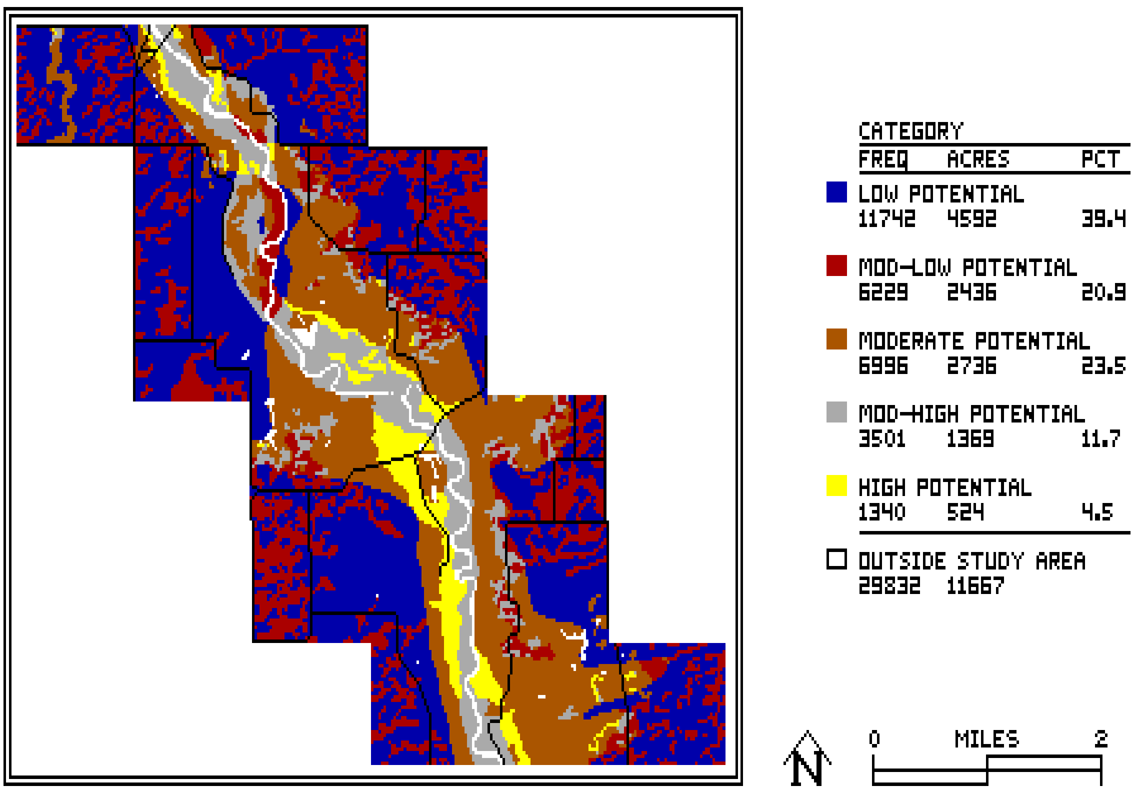
GIS Modeling of

Archaeological Sites in the Raccoon River Greenbelt

Dallas County, Iowa

Phase 2: Minburn Unit



for the Dallas County Conservation Department
Perry, Iowa

by the Department of Landscape Architecture
Department of Agronomy
Land Use Analysis Laboratory
Iowa State University

December 1996

GIS Modeling of Archaeological Sites in the Raccoon River Greenbelt

Dallas County, Iowa

Phase 2: Minburn Unit

prepared for
Jeff Logsdon and Donna Howe
Dallas County Conservation Department

prepared by
Paul F. Anderson
Department of Landscape Architecture and Department of Agronomy
Land Use Analysis Laboratory Iowa State University
December 1996

Abstract:

The Raccoon River Greenbelt was created in 1989 to develop and implement management policies and plans for cultural and natural resources. Concern for preserving archaeological sites in the Greenbelt increased recently due to extensive flooding in 1993 and rapid urbanization in the past decade. In 1993, the Dallas County Conservation Department initiated an archaeological study to document known sites and locate additional sites.

A GIS database, developed primarily from 1989 to 1992, was used for modeling 106 known prehistoric archaeological sites in the Greenbelt. The raster database, which covered 208 square miles, included data on land cover, soils, elevation, and significant sites (plant, animal, historic, archaeologic, geologic, and hydrologic).

In previous research (Phase 1), descriptive models were developed for 85 known sites along the South Raccoon River corridor. Two predictive models were developed based on the results of descriptive modeling. Model 2 showed a 55.7 percent improvement over chance. Model 2 showed a 51.4 percent improvement over chance. These two models were later used to locate archaeological sites in the Minburn Unit of the Greenbelt along the North Raccoon River corridor.

In Phase 2, an additional 21 sites were located by the archaeological survey field team in the Minburn Unit. Descriptive modeling compared characteristics of the additional 21 known sites with those in a random sample of non-sites. Measures of frequency, cumulative percentage, Chi-square, significance, and areal correspondence identified four variables on which to base additional predictive models: proximity to stream confluences, proximity to stream valleys, soil landscape position, and historic vegetation.

Predictive modeling used logistical multiple linear regression (logit modeling) techniques to identify areas with high potential for additional sites. Results of three new models were compared using measures of mean, significance, cumulative percentage, percentage correctly classified, and improvement over chance. Results from Model 4 showed a 26.0 percent improvement over chance. Results from Model 5 showed a 46.0 percent improvement over chance. Model 6 was then developed in an effort to increase the improvement over chance. Descriptive modeling of the 21 additional known sites compared their characteristics with those of 37 *known* non-sites (sites examined by the field team in which no archaeological resources were found). Results from Model 6 showed a 44.3 percent improvement over chance, slightly lower than Model 5.

Results from all the predictive models show that it is important to customize descriptive and predictive models to each part of the Greenbelt. Because the North Raccoon valley has markedly different landscape characteristics than the South Raccoon valley, models customized for the North Raccoon valley showed greater improvement over chance. County Conservation Department officials plan to extend the survey and modeling work to the remainder of the Greenbelt to assist landowners and local officials in making decisions about cultural resource preservation and landscape management.

GIS Modeling of Archaeological Sites in the Raccoon River Greenbelt
Dallas County, Iowa

I. Introduction

A. Need for research

Knowing the location of cultural resources is important information in developing landscape resource management plans. Location information is often more difficult to obtain for archaeological resources than for other types of cultural resources. For many types of archaeological sites, little physical evidence currently remains visible. Through time, artifacts and other evidence are removed or buried by natural forces or human activity (Schiffer 1987). Examples of natural forces include soil erosion and deposition, ground movement, vegetation growth and change, and wildlife burrowing (Bettis 1988). Examples of human activities include crop production, construction of roads and buildings, and collection of artifacts.

A second difficulty in locating archaeological sites is recognizing physical clues even when they are clearly visible. A trained archaeologist identifies archaeological sites through convergence of evidence: landform, surface characteristics, proximity to other features in the landscape, and cultural patterns. Evidence may include pieces of physical artifacts. Without this training, most land owners, managers, and users would be unable to recognize archaeological resources.

A third difficulty in locating archaeological sites is the limited availability of the information. A protection and preservation strategy often used by cultural resource managers is to limit distribution of information about archaeological site locations. Often, the exact locations of archaeological sites are protected by law. The purpose is to minimize the chances that people will remove evidence or otherwise disturb the sites.

These three factors mean that the locations of archaeological sites are usually not common knowledge. Therefore, to aid land managers in protecting the quality and quantity of existing archaeological sites, systematic studies involving field surveys by professional archaeologists are necessary to map the locations of archaeological sites.

Knowledge of the locations of archaeological sites is especially critical when an area is threatened by major changes. Major changes in Dallas County, Iowa are of two types: natural and cultural. Natural changes have occurred recently due to major flooding along the three branches of the Raccoon River (Figure 1). In 1993, many reaches of the Raccoon River in Dallas County flooded large portions of the floodplain (Fruhling 1994). In some reaches, it was the flood of record. In other reaches, the flood level was equivalent to the 500 year flood frequency. Major cultural changes include urban expansion and development. These have occurred since the 1970s, when several Des Moines suburbs began annexing land in Dallas County. In that same time period, many new rural subdivisions were built in Dallas County. Additional proposals for beltway highways and a gambling casino have recently been announced. In these situations, construction of buildings, homes, streets, roads, and accompanying recreation facilities can remove, disturb, or destroy archaeological sites. In other situations, sites become less accessible because of paving or building over them.

Because flooding, urban expansion, and other agents of landscape change are likely to continue in the future, the Dallas County Conservation Department is developing additional strategies to manage natural and cultural resources, including archaeological sites. In 1989, Dallas County began developing a resource management plan for the Raccoon River Greenbelt. In 1990, the county instituted an environmental review process as part of zoning change requests.

In 1993, the county began a study of archaeological sites. This study was funded by the county and the Iowa Resource Enhancement and Protection Act, through the Historical Resources Development Program of the State Historical Society of Iowa. The funding was used to support the following:

1. Study of existing archaeological sites through archival sources
2. Field survey of additional archaeological sites through surface and excavation techniques
3. Study of existing and potential archaeological sites through use of geographic information systems (GIS) technology and statistical modeling

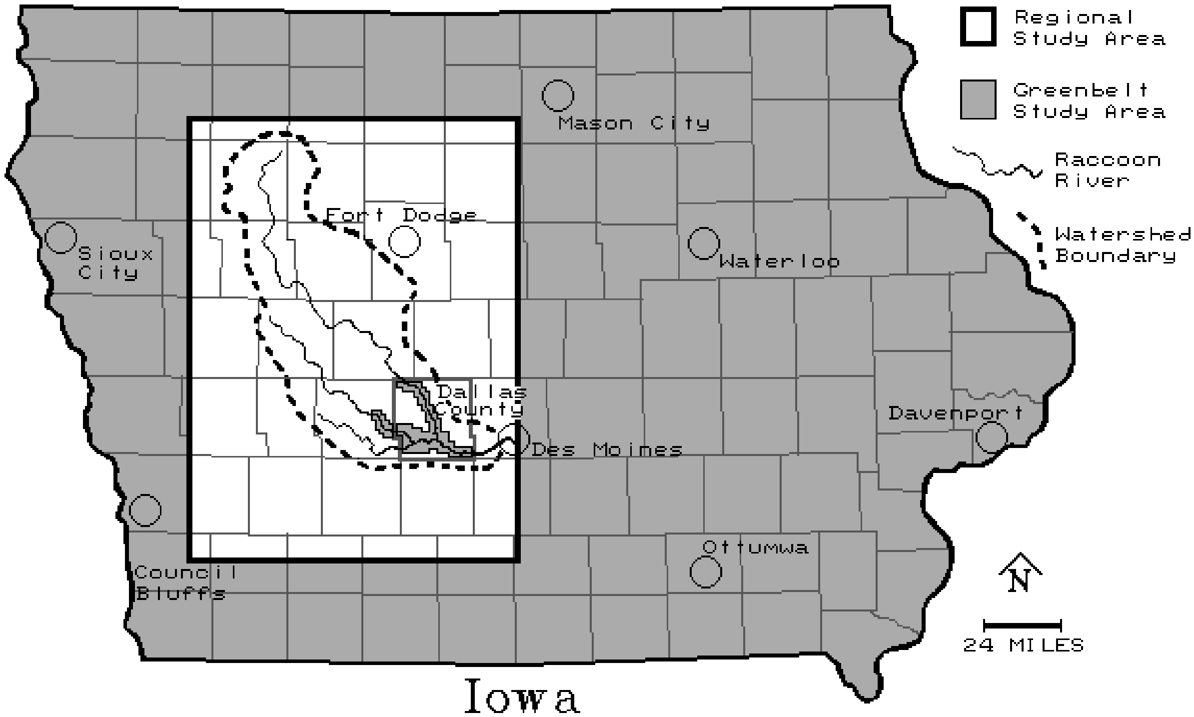


Figure 1. Location of the Raccoon River basin, three Raccoon River branches, and Dallas County

B. Research objectives and hypotheses

This report describes the third study listed above involving GIS technology and statistical modeling. Information from county employees, archaeologists, and a geomorphology consultant working on the first two studies became inputs into the third study. Three objectives of this third study included the following:

1. Document landscape characteristics of known archaeological sites through descriptive modeling
2. Help locate additional archaeological sites through predictive modeling
3. Provide context for landscape planning and management decisions

In previous research (Phase 1), information developed in the first two studies listed above covered much of the southern tier of townships in Dallas County. This area of approximately 150 square miles contains much of the south branch of the Raccoon River and its confluences with the north branch and the middle branch. It also contains several areas owned and managed by the Dallas County Conservation Department, the largest of which are Kuehn Conservation Area and Hanging Rock Conservation Area. Also, the southern tier of townships is where most of the urban expansion and rural non-farm development are taking place. For these reasons, the Dallas County Conservation Department started their archaeological studies in the southern part of the county, concentrating on prehistoric sites because less was known about prehistoric sites than historic sites.

In the current research (Phase 2), archaeological surveys and modeling were extended into the Minburn Unit along the North Raccoon River. This area of approximately 18 square miles contains the Voas Nature Area, Spring Valley Access, Snyder Access, and Crellin Wildlife Refuge. This portion of the Greenbelt has a younger landscape (geologically) than the study area for Phase 1 of the research.

