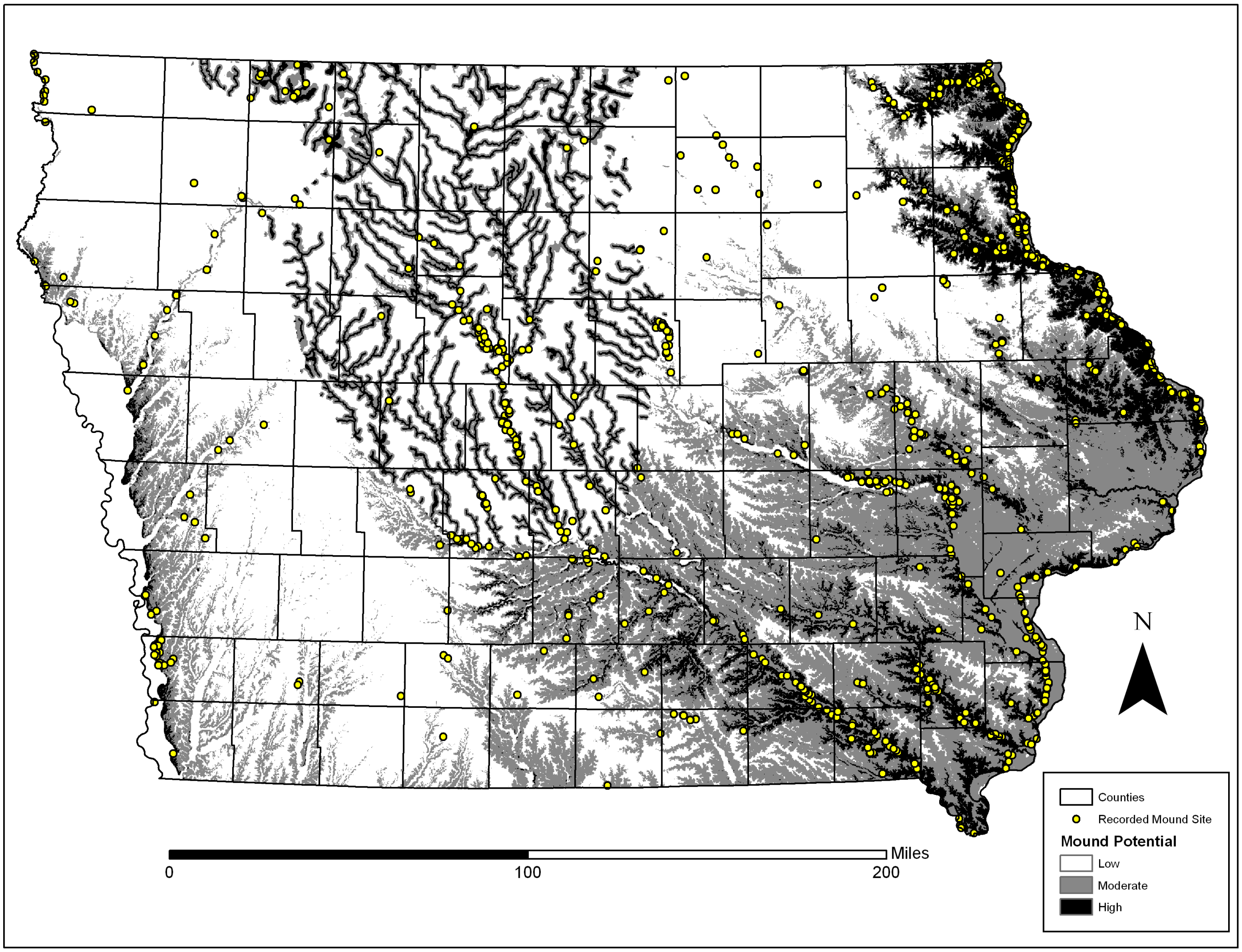

Figure 4. Map showing model results in relation to recorded mound locations.

the independent test mounds, 72 percent are located in the ten percent of the state mapped as high probability based on the model. This results in a gain of 0.86 when using Kvamme's gain statistic (Kvamme 1988). The gain statistic is useful for showing how much the model improves mound location predictability over chance models. A value of zero means the model is no improvement over chance and a value of one means the model is near perfect based on the data used.