Hypotheses for Phase 2 of the research included the following:

1. Descriptive models which compare characteristics of known sites to *known non-sites* (rather than to a random sample of non-sites) result in predictive models which have greater predictive power (as measured by the logit modeling improvement over chance)

2. Predictive models customized specifically for the Minburn Unit (in the North Raccoon portion of the Greenbelt) have greater predictive power than those developed during Phase 1 for the South Raccoon portion of the Greenbelt

Several Dallas County departments are planning to use information on known and potential prehistoric archaeological sites to plan and manage the Raccoon River Greenbelt, review development plans and rezoning requests, develop management plans for watersheds and river basins, and develop resource management plans for individual landowners.

C. Participants

This research was prepared under contract to the Dallas County Conservation Department: Jeff Logsdon (director) and Donna Howe (project manager). Archaeological consultants to the Dallas County Conservation Department included Cindy Peterson, John Doershuk, and Fred Finney of the Iowa Office of the State Archaeologist. Geomorphology consultant to the Dallas County Conservation Department was Rolfe Mandel of the University of Nebraska. GIS data entry, mapping, modeling, and statistical analysis were completed by Paul Anderson of Iowa State University.

Funding was provided by Dallas County, the Iowa Resource Enhancement and Protection (REAP) program (through the Historic Resources Development Program of the State Historical Society of Iowa), and Iowa State University.

II. Previous work

A. Raccoon River Greenbelt GIS

In 1989, the Dallas County Conservation Department began a resource master planning process for the Raccoon River Greenbelt. The purpose of the planning effort was to guide wise use of the valley and adjacent corridor for the three branches of the Raccoon River in Dallas County. The planning effort identified areas for preservation, conservation, restoration, and development. Planning decisions were based on surveys of local residents' needs and attitudes, existing landscape uses and characteristics, and future opportunities and limitations of the area's cultural and natural resources. The plan also assessed recreation supply and demand within the county and surrounding region.

The database developed for the Raccoon River Greenbelt Master Plan included ten GIS data layers:

1. Soils
2. Land cover/land use
3. Transportation
4. Elevation
5. Slope aspect
6. Zoning
7. Public lands
8. Private conservation lands
9. Utilities
10. Significant sites

Significant sites included habitats of rare and endangered plants and animals. Significant sites also included known archaeological, historical, geological, and hydrological sites.

GIS data layers were prepared in raster (cell) format from a combination of published and unpublished sources. Each of the approximately 351,000 rectangular cells was defined by 1.5 arc seconds of latitude and longitude (151.9 feet by 112.9 feet at 42 degrees latitude). Each cell covered an area of approximately 0.39 acres per cell (0.158 hectares). Data sources (see Appendix A) were compiled on map tracings, then optically scanned using a DataCopy digital scanner, then labeled using Map Editing System software at the Iowa State University Land Use Analysis Laboratory. Data were then prepared in a raster format for Geodesy, a locally-written DOS software package based on concepts of map overlay and map arithmetic (Anderson 1992). It was then converted for use with Map II, a Macintosh GIS package (Pazner and others 1989).

The GIS database covered a study area of approximately 208 square miles (137,000 acres or 55,500 hectares). This study area was defined by first including all floodplain areas adjacent to the Raccoon River, then by adding the adjacent areas with valley walls and dense woodland, finally by adding adjacent areas within the same sections of land. In other words, the boundary of the GIS study area followed section lines. This defined the GIS study area as a corridor along the Raccoon River from 3 to 5 miles wide, depending on the width of floodplain, valley walls, and dense woodland (Figure 2).

The GIS study area was divided into 11 units for more detailed study and GIS analysis. GIS analysis models included inventory maps for each data layer and interpretation maps of land cover/land use change, river viewshed, landscape position, agricultural suitability, wildlife habitat potential, woodland restoration potential, wetland restoration

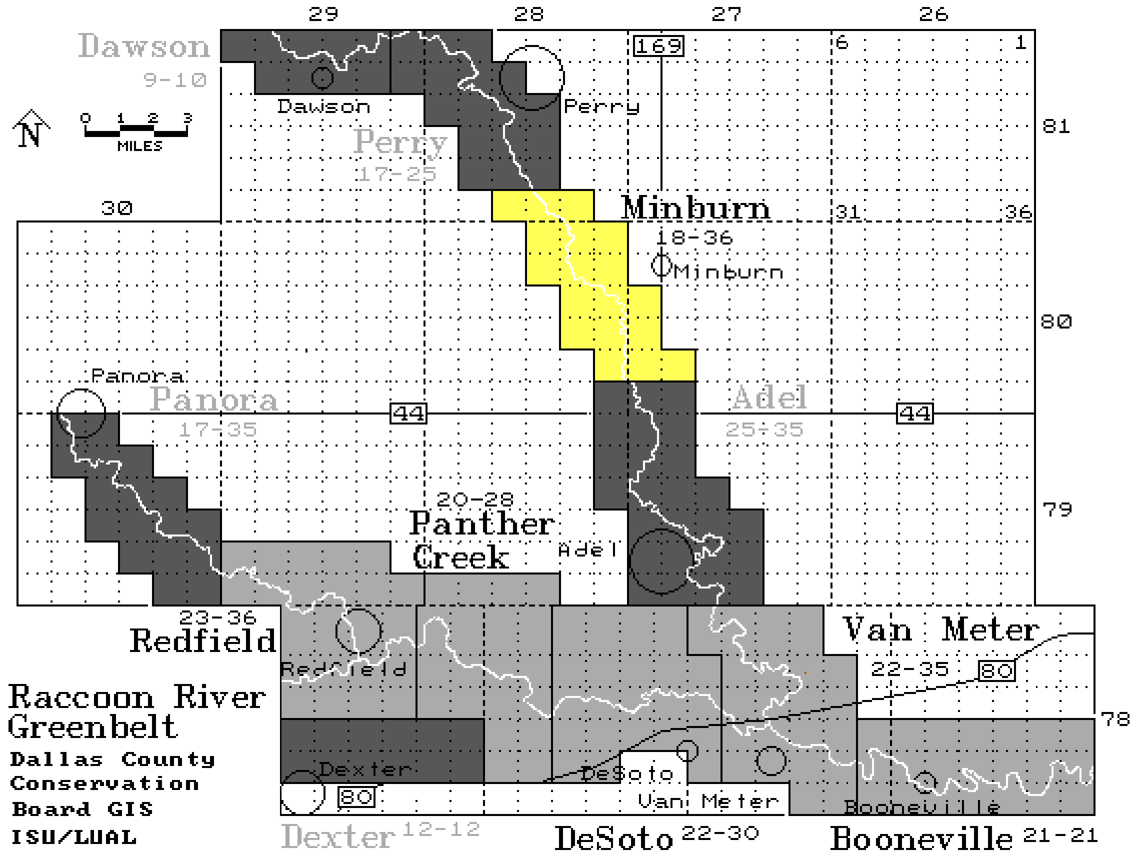


Figure 2. GIS study area for the Raccoon River Greenbelt, Dallas County

potential, prairie restoration potential, active recreation potential, composite recreation potential, and critical resource areas.

Five GIS data units were selected as the study area for the first archeological site modeling study in Phase 1 (1994-1995): Redfield, Panther Creek, DeSoto, Van Meter, and Booneville (Figure 2). These were selected because they cover the portion of the southern tier of townships used for the first archaeological field survey. In Phase 2 (1995-1996), the field survey and GIS modeling study were extended into the Minburn Unit along the North Raccoon River.

B. GIS predictive modeling

Earlier GIS predictive modeling in Iowa centered on habitats of plant and animal species. Dean Roosa, former state ecologist with the Iowa Preserves Department and the Iowa Department of Natural Resources, located potential sites in northeast Iowa for the plant species northern wild monkshood (*Aconitum noveboracense*) using the statewide GIS database and MSDAMP software at the Iowa State University Land Use Analysis Laboratory (Beavers 1977). Using the same GIS database and software, Bednarz (1979) located potential nesting sites for red-shouldered hawks (*Buteo lineatus*) in Iowa. In each of these studies, the scientist provided a description of habitat requirements for the species. This was translated into a GIS numerical model of habitat potential. The process involved identifying data layers (variables) that relate to species habitat requirements, ranking the importance of each data layer, then rating the categories within each data layer based on their relative potential using an ordinal measurement scale.

Kvamme (1990, 1992) was one pioneer in the use of GIS predictive modeling for archaeological sites. His models were based on a form of multiple linear regression called logit models. Logit models are appropriate when "the independent variables are categorized using a nominal or ordinal measurement scale rather than an interval or ratio scale" (Wrigley 1976, p. 3). Logit models are based on Boolean logic and result in a dichotomous (binary) answer--yes/no--which indicates whether there is a high or low probability. Kvamme's application of logit models to predictive modeling of archaeological sites involved development of a multiple linear regression equation based on Chi-square and other measures of the differences in the geographic and landscape characteristics between known archaeological sites and non-sites (the control group). It compared probability scores at known sites with scores at non-sites to identify a cutpoint score (often expressed as a decimal number from 0 to 1) at which the combination of correctly classified known sites and correctly classified non-sites was maximized. The percent of known sites scoring at or above the cutpoint score was then compared to the percent of the entire study area classified as high potential. This yielded a measure of the improvement in predictive power of the model over chance (random selection of sites). The result is what statistical geographers call a “probability surface model,” which is appropriate for categorized data (Wrigley 1977). “Such an approach allows analysis of the large amount of archaeological data which can only be assessed as presence/absence (the presence or absence of a site or a particular artifact type)” (Hodder 1979, p.192). “Past applications of logistic regression [logit modeling] have produced slightly better predictive results than other multivariate classification techniques such as discriminant analysis” (Kvamme 1988 quoted in Carmichael 1990, p. 221).

C. Landscape archaeology

Much of the archaeological record in Iowa has resulted from cultures inhabiting the region since the last major ice sheet melted 11,000 to 13,000 years ago. Some of the earliest evidence of prehistoric Native Americans in Iowa are Clovis stone spear or projectile points (Gradwohl 1978, p. 37). These were used during a time period called the Paleo-Indian Period or Big Game Hunting Tradition. Later cultures left evidence of hunting, habitation, incipient agriculture, burials, horticulture, agriculture, food storage, trade, and transportation.

Iowa dates	Culture-historical period	Dallas County dates
6,000 to 12,000 BC	Paleo-Indian	7,500 to 8,000 BC
500 to 6,000 BC	Archaic	500 to 7,500 BC
500 BC to 1000 AD	Woodland	500 BC to 1000 AD
1,000 to 1,600 AD	Post-woodland (Late Prehistoric)	1,000 to 1,600 AD

Two major types of sites identified by archaeological investigations include burial sites and activity sites. Activity sites include habitation, agriculture, hunting, and ceremonial sites. Features of the landscape influenced activities and evidence left by these people. Water, soil, topography, and vegetation were major influences on climate, drainage, and location of wildlife. These presented both hazards and resources which varied over space and time. Space was not homogeneous or of uniform value for many of these activities, such as hunting (Butzer 1982). Distance and concentration were major variables in deciding where to camp, hunt, plant, and bury their dead. Butzer (1982, pp. 215-229) described four major categories of models used by archaeologists to explain locations of archaeological sites:

1. Central-place theory -- permanent economic hierarchy
2. Resource concentration -- nomadic movement
3. von Thunen -- concentric rings decreasing in intensity of use
4. Gravity -- distance to populations, food, and other resources

Because of the general lack of knowledge about large, permanent settlements and dependence on both hunting and agriculture, gravity models are appropriate for descriptive and predictive models in Iowa. This was the reason that two major variables in the models described later were based on distance (proximity).

III. Methods

A. Research tools

The two major tools used in the research were statistics and GIS modeling. Statistical measures provided a way of summarizing and characterizing the patterns represented by the known archaeological sites in the study area. For example, descriptive statistics included mean, minimum, maximum, and frequency distribution of landscape characteristics such as landscape position. Statistics also provided a measure of the probability for archaeological sites. Regression models are probably the most widely used statistical models in geographic research (Wrigley, 1976, p. 3). As described earlier in Section II.B, logit models are a special form of logistic multiple linear regression models used when the independent variables (such as landscape position) are categorized.

GIS modeling provided the spatial link between the probable characteristics of potential archaeological sites and specific locations within the study area. When data layers (such as soils and hydrology) were combined using GIS software, combinations of patterns were examined for their potential and then mapped to guide future archaeological investigations in the field. GIS mapping functions useful in predictive modeling included numerical functions, logical (Boolean) functions, and geographic functions. In gravity models, a particularly useful geographic function was proximity (buffer). Proximity provided a measure of distance from significant geographic features, such as stream confluences. Proximity and other GIS functions were useful both in descriptive modeling and in predictive modeling. As used in this research, GIS descriptive modeling summarized the geographic characteristics of known archaeological sites using descriptive statistics. Predictive modeling used GIS software and data layers to locate other areas which had characteristics similar to known archaeological sites.

In addition to Geodesy for GIS modeling (Anderson 1992), other tools used in this research included software for database preparation, statistical analysis, and data display. Software packages included Microsoft Excel, Minitab, and Microsoft Word.

B. Research procedure

Four major steps (Figure 3) were used in the research procedure:

1. Obtain information on known archaeological sites
2. Create descriptive models of known sites
3. Create predictive models of potential sites
4. Use predictive models to guide additional field surveys

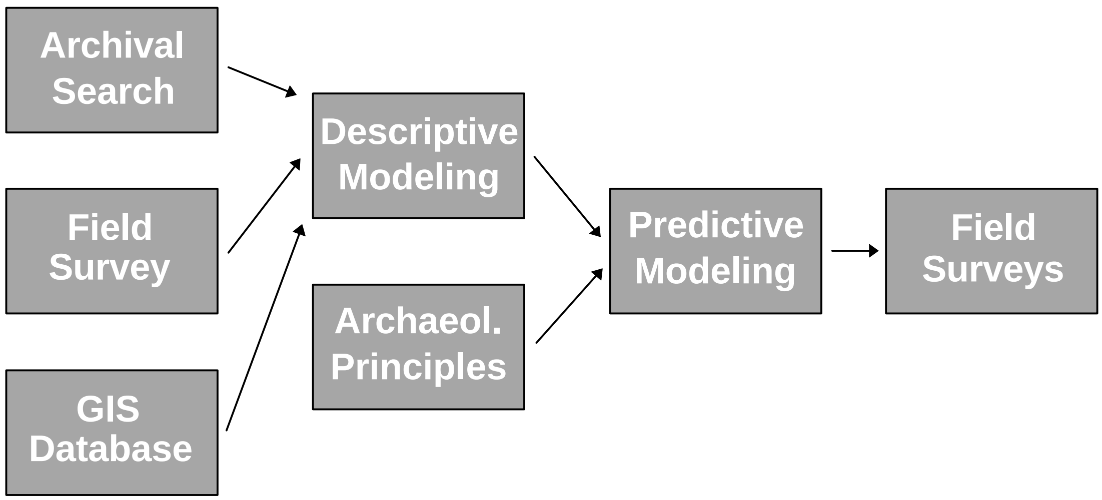


Figure 3. Research procedure

Obtain information on known archaeological sites. Information on known archaeological sites was obtained from Cindy Peterson, John Doershuk, and Fred Finney of the Office of the State Archaeologist (OSA) in Iowa City. They provided a list of known archaeological sites which included the location, brief description, and identification number of each. For some of the sites, they provided information on artifacts found at the site (such as projectile points or pottery) or type of activity at the site (such as habitation or burial). However, this artifact and activity information was not used directly in creating the models because the model did not stratify sites into categories. Information provided by Peterson, Doershuk, and Finney came from archival records at OSA and from field surveys conducted by Peterson and others as part of the archaeological study.

Peterson also provided information on geographic location and extent for each of the known sites. Known sites were outlined on USGS 7.5-minute quadrangle maps (1976-1982, scale 1:24,000) and on USDA panchromatic aerial photographs (1966-1967, scale 1:20,000). Locations of known sites were digitized by the author to create a new GIS data layer. The digitizing process involved using proportional measurement on the USGS maps, aerial photographs, and an existing GIS data layer of land cover/land use digitized from USGS color-infrared aerial photographs (1983, scale 1:58,000) and panchromatic aerial photographs (1990, scale 1:4,800). Digitizing was done using EdCell, a DOS-based GIS software package for data digitizing and editing, previously written by the author at the Land Use Analysis Laboratory. Attribute data entered for each site included the site identification number (Site ID) assigned by the OSA staff, proximity to stream confluences, landscape position, proximity to valley, soil mapping unit, 1990 land cover, 1847-51 General Land Office (GLO) historic vegetation, and native vegetation from soils. Attribute data were entered in data tables using Microsoft Excel (see Appendix B).

Create descriptive models of known sites. Descriptive models were then created by mapping landscape characteristics of the known sites. At the beginning of Phase 1 of the research, Peterson and Finney identified three landscape characteristics of the sites significant in their field surveys and in other similar studies in Iowa:

1. Proximity to stream confluences
2. Landscape position (landform)
3. Proximity to the river valley

New data layers for these three landscape characteristics were derived from the soils data layer, which was originally digitized in 1989 for the Raccoon River Greenbelt Master Plan GIS database. Soils data were digitized from the USDA soil survey of Dallas County (US Department of Agriculture 1983). Of the ten original data layers in the GIS database, the soils layer provided the best indication of stream location and landform in the study area. The new data layers for proximity to stream confluences and proximity to the river valley were derived from soils data using ProMap, a DOS-based GIS software package for proximity mapping previously written by the author at the Land Use Analysis Laboratory. The software allowed up to nine distance zones to be calculated in making new data layers. Distance zones selected for proximity to stream confluences included the following:

1. Stream confluence
2. 0.01 - 0.25 miles
3. 0.26 - 0.50 miles
4. 0.51 - 1 mile
5. 1.01 - 2 miles
6. > 2 miles

Distance zones were selected based on the size of the study area and the limitation of the software. Initially, stream confluences selected for this new data layer included only those formed by perennial streams shown with a double line on the soil survey map sheets. After viewing these initial descriptive modeling results, Peterson and Finney suggested that smaller perennial stream confluences would also be significant in predictive modeling. Therefore, the definition of stream confluences was changed to include confluences of all perennial streams shown on the soil survey map sheets, whether shown with a double-line or single-line (Figure 4).

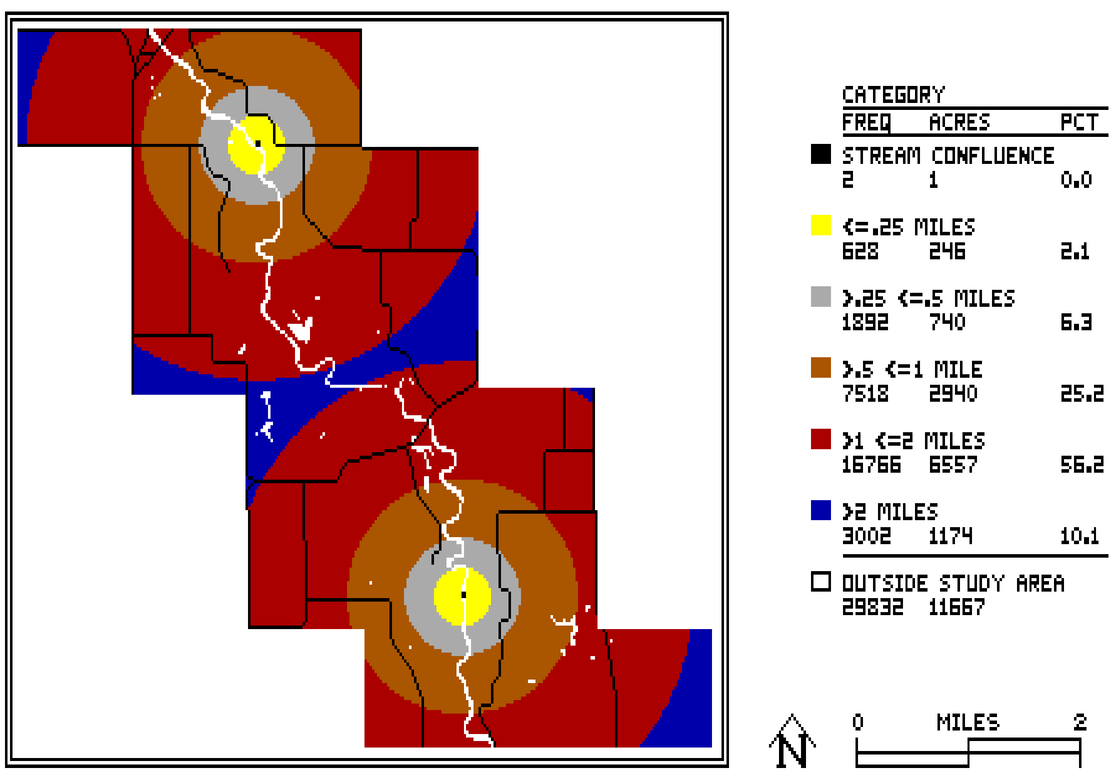


Figure 4. Proximity to stream confluences in the Minburn Unit

Distance zones selected for proximity to river valley included the following:

1. Valley
2. 1 - 200 feet
3. 201 - 500 feet
4. 501 - 1000 feet
5. > 1000 feet

For this new data layer, valley was defined as the river, adjacent floodplain, terraces, and valley walls. Using this definition, valley occupied approximately 31 percent of the study area (Figure 5).

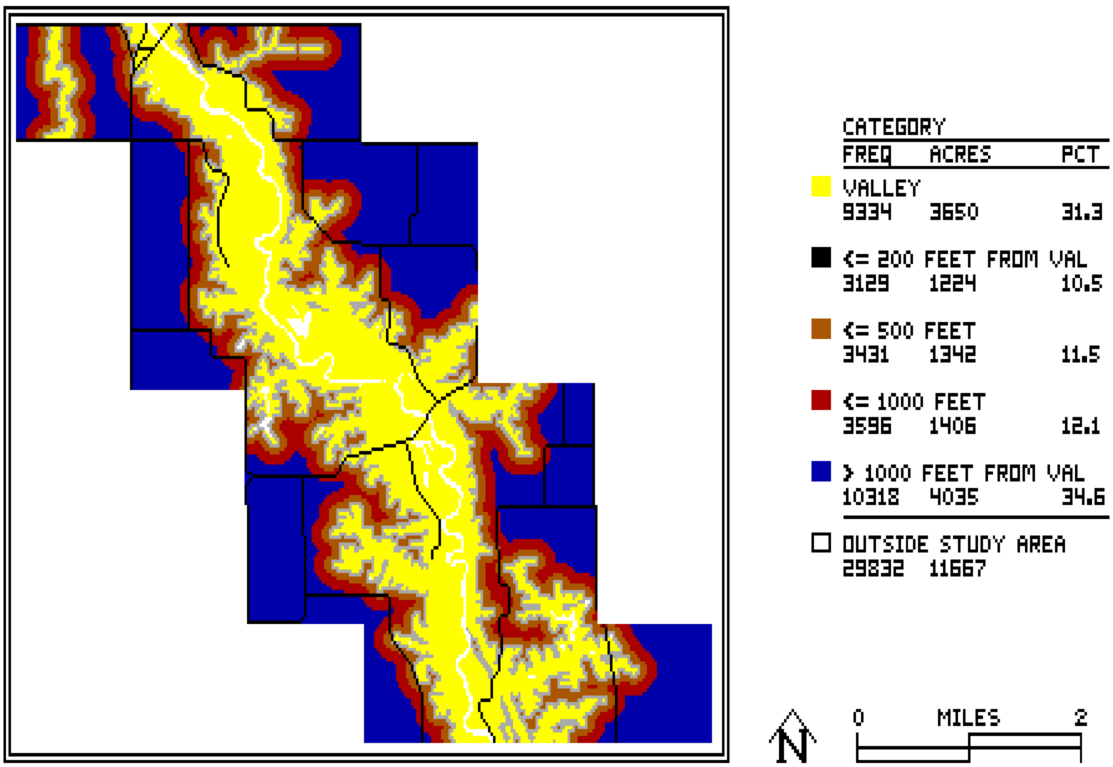


Figure 5. Proximity to river valley in the Minburn Unit

Landscape position was interpreted from the soils data layer digitized for the original Raccoon River Greenbelt database. Soil mapping unit descriptions in the report (US Department of Agriculture 1983) provided information on landscape position. Soils in the study area were aggregated into the following landscape position categories (Figure 6):

1. Alluvial fans
2. Terraces
3. Floodplain ridges
4. First bottoms
5. Wetlands
6. Valley rim
7. Valley wall, upland, and other landscape positions

In addition to proximity to stream confluences, proximity to valley, and landscape position, several other characteristics were recorded for each site. These included existing land cover and two sources of historic vegetation (GLO vegetation and native vegetation from soils). These were included in descriptive modeling because of their potential for inclusion in future predictive models. Each of these six characteristics was recorded for each of the

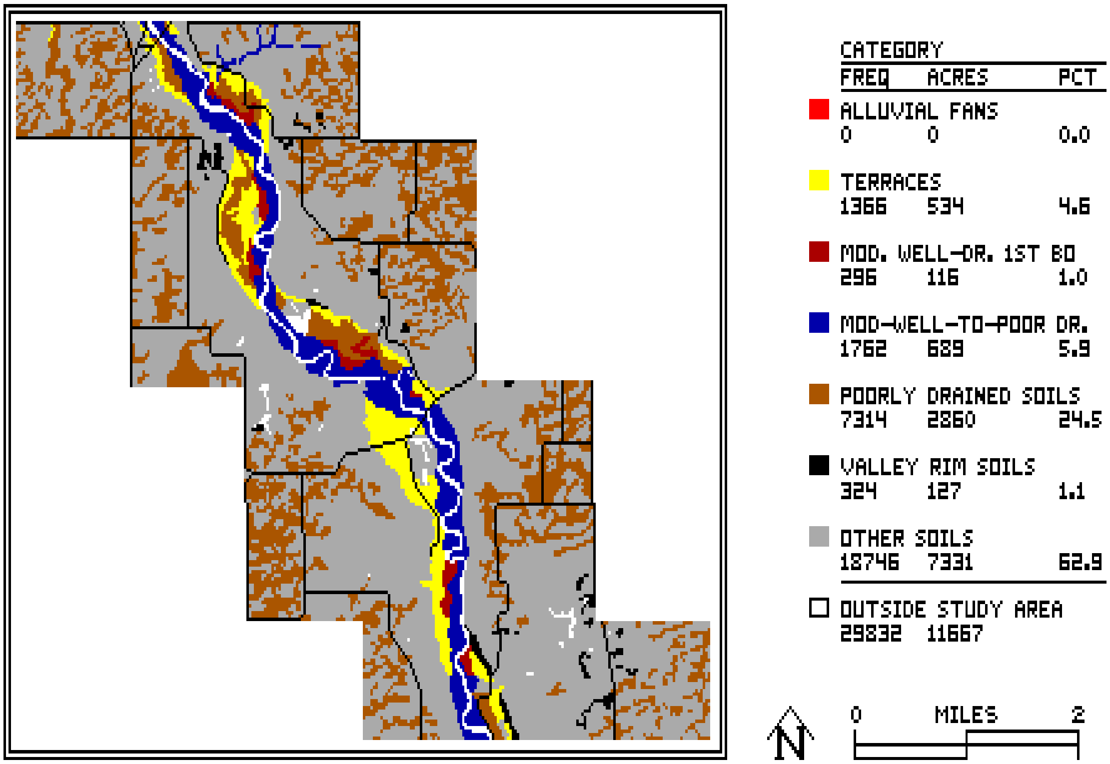


Figure 6. Landscape position in the Minburn Unit

known sites using a Microsoft Excel data table (see Appendix B). Frequency tables were then computed and graphs were drawn, also using Microsoft Excel.