Discussion

The model presented here departs from other site location models in Iowa archaeology in that it is an objective statewide model for one specific type of archaeological site: mounds. Furthermore, it is the first to take in to account the geographic aspect of archaeological site data and incorporates techniques to deal with spatial autocorrelation and its effects on statistical models.

The model was tested using known mound locations that were not used to build the model itself. This means the test mounds were essentially independent of training mound points. Many of the previous models in Iowa were tested with same data that was used to build the model. While low sample sizes or other restraints may play a factor in this, testing with independent data further validates the model and demonstrates its more predictive nature.

Three useful aspects of this model are (1) it is objective; (2) it has been independently tested; and (3) it exists in map form for the entire state of Iowa. Pre-GIS models did not have the capacity to provide statewide results. Modern GIS technology allows this model to be used by any person or organization with GIS capabilities. The 30-x-30-meter grid can be placed over any reference layers allowing the user to determine the likelihood of any location to contain prehistoric mounds based on the samples, variables, and test results of the model.

Archaeologists with experience in Iowa may have their own cognitive models of mound location that may well predict the likelihood of mounds to exist at a particular location. However, planners, developers, state and federal organizations are not able to do this. The model presented here could be used as a guide for archaeologists and non-archaeologists alike to increase awareness of the potential for these important cultural resources when planning for surveys or development. Field reconnaissance work could then quickly determine the outcome.

GIS layers that show current and historic vegetation for Iowa can be used with this model. A simple overlay of the model and these vegetation layers can show the few areas in Iowa that have not in been in agricultural production which may be more likely to have unrecorded mounds still intact today. In a similar vein, the state of Iowa has also recently begun acquiring LIDAR data. LIDAR provides very detailed mapping of elevations across the state. Riley (2009) has developed a way of identifying conical mounds from this data. As Riley shows, one current problem with LIDAR data is it is very cumbersome and processing of the data is time-consuming. But where the data is available, this model used in conjunction with Riley's method could prove to be very efficient and powerful.

Conclusions

Careful statistical analysis of geographic and environmental variables derived from existing datasets has resulted in a GIS-based mound location model that may be highly useful in new situations in the world of cultural resource management. The model avoids the drawbacks of earlier modeling approaches and has the major advantage of statewide applicability. Of the approximately 10,000 mounds in Iowa that Keyes (1928) estimated, many are lost. Use of this model can help ensure the preservation of the remaining mounds in the state as well as aid the identification of previously unknown mounds.

Acknowledgements

I would like to thank the following people for offering comments on this project as I have worked on it over the last five years: Shirley Schermer, Ken Kvamme, Johnathan Buffalo, Dan Higgenbottom, Derek Lee, Art Bettis, Melanie Riley, Mike Perry, and Beth Goings.

References Cited

Abbott, Larry R.
1982 *A Predictive Model for Location of Cultural Resources, Mines of Spain Area, Dubuque County, Iowa*. Contract Completion Report 200. Office of the State Archaeologist, University of Iowa, Iowa City.

Abbott, Larry R., and Joyce McKay
1978 *A Cultural Inventory for the Crow Creek Expanded Flood Plain Information Study, Scott County, Iowa*. Contract Completion Report 140. Office of the State Archaeologist, University of Iowa, Iowa City.

Benn, David W. (editor)
1987 *Big Sioux River Archaeological and Historical Resources Survey, Lyon County, Iowa*, Vol. 1. Center for Archaeological Research, Southwest Missouri State University, Springfield.

Chadderdon, Thomas J.
2003 *Phase IA Archaeological Modeling and Phase I Archaeological Survey. Page, Fremont, and Montgomery Counties, Iowa: Cultural Resource Investigations for the Page 1 Rural Water District Add on Users*. The Louis Berger Group, Marion, Iowa. Submitted to Page 1 Rural Water Group. Copy on file, Office of the State Archaeologist, University of Iowa, Iowa City.