Create predictive models of potential sites. Predictive models were created by mapping locations in the study area with characteristics similar to the known sites. Before descriptive modeling was completed, Finney and Peterson ranked proximity zones and landscape positions based on their potential for archaeological sites. Rankings were based on their experiences during field surveys in Dallas County and elsewhere in Iowa, on literature, and on landscape archaeology principles. An ordinal rating scale of 1 to 5 was used; a rating of 1 point indicated low potential and a rating of 5 points indicated high potential:

5 pts.	Stream confluences
5 pts.	0.01 - 0.25 miles
4 pts.	0.26 - 0.50 miles
3 pts.	0.51 - 1 mile
2 pts.	1.01 - 2 miles
1 pt.	> 2 miles

5 pts.	Terraces
4 pts.	Alluvial fans
4 pts.	Floodplain ridges
2 pts.	First bottoms
2 pts.	Wetlands
1 pts.	Valley wall, upland, and other landscape positions

5 pts.	1 - 200 feet
4 pts.	201 - 500 feet
3 pts.	Valley
2 pts.	501 - 1000 feet
1 pts.	> 1000 feet

Then, Peterson and Finney evaluated and weighted the three data layers based on their relative importance in determining potential for archaeological sites. Again, weighting was based on their experiences during field surveys in Dallas County and elsewhere in Iowa, on literature, and on landscape archaeology principles. They assigned a multiplier (weight) to each data layer:

Multiplier 3	Proximity to stream confluences
Multiplier 2	Landscape position (landform)
Multiplier 1	Proximity to river valley

Next, the multipliers and points were combined to compute a numerical score for each cell in the study area based on its combination of landscape characteristics. GIS arithmetic functions combined the multipliers and points using Geodesy, a DOS-based GIS software package for map arithmetic, previously written by the author at the Land Use Analysis Laboratory (Anderson 1992):

$$\text{Composite score} = (M^1 \times P^1) + (M^2 \times P^2) + (M^3 \times P^3)$$

Where

M^1 = multiplier for data layer 1

P^1 = points assigned to the landscape characteristic in the cell on data layer 1

Using multipliers of 1, 2, and 3 with points from 1 to 5, the maximum composite score that any cell could receive was 30 = (3 x 5) + (2 x 5) + (1 x 5). The minimum composite score that any cell could receive was 6 = (3 x 1) + (2 x 1) + (1 x 1).

Using this approach to map arithmetic, a total of six predictive models were then developed. Models 1, 2, and 3 were developed during Phase 1 of the research for 85 known sites in five data units in the southern portion of the Greenbelt (South Raccoon River corridor). Models 4, 5, and 6 were developed during Phase 2 of the research for 21 known sites in the Minburn Unit in the northern portion of the Greenbelt (North Raccoon River corridor).

Model 1 was applied to only a part of the Greenbelt (Redfield Unit) as an initial test. Based on these results, Peterson and Finney suggested several changes to the model:

1. Change definition of stream confluences to include confluences of single-line perennial streams
2. Change alluvial fans from 3 pts. to 4 pts. for landscape position

These changes were incorporated into Model 2. Model 2 was then applied to the entire Phase 1 study area. To measure the effectiveness of Model 2, techniques of logit modeling were used (Wrigley 1976). The purpose was to measure how much Model 2 improved predictive power over chance. This required comparing Model 2 composite scores at known archaeological sites with composite scores at a random sample of non-sites. The random sample represented the control group against which the known sites were compared. Pereira and Itami (1991) recommended that the number of sample sites exceed the number of known sites by a slight margin, because a larger variation was expected in non-sites than in known sites. There were 741 cells in Phase 1 known sites, so a total of 850 cells were selected for the random sample. The sampling ratio was 1 in 223 (189,262 cells in the entire Phase 1 study area). There were 143 cells in Phase 2 known sites, so a total of 136 cells were selected for the random sample in the Minburn Unit. The sampling ratio was 1 in 219 (29,808 cells in the Phase 2 study area). The sample was a random-start, systematic sample. Every fifteenth cell in every fifteenth row was sampled (Figure 7). A random number table was used to determine the row and column of the first cell in the sample (row 4, column 7). Based on the sampling pattern, if a sample cell fell within a known site, the first non-site cell to the east (along the row of cells) was selected for the sample.

For known sites, the distribution of cell scores was computed, then graphed on a percentage basis. In a similar way for the random sample, the distribution of cell scores was computed, then graphed. Then a cutpoint score was identified which maximized the number of sites correctly classified as high potential and the number of non-sites correctly classified as low potential. Identifying the cutpoint score required combining the two curves together. The peak of the resultant curve identified the cutpoint score. The peak also measured the percent improvement in predictive power that the model had over chance.

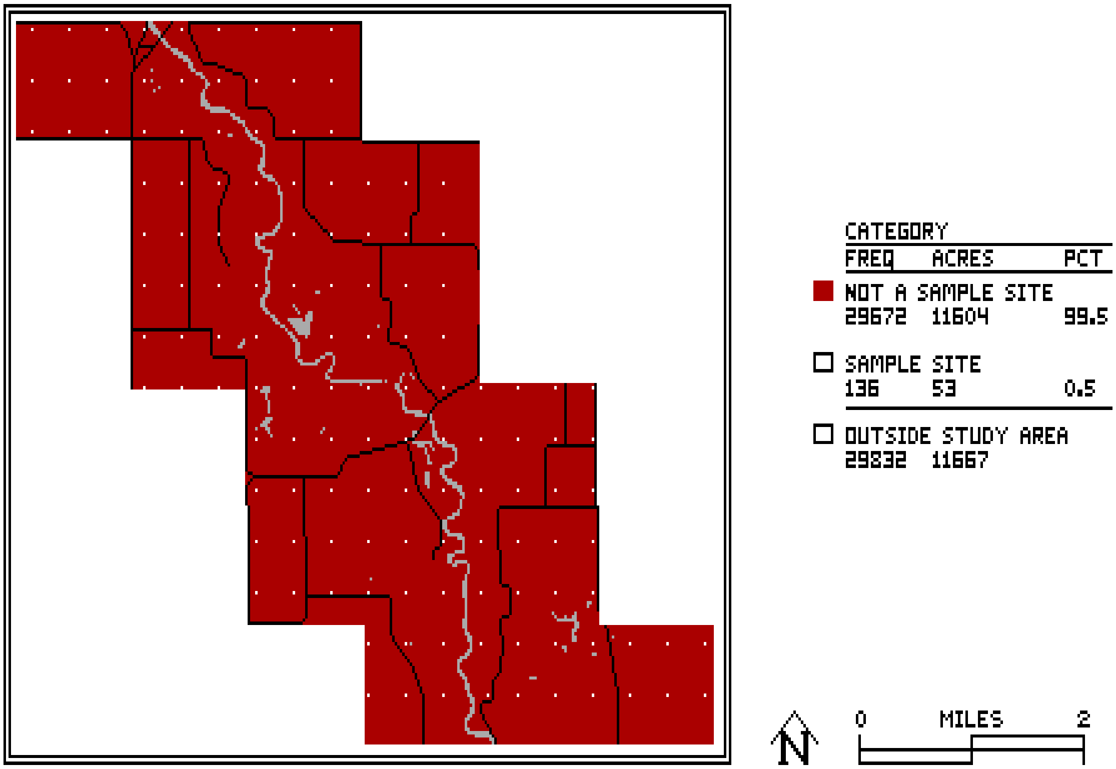


Figure 7. Sample of non-sites in the Minburn Unit

Again, Peterson and Finney inspected the results of Model 2 and suggested further refinements:

1. Change proximity to valley criteria
   a. Valley from 3 to 5 pts.
   b. 0-200 feet from 5 to 4 pts.
   c. 201-500 feet from 4 to 3 pts.
2. Change landscape position criteria
   a. Alluvial fans from 3 to 4 pts.
   b. Floodplain ridges from 4 to 3 pts.
   c. Add valley rim 4 pts.
   d. Change Nodaway soil from alluvial fan to floodplain ridge
3. Change multipliers
   a. Landscape position from 2 to 1
   b. Proximity to valley from 1 to 2

These changes were incorporated into Model 3. Model 3 was then applied to the entire Phase 1 study area. For Model 3, a cutpoint score was also computed using logit model techniques described above. A summary of the criteria (multipliers, weights, and points) is contained in Appendix C.

To further refine the model, Chi-square measures were made with the results of descriptive modeling. A relatively large Chi-square measure indicated a large difference between known sites and non-sites. This suggested an important distinction between the two groups and, therefore, a higher weight (multiplier) in the predictive model.

To further refine the model, Peterson and Finney considered adding another variable: historic vegetation. To aid in their decision, two new data layers for historic vegetation were made (Figures 8 and 9):

1. Vegetation from Government Land Office (GLO) township plat maps (see Anderson 1996a)
2. Native vegetation (an interpretation from the existing soils data layer)

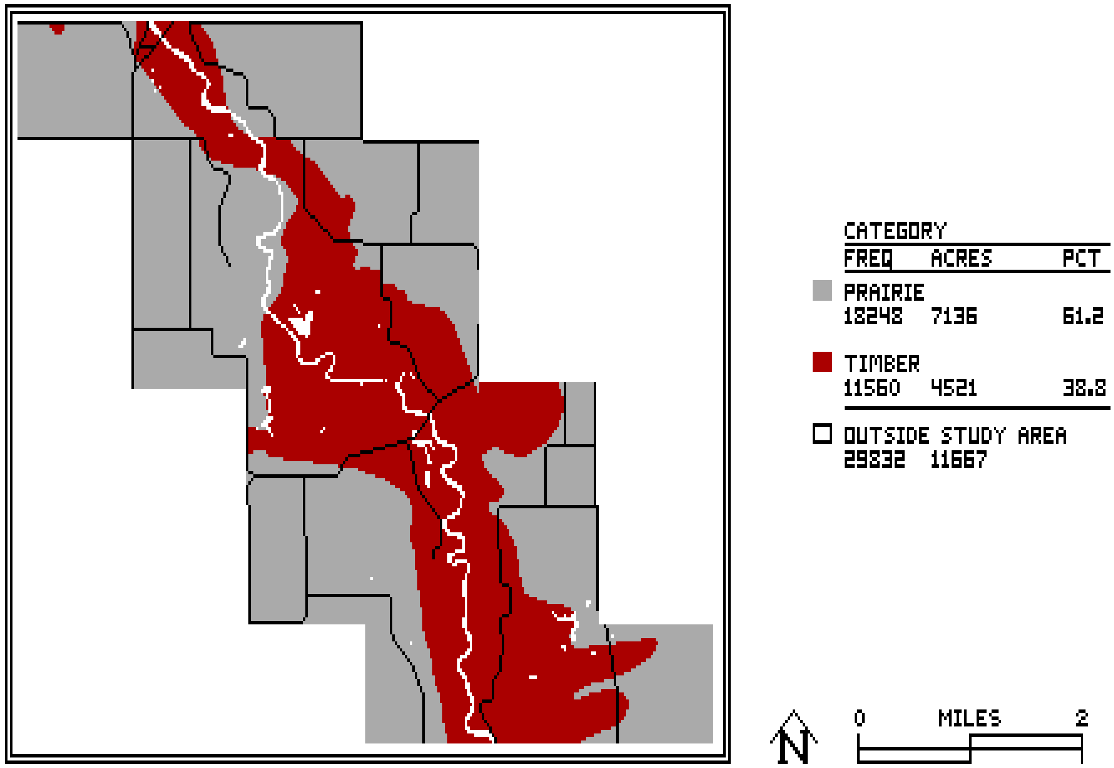


Figure 8. GLO vegetation in the Minburn Unit

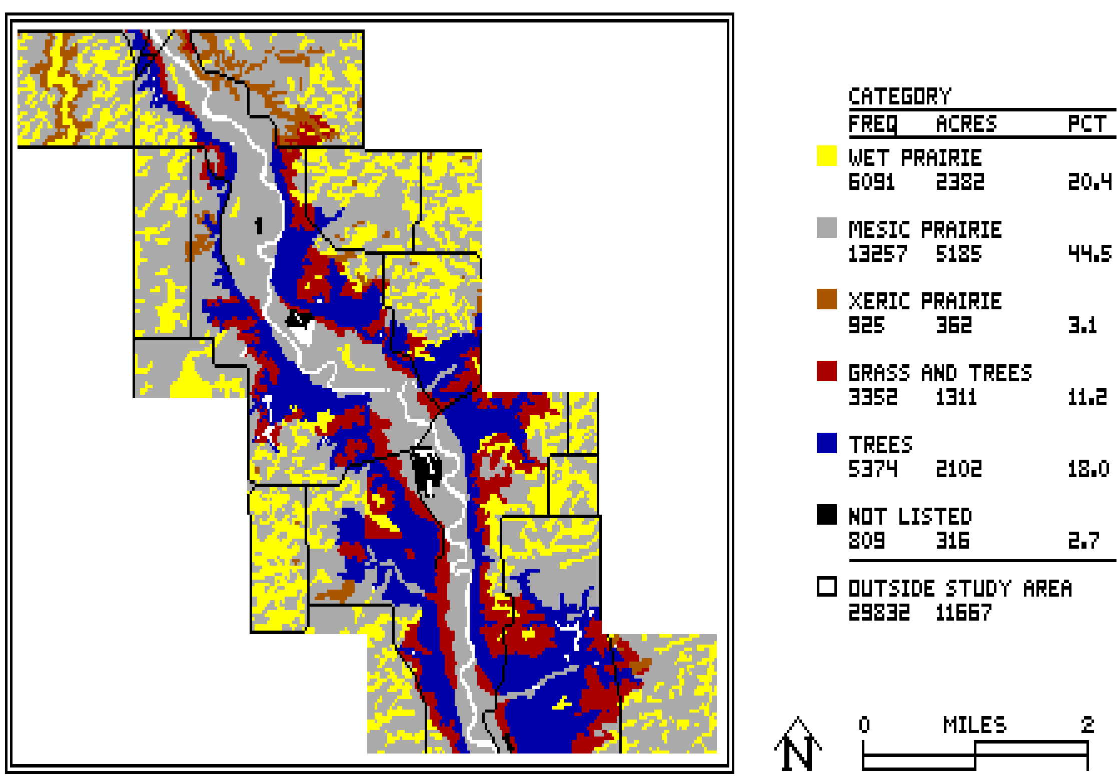


Figure 9. Native vegetation in the Minburn Unit

These two new data layers were first compared to each other. This helped measure the similarities and differences between the two vegetation patterns. This comparison also measured the similarities to the three variables already included in the model. The statistical measure used for this purpose was the Coefficient of Areal Correspondence (CAC). CAC measures the amount of overlap (spatial intersection) between two sets of polygons (Unwin 1981, p. 189-192; Minnick 1964). For example, if A^u represents the union of the two areas (total area covered by woodland or native vegetation) and A^i is the intersection of the two areas (area of overlap), then the CAC can be computed:

$$CAC = A^i / A^u$$

CAC can be expressed as a decimal or as a percentage. CAC values closer to 1.0 (100%) indicated a high amount of overlap between the two sets of polygons and, therefore, similarities both in quantity and spatial distribution. This was useful in making sure that any new variable added to future models was significantly different from one already used in the models (to avoid spatial autocorrelation).

Phase 2 of the research began by applying criteria from Model 2 and Model 3 to the Minburn Unit and Adel Unit in the northern portion of the Greenbelt (North Raccoon River corridor). To help guide field surveys, GIS maps of the models were prepared which showed the results of both models on one map (Figure 10). In this way, the field team could more easily refer to the results of both models as they conducted their field surveys. The GIS maps were printed on transparency material at a scale of 1:24,000; this allowed them to be placed on USGS 7.5-minute quadrangle maps used in the field by the survey team. To accomplish this, GIS raster images in PCX format were first created using the "save PCX" feature in the Geodesy GIS software (Shift-<F7>). Each PCX image was imported into AutoCAD, where it was scaled to match the USGS quadrangle maps and plotted on transparency material in color. Using a similar procedure, maps of landscape position were also produced to aid the survey team in their field work (Figure 11).

After the field surveys were completed during the spring and summer of 1996, Peterson provided information about the location and extent of 21 additional sites in the Minburn Unit. Though the field team surveyed the Adel Unit also, only one site was located there. Team members said that urban and agricultural development in the vicinity of the City of Adel eliminated archaeological evidence or limited access to it. They decided to limit descriptive and predictive modeling work in Phase 2 to the Minburn Unit.

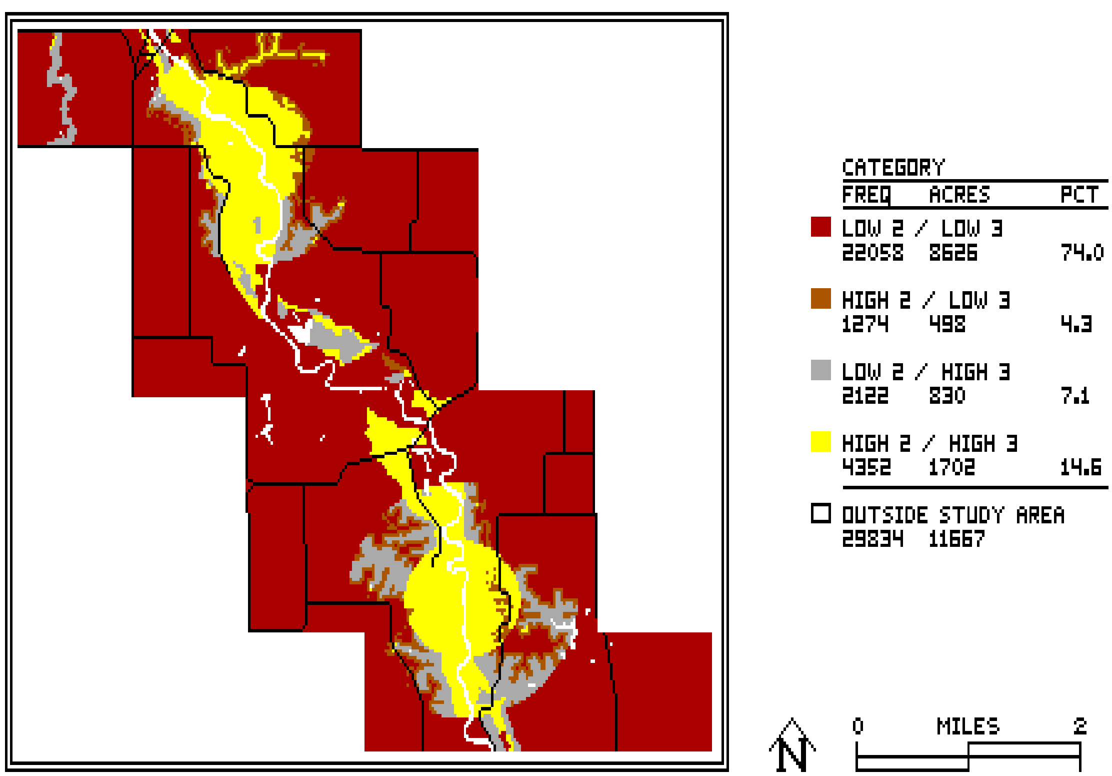


Figure 10. Models 2 and 3 in the Minburn Unit

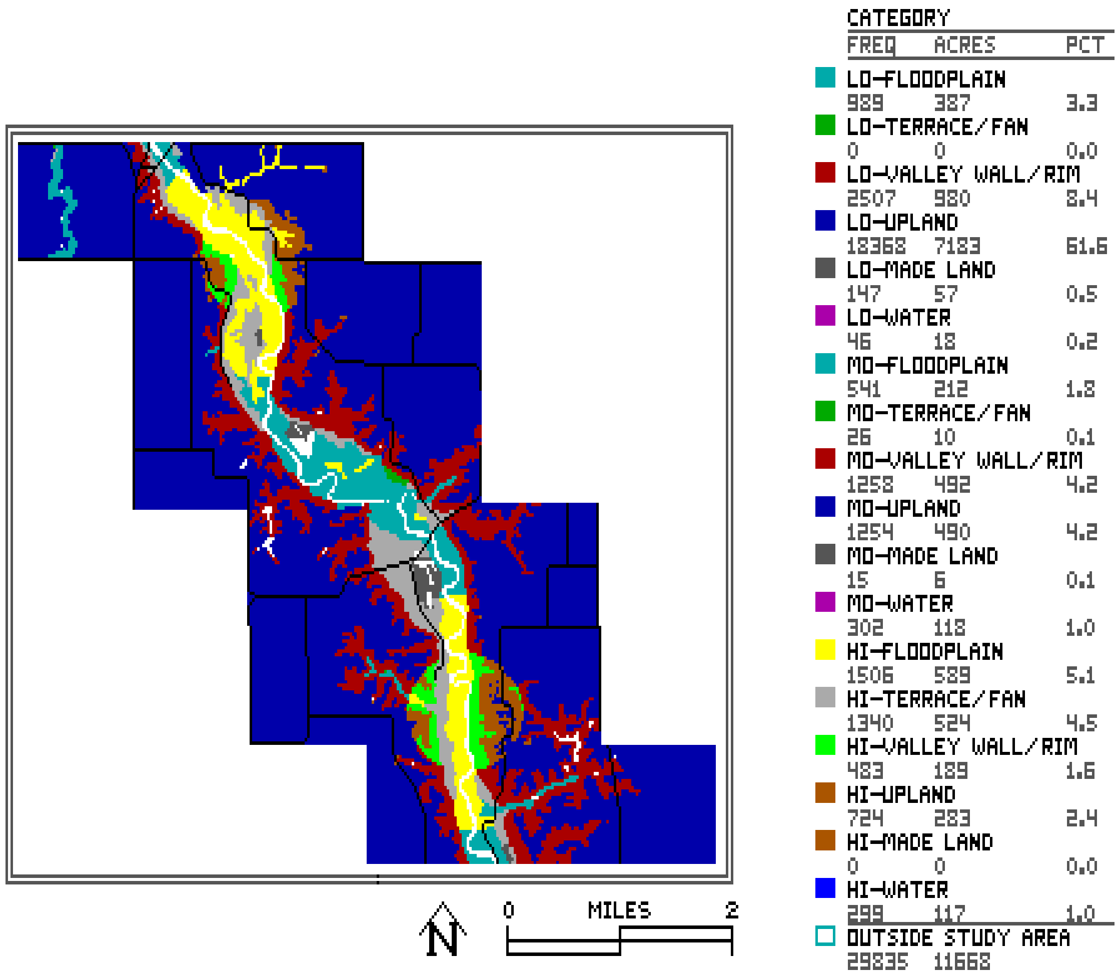


Figure 11. Landscape position for Models 2 and 3 in the Minburn Unit

Peterson outlined and numbered each of the 21 sites in the Minburn Unit on copies of USGS 7.5 minute quadrangle maps. These were digitized using EdCell software written by the author (Figure 12). Also digitized were approximately 37 known non-sites outlined by Peterson. These areas were surveyed by the field team but did not yield any archaeological evidence.

Following descriptive modeling of the 21 additional sites, three additional predictive models were developed. Model 4 used the same variables as Models 1, 2, and 3. However, multipliers were changed based on Chi-square measures produced during descriptive modeling. Chi-square measures also provided the basis for Model 5. GLO historic vegetation was used instead of proximity to confluences.

Model 6 was then developed in an effort to increase the improvement over chance. Descriptive modeling compared the characteristics of the 21 additional known sites with those of *known* non-sites (sites examined by the field team in which no archaeological resources were found). On this basis, native vegetation from soils was substituted for GLO historic vegetation in Model 6.

Use predictive models to guide additional field surveys. Results of Model 2 and Model 3 were used as a guide in conducting the field surveys along portions of the North Raccoon River valley. This was one form of validating the predictive models. The results of Models 2 through 6 will be incorporated into procedures for zoning review and subdivision plat review in Dallas County. In addition, model results are being used in developing resource management plans for individual landowners in the Greenbelt area.