Gallant, John P., and John C. Wilson
2000 Primary Topographic Attributes. In *Terrain Analysis: Principles and Applications,* edited by John P. Wilson and John C. Gallant, pp. 51–85. John Wiley & Sons, New York.

Goings, Chad A.
2003 A Predictive Model for Lithic Resources in Iowa. *Plains Anthropologist* 48:53–67.
2005 Modeling Prehistoric Site Locations in Iowa. Paper presentation at the 2005 Iowa Geographic Information Council Conference in Ames, Iowa.

Keyes, Charles
1928 The Hill-Lewis Survey. *Minnesota History* 9(2):96–108.

Kvamme, Kenneth L.
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. James Judge and Lynne Sebastian, pp. 325–428. U.S. Department of the Interior, Bureau of Land Management, Denver.

Lensink, Stephen C.
1984 A Quantitative Model of Central-Place Foraging among Prehistoric Hunter-Gatherers. Unpublished Ph.D. dissertation, Department of Anthropology, University of Iowa, Iowa City.

Leopold, Luna B., M. Gordon Wolman, and John P. Miller
1964 *Fluvial Processes in Geomorphology.* Dover Publications, New York.

Mallam, R. Clark
1976 *The Iowa Effigy Mound Manifestation: An Interpretive Model.* Report 9. Office of the State Archaeologist, the University of Iowa, Iowa City.

Parrish, Stephen Wallace
1998 Middle/Late Woodland Settlement Systems: An Analysis of Prehistoric Locational Behavior in the Central Des Moines Valley. Unpublished Master's Thesis, Department of Anthropology, Iowa State University, Ames, Iowa.

Riley, Melanie A.
2009 Automated Detection of Prehistoric Conical Burial Mounds from LIDAR Bare-Earth Digital Elevation Models. Unpublished Master's Thesis, Department of Geology and Geography, Northwest Missouri State University, Maryville, Missouri.

Schermer, Shirley J.
1982 Environmental Variables as Factors in Site Location: a Study of Woodland sites in the Iowa River Valley. Unpublished Master's thesis, Department of Anthropology, University of Iowa, Iowa City.

Schermer, Shirley J., and Joseph A. Tiffany
1985 Environmental Variables as Factors in Site Location: An Example from the Upper Midwest. *Midcontinental Journal of Archaeology* 10:215–240.

Tiffany, Joseph A., and Larry R. Abbott
1982 Site Catchment Theory: Applications to Iowa Archaeology. *Journal of Field Archaeology* 9:313–322.

Zimmerman, Larry J.
1977 *Prehistoric Locational Behavior: A Computer Simulation*. Report 10. Office of the State Archaeologist, University of Iowa, Iowa City.

Notes

1. A jackknife procedure can be run in statistical applications by systematically removing one member of the sample at a time to find the bias and standard error in the overall outcome. In this case, coefficients were generated for a logistic regression. Instead of using the entire sample, a jackknife procedure shows the influence that each member of the sample is having on the overall outcome. This can help in determining outliers or certain values in the sample that are heavily biasing the outcome.
2. Non-linear effects: In this case, non-linear effects occur when an independent variable does not vary in a linear fashion between where mounds are found and where they are not found in the environment. For example, as you move systematically away from a major river, the probability of encountering a prehistoric mound should not decrease in a linear fashion. It should be high adjacent to the major stream and then dramatically decrease at a certain distance from the major stream. This can be taken in to consideration by using polynomials and testing to see if that increases the t-value which gives a better result.
Interaction effects: It is possible that combining independent variables may result in a better model than considering them separately. In other words, by combining relief and slope, one might find that in high relief areas, low slopes are more important. This can be tested by using the products of two independent variables and seeing how that affects the t-value. If there is an increase in the t-value, it might be worth using the product instead of using each separately.
3. i.e., normally distributed, in other words, a histogram distribution that is symmetrical.