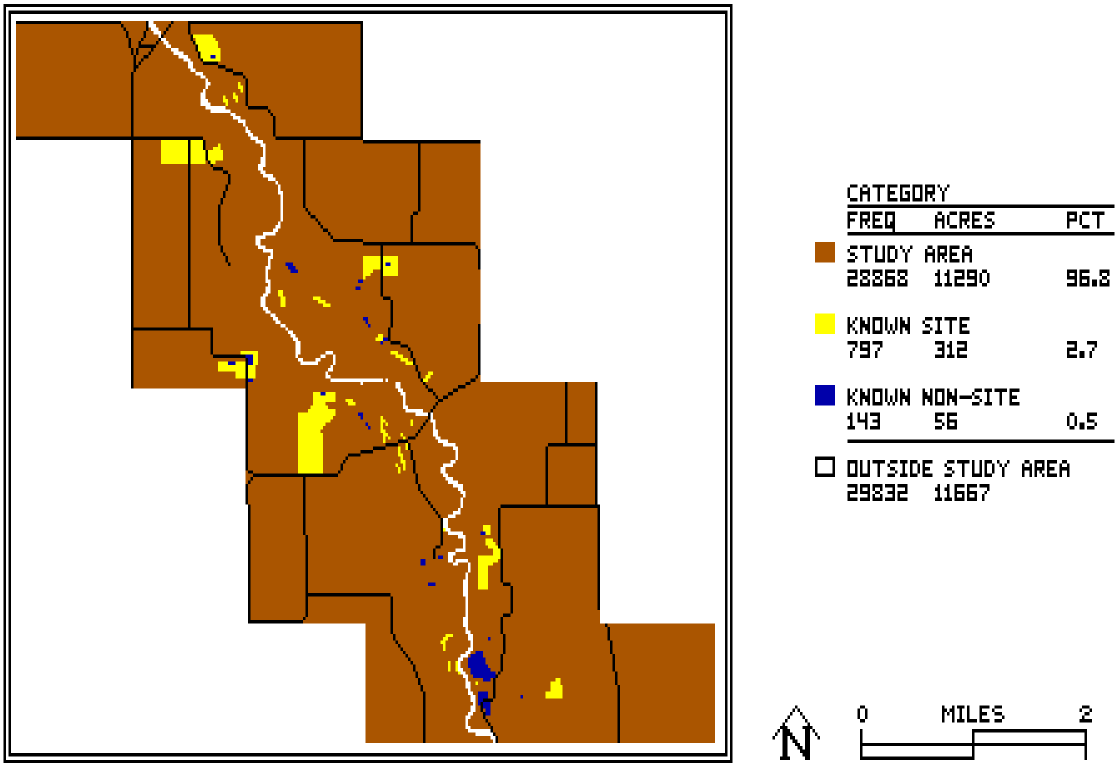


Figure 12. Known sites and known non-sites in the Minburn Unit

C. Assumptions

The first assumption was that the predictive models assist in finding archaeological sites similar to the known sites. This was because the predictive models were based in part on the descriptive models of known sites. This was a built-in bias of the model. In particular, the easiest sites to find were those in areas currently disturbed in the landscape. The most common and widespread landscape disturbances are cultivation and road construction, which remove vegetation or move soil or other earth materials.

Related to the first, the second assumption was that the known sites were representative of all potential archaeological sites in the study area. This was unlikely. However, until more archaeological sites are located in the study area, this will be unknown and will likely remain so for the foreseeable future. If a known bias is identified in the future, the model could be refined to incorporate this bias. Adjustments could be made to the model to target types of sites not represented in the model.

The third assumption was that the quality of data was equal for all known sites. However, there were differences between sites in the level of uncertainty about site location and extent. In the models, each site contributed equally regardless of type, age, size, extent, location, amount of documentation, or other factors.

The fourth assumption was that locations of river confluences (as shown on the soil maps) were similar to those in the past. Though rivers are dynamic systems that move in response to hydrologic processes, there was no indication that the confluences had moved significantly. For example, when comparing the GLO township plat maps and soil survey maps with proportional measurement techniques, the confluence of the Middle Raccoon River and South Raccoon River had moved over 600 feet. There have been some documented cases of major changes in river channel alignment and confluence location elsewhere in Iowa. The Little Maquoketa River and its confluence with the Mississippi River is an example of a change of over five miles (Bettis 1987).

The fifth assumption was that the sample of non-sites included sites that did not have archaeological evidence. However, these sites were not surveyed or confirmed as non-sites. In other words, they were *assumed* non-sites but not *known* non-sites. This assumption does not apply to the second descriptive modeling procedures for Phase 2 research, in which characteristics of known sites were compared to characteristics of approximately 37 known non-sites.

IV. Results

A. Descriptive modeling of known sites

Frequencies were initially computed using two methods: by sub-sites and by cells. Because many of the known sites were composed of more than one cell in the raster map data layer, frequency distribution was different for sub-sites than for cells. (In Phase 1, the mean size of the 85 known sites was 8.7 cells; median 4 cells; mode 2 cells; minimum 1 cell; maximum 165 cells. In Phase 2, the mean size of the 21 known sites was 6.8 cells; median 3 cells; mode 2 cells; minimum 1 cell; maximum 60 cells.) For example, Phase 1 frequencies for proximity to stream confluences were the following:

Proximity to	**Sub-sites**		**Cells**	
confluences	**Frequency**	**Percent**	**Frequency**	**Percent**
0-0.25 mi.	28	29	311	42
0.26-0.50 mi.	46	47	336	45
0.51-1.00 mi.	20	20	85	11
1.01-2.00 mi.	3	3	7	1
>2.00 mi.	1	1	2	<1
Totals	98	100	741	100

Figure 13 illustrates the difference between sub-sites and cells. Finney and Peterson suggested that frequencies by sub-sites were more appropriate than by cells. The reason was that the areal extent of archaeological sites was often uncertain. Frequency by sub-sites indicated composition only; frequency by cells indicated both composition and area. Therefore, frequency by sub-sites was a less area-sensitive measure and more reflective of what was known about the archaeological sites. Therefore, the following descriptive modeling results for Phase 2 are presented using frequencies by sub-sites.

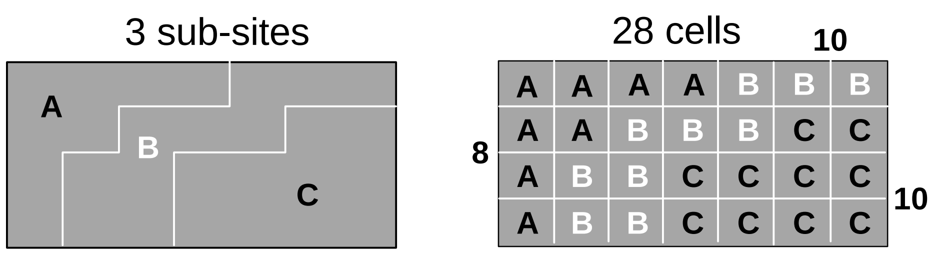


Figure 13. Comparison of frequency by sub-sites (left) and cells (right)

The following Phase 2 descriptive modeling results compare frequency distribution of landscape characteristics by sub-sites at the 21 known sites with the random sample of non-sites.

Proximity to	**Known sites**		**Sample of non-sites**	
confluences	**Frequency**	**Percent**	**Frequency**	**Percent**
0-0.25 mi.	0	0	3	2
0.26-0.50 mi.	4	18	12	9
0.51-1.00 mi.	4	18	33	24
1.01-2.00 mi.	11	50	74	54
>2.00 mi.	3	14	14	10
Totals	22	100	136	100

Proximity	**Known sites**		**Sample of non-sites**	
to valley	**Frequency**	**Percent**	**Frequency**	**Percent**
In valley	7	28	40	29
0-200 ft.	11	44	16	12
201-500 ft.	4	16	15	11
501-1000 ft.	3	12	19	14
>1000 ft.	0	0	46	34
Totals	25	100	136	100

Landscape position	Known sites Frequency	Known sites Percent	Sample of non-sites Frequency	Sample of non-sites Percent
Water	0	0	1	1
Poorly dr. floodplain	2	6	7	5
Mod-poor dr. floodplain	1	3	8	6
Mod well dr. floodplain	1	3	1	1
Made land	0	0	0	0
Terrace	4	13	8	6
Alluvial fan	0	0	0	0
Valley wall	1	3	16	12
Valley rim	3	10	0	0
Poorly drained upland	1	3	26	19
Upland sideslope	3	10	17	13
Upland ridge	15	48	52	38
Totals	31	100	136	100

In addition to the three variables above, several other variables were selected for descriptive modeling: 1990 land cover, 1847-1851 GLO vegetation, and native vegetation (based on soils).

Land cover	Known sites Frequency	Known sites Percent	Sample of non-sites Frequency	Sample of non-sites Percent
Cropland	11	41	78	57
Woodland	6	22	34	25
Pasture	4	15	8	6
Scattered trees	4	15	7	5
Road	1	4	2	1
Old field	1	4	1	1
Farmstead	0	0	2	1
Stream	0	0	1	1
Residential	0	0	1	1
Open lawn	0	0	0	0
Pond/lake	0	0	0	0
Reservoir	0	0	0	0
Commercial/industrial	0	0	0	0
Mining	0	0	0	0
Cemetery/institutional	0	0	0	0
Recreation	0	0	0	0
Prairie	0	0	0	0
Totals	27	100	136	100

GLO vegetation	Known sites Frequency	Known sites Percent	Sample of non-sites Frequency	Sample of non-sites Percent
Timber	17	81	50	37
Rough	0	0	0	0
Field	0	0	0	0
Prairie	4	19	86	63
Totals	21	100	136	100

Native vegetation	Known sites Frequency	Known sites Percent	Sample of non-sites Frequency	Sample of non-sites Percent
Wet prairie	1	3	29	21
Mesic prairie	10	32	67	49
Dry prairie	1	3	2	1
Grass and trees	12	39	16	12
Trees	7	23	21	15
Not listed	0	0	1	1
Totals	31	100	136	100

Graphs of cumulative frequencies were made to help summarize and compare frequency data for known sites and non-sites (Figures 14 (proximity to confluences), 15 (landscape position), 16 (proximity to valley), 17 (land cover), 18 (GLO vegetation), and 19 (native vegetation)).

Chi-square was used to measure differences in frequency distribution between the 21 known sites and sample of non-sites in the Minburn Unit. Tests of significance yielded *p* values of 0.001 or less for all Chi-square measures except (by sub sites) landscape position (0.367), proximity to confluences (0.713), and land cover (0.114).

Variable (Phase 2)	**Chi-square**		**Chi-square**	
Data layer	**Cells**	**Normalized**	**Sub-sites**	**Normalized**
Landscape position	97.340	100	6.534	30
GLO vegetation	81.412	84	14.519	66
Proximity to valley	73.654	76	22.085	100
Native vegetation	48.986	50	17.357	79
Proximity to confluences	35.127	36	1.370	6
Land cover	21.217	22	7.461	34

Normalized values helped study the relationships (relative values) between Chi-square measures. This was useful for refining the selection of multipliers for the predictive models.

Chi-square measures were different for sub-sites than for cells because frequencies were quite different. Also, in contrast to Phase 1, the ordered sequence of the variables was different. This may have been due to the fact that there were only 21 known sites in the Minburn Unit, compared to a total of 85 known sites in the Phase 1 study area. When frequencies are low for categories within a variable, the Chi-square measure is more likely invalid and *p* values increase, which increases the probability that repeated sample will likely yield much different results. Stated another way, when more than a few frequencies are low, the Minitab statistical software package displays warning messages about “expected counts less than 5” and “cells with expected counts less than 1. Chi-square approximation is probably invalid.”

The Chi-square measures (by cells) indicated that the greatest differences between known sites and non-sites in the Minburn Unit were in *landscape position, GLO vegetation*, and *proximity to valley*. In contrast, Chi-square values from Phase 1 were highest for *proximity to confluences, landscape position, GLO vegetation,* and *proximity to valley*:

Variable (Phase 1)	**Chi-square**		**Chi-square**	
Data layer	**Cells**	**Normalized**	**Sub-sites**	**Normalized**
Proximity to confluences	716.871	100	131.782	100
Landscape position	528.420	74	59.251	45
GLO vegetation	452.470	63	55.414	42
Proximity to valley	350.552	49	54.056	41
Land cover	209.852	29	49.900	38
Native vegetation	120.580	17	15.017	11

It was not surprising that *landscape position* had the highest Chi-square value in the Minburn Unit because Peterson reported that it was a major factor in the field survey work. It was also not surprising that *proximity to confluences* had a relatively low Chi-square value in the Minburn Unit because the North Raccoon River valley is in a relatively young landscape (geologically), has a less well-developed drainage system, and therefore fewer stream confluences.

In rank, *GLO vegetation* and *proximity to valley* both moved higher on the Chi-square list for the Minburn Unit (from third and fourth, respectively, to second and third). This was logical, based on the fact that *proximity to confluences* moved from first to fifth on the list. The spatial distributions of GLO vegetation and proximity to valley were somewhat similar and (to minimize the effects of autocorrelation) should not be used together in the same model. Therefore, CAC was used to measure the overlap in spatial distributions. The CAC measure for the Minburn Unit was 49 percent, which indicated a low to moderate overlap. For similar reasons, the spatial distribution of GLO vegetation was also compared to that of native vegetation. The CAC was 46 percent when GLO timber was compared to the combination of native trees and native grass and trees. The CAC was even lower, 31 percent, when GLO timber was compared to native trees.

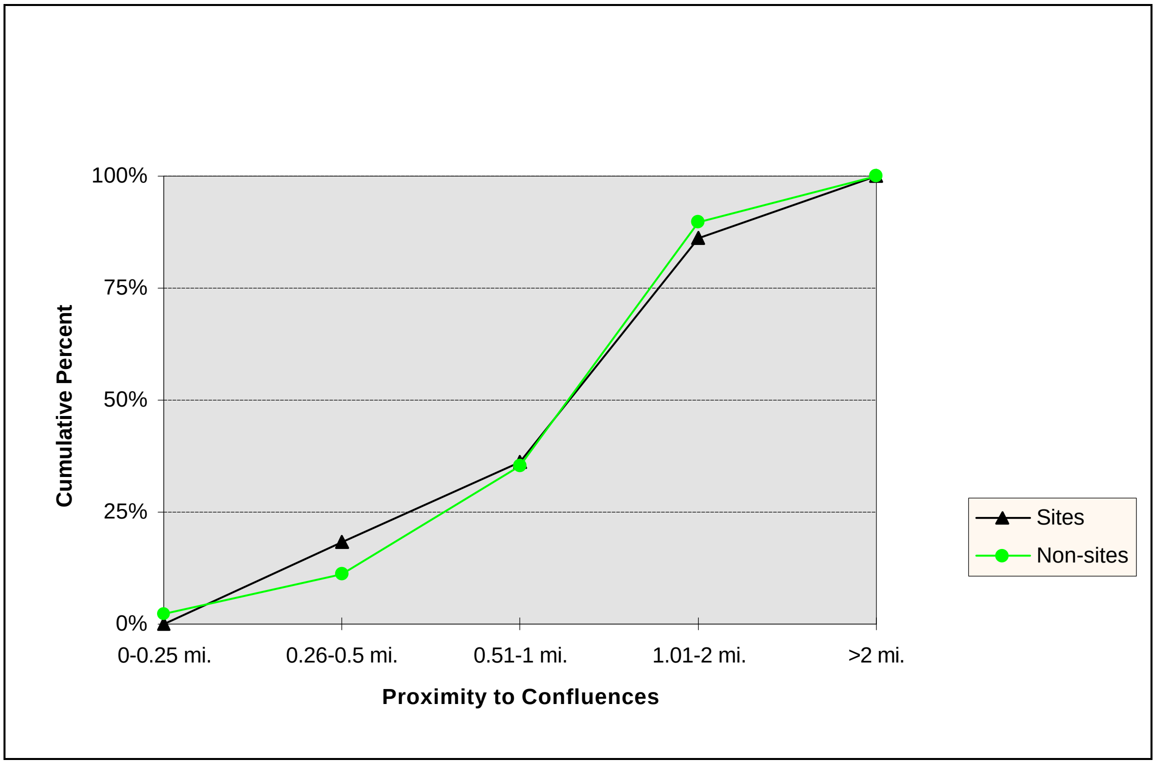


Figure 14. Cumulative percent distribution for proximity to confluences

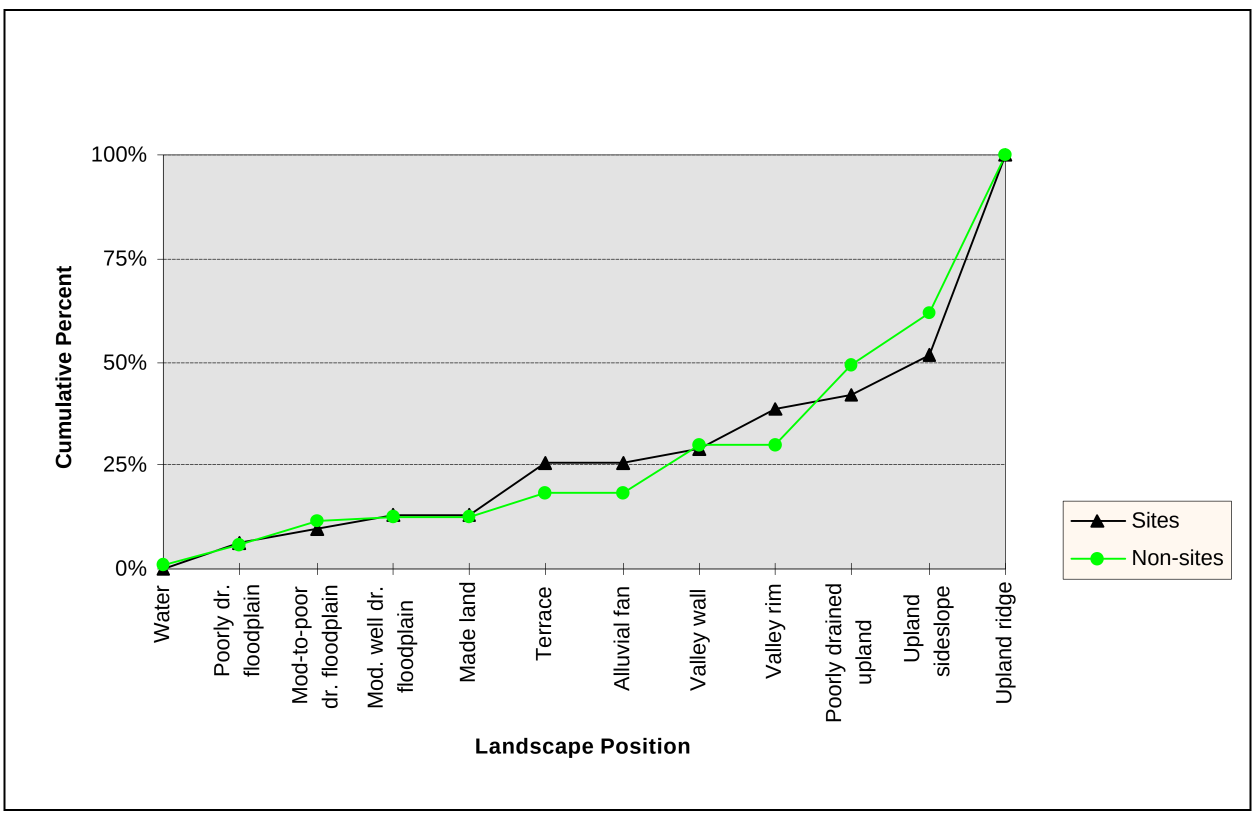


Figure 15. Cumulative percent distribution for landscape position

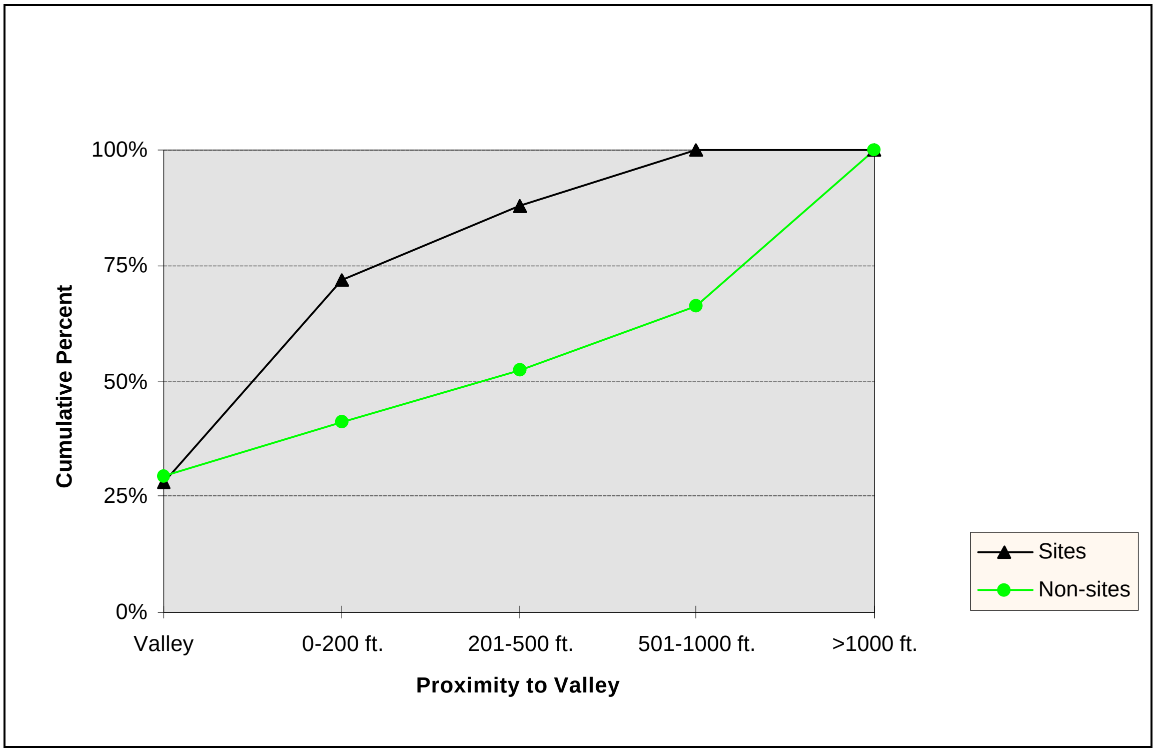


Figure 16. Cumulative percent distribution for proximity to valley

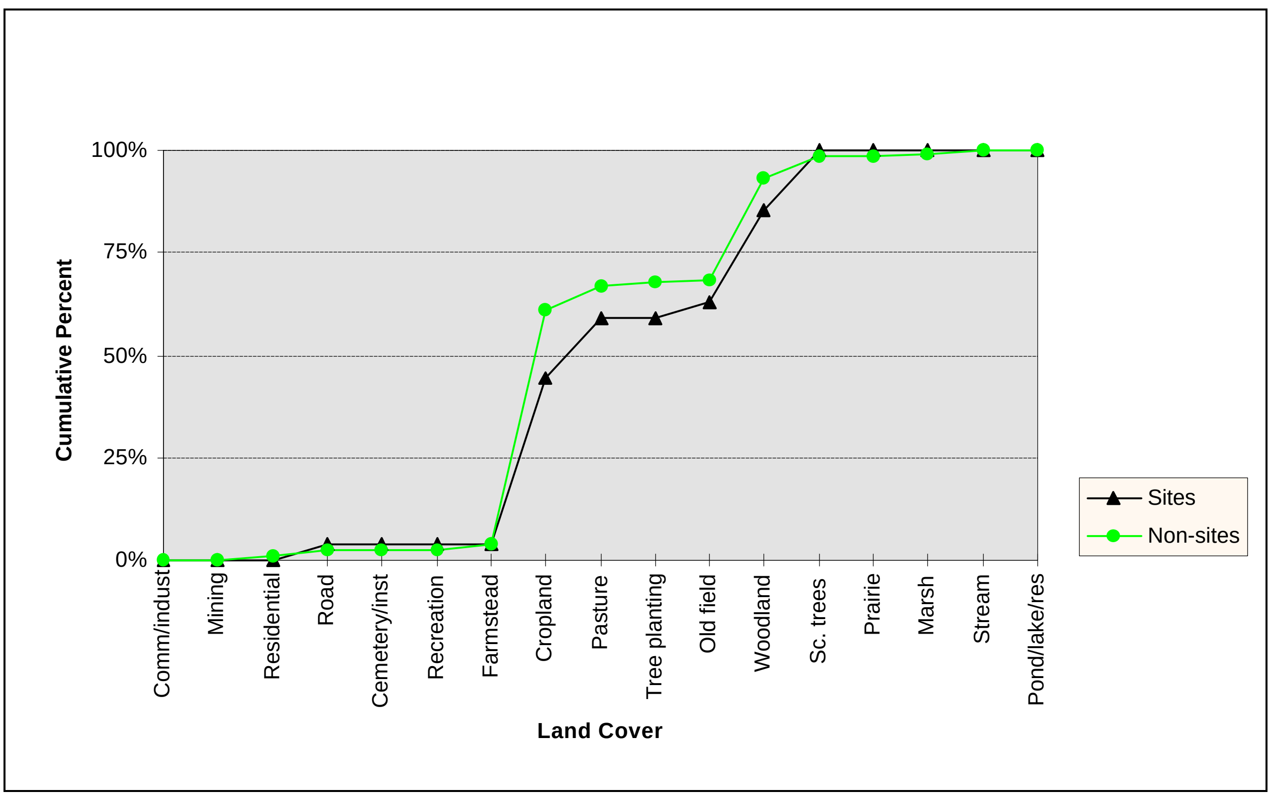


Figure 17. Cumulative percent distribution for land cover

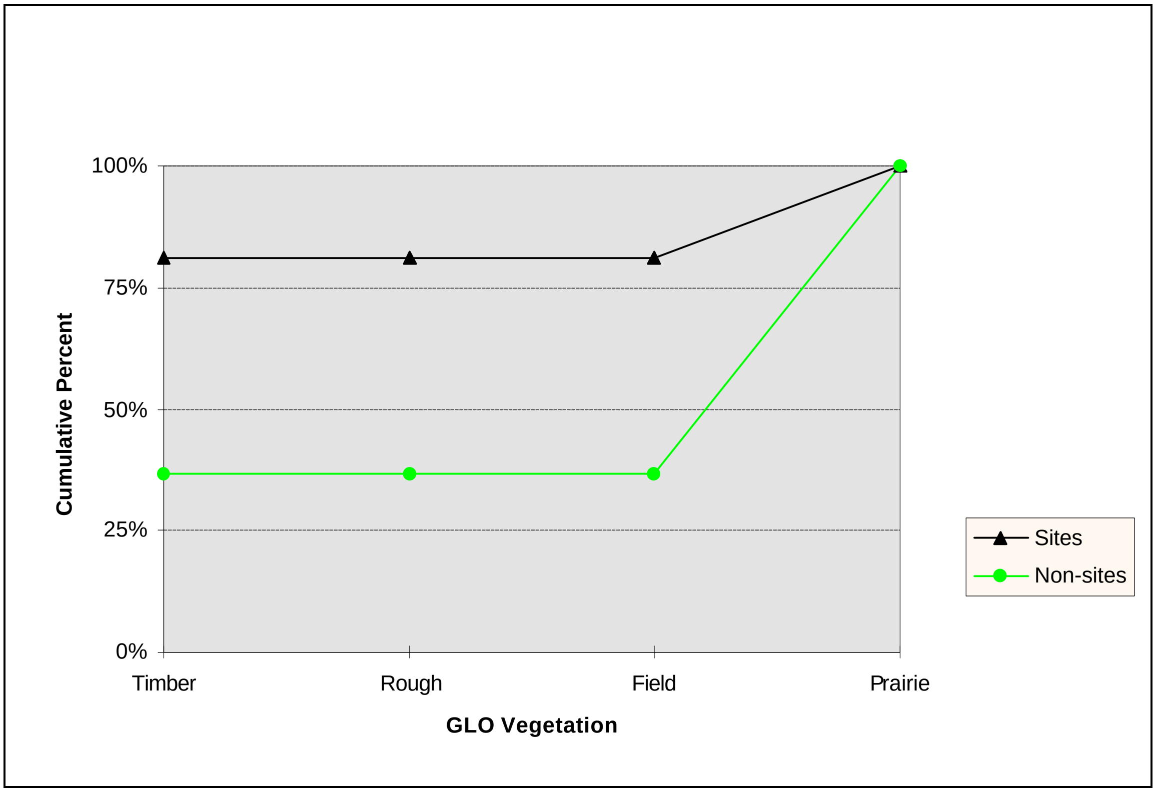


Figure 18. Cumulative percent distribution for GLO vegetation

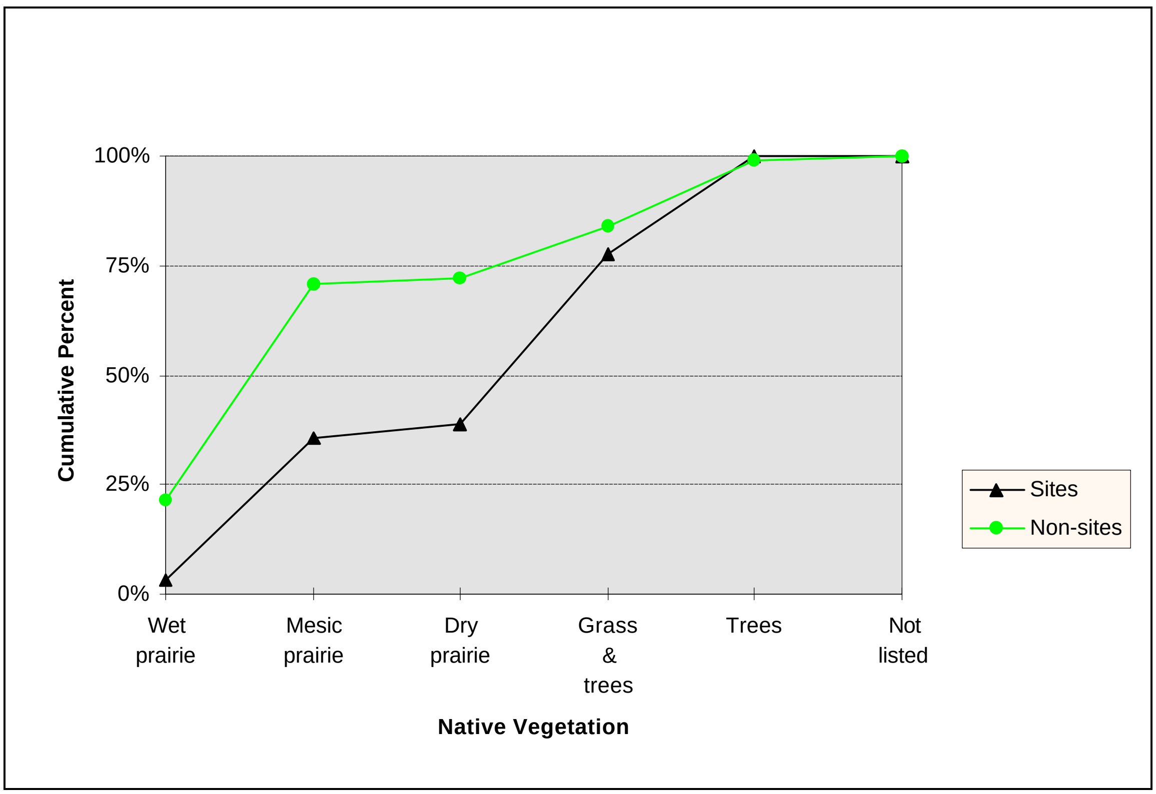


Figure 19. Cumulative percent distribution for native vegetation

Chi-square values for *native vegetation* increased in the Minburn Unit and moved from sixth to fourth on the list. Native vegetation is an interpretation of soils which indicates the predominant vegetation over the past 3,000 years (since the last major erosion cycle). In Phase 1, it was surprising that this indicator of historic vegetation had such little difference between known sites and non-sites. However, this was not the case in Phase 2. Its Chi-square measure by cells was still not as high as that for GLO vegetation, but was higher by sub-sites.

As in Phase 1, *land cover* showed relatively little difference between known sites and non-sites. This was expected because of its uncertain relationship with land cover in the past 1,000 to 12,000 years. However, existing land cover was a factor in accessibility during field surveys. In the Minburn Unit, 41 percent of the known sites (and 74 percent of the cells) occurred in existing cropland, where seasonal cover and tillage practices made archaeological evidence quite visible at certain times of the year (particularly spring). A similar value (approximately 57 percent) of the sample of non-sites occurred in existing cropland (also 57 percent of the cells). Because the statistical distributions were similar, the Chi-square measure was relatively low.

B. Predictive modeling of potential sites

Models 1, 2, and 3 were developed during Phase 1 of the research. They were applied to five units of the Raccoon River Greenbelt along the South Raccoon River corridor. In Phase 2 of the research, Models 2 and 3 were applied to the Minburn Unit. After field surveys and descriptive modeling were completed for Phase 2, Models 4, 5, and 6 were developed in a effort to customize the models to the landscape of the Minburn Unit, increasing their predictive power.

Model 1. Model 1 was applied to only the Redfield Unit of the Phase 1 study area as a test of the initial criteria supplied by Finney and Peterson. High potential areas were concentrated only around major stream confluences. In the Redfield area, this included the confluence of the Middle Raccoon River and Mosquito Creek and the confluence of the Middle Raccoon River and the South Raccoon River.

As described previously in Section III.B, several refinements were made based as a result of the test: redefinition of stream confluences and increasing points for alluvial fan landscape positions. These were incorporated into Model 2.

Model 2. Model 2 was applied to the entire study area. Model 2 results for the Minburn Unit are shown in Figure 20. High potential areas were located at two confluences, one near the north end of the unit and one near the south end.

The statistical distribution of Model 2 scores (without aggregation) is shown in Figure 21. The graph shows the trend that non-sites scored predominantly low scores and known sites scored predominantly moderate and high scores. Another way to display the statistical distribution of Model 2 scores was by cumulative percentage (Figure 22). Again the trend was visible on the graph: non-sites scored predominantly low scores and known sites scored predominantly moderate and high scores.

This graph was modified to create another graph showing percent correctly classified at each potential cutpoint score from 6 to 30 (Figure 23). To make this graph, percentage values for known sites were subtracted from the value 100 at each score from 6 to 30. This had the graphic effect of mirroring the curve on the previous graph for known sites using a horizontal axis at the 50% line. At each potential cutpoint score from 6 to 30, the percent of non-sites and percent of known sites correctly classified were displayed on the graph. For example, at cutpoint score 16, 84.8% of known sites were correctly classified as high potential; 70.9% of non-sites were correctly classified as low potential.

This was the characteristic feature of logit models: classifying each part of the study area into either high potential or low potential. This Boolean (binary) approach resulted in only two categories of potential: high and low. The optimum cutpoint score was the one which maximized both percentages:

- percent of known sites correctly classified as high potential
- percent of non-sites correctly classified as low potential

To find the optimum, both percentages were combined at each potential cutpoint score from 6 to 30 (Figure 24).

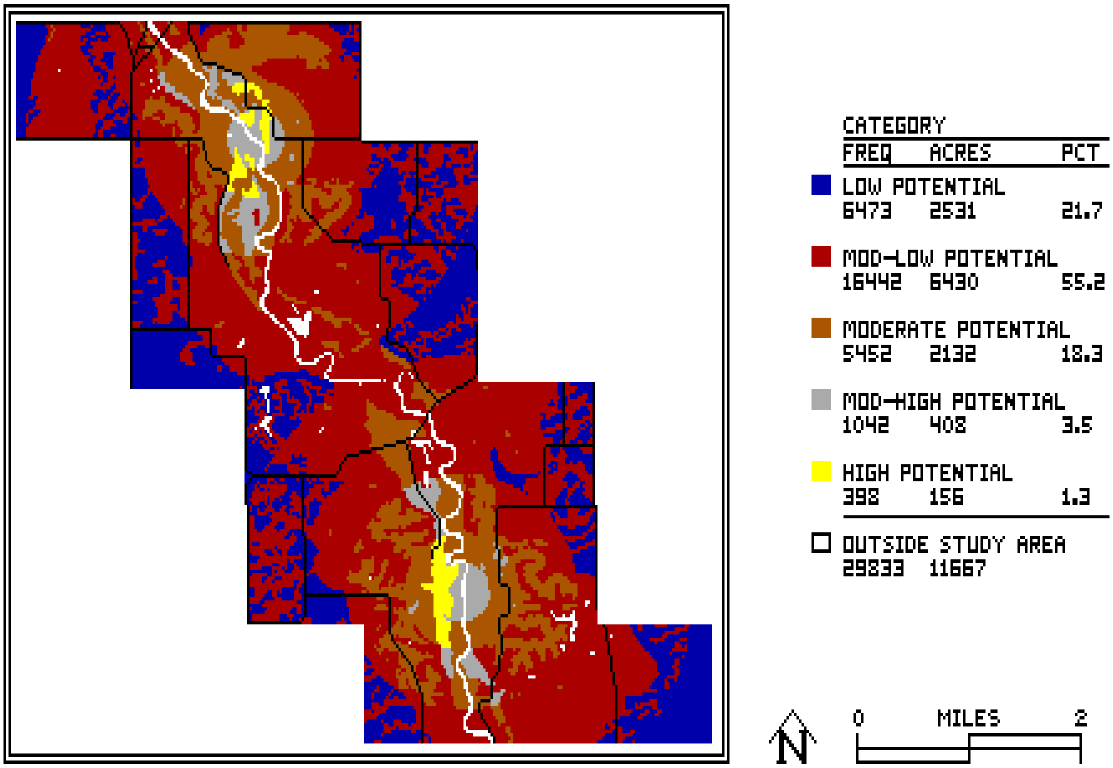


Figure 20. Model 2 for the Minburn Unit

For Model 2 in the Minburn Unit, the optimum cutpoint score was 13 (0.29 on a scale of 0 to 1), where the curve peaked on the graph (Figure 24). At a cutpoint score of 13, Model 2 provided an improvement of 36.1% over chance. This means that at a cutpoint score of 13, Model 2 correctly classified 66.7% of the known sites as high potential. At that cutpoint score, 69.4% of the non-sites were correctly classified as low potential and, therefore, only 30.6% were (incorrectly) classified as high potential. At a cutpoint score of 13, this would have produced a model of high potential which covered approximately 30.6% of the study area when mapped cell by cell (Figure 25). This represented a 36.1% improvement in predictive power over chance (66.7 minus 30.6) because a model with no predictive power that covered 30.6% of the study area should, by chance, have classified only 30.6% of the known sites correctly. The fact that 66.7% of known sites were correctly classified yielded the 36.1% improvement over chance (improvement from 30.6% to 66.7%).

In other words, though only approximately 31% of the study area scored 13 or higher, approximately 67% of the known sites scored 13 or higher. If known sites were distributed randomly throughout the study area, only approximately 31% of the known sites would have scored 13 or higher. Because known sites were not distributed randomly, Model 2 improves predictive power because it correctly classified approximately 67% of the known sites correctly (which scored 13 or higher) in the high potential area.

As described previously in Section III.B, three major refinements were made by Finney and Peterson based on the results of Model 2. These included adjustments in points for proximity to valley categories, adjustment in points to landscape position categories, and change in multipliers for proximity to valley and landscape position. These changes were incorporated into Model 3.

Model 3. Model 3 was then applied to the entire study area. Results for the Minburn Unit are shown in Figure 26. High potential areas were located in a slightly more continuous linear pattern along the North Raccoon River. However, high potential areas were still quite concentrated around two confluences.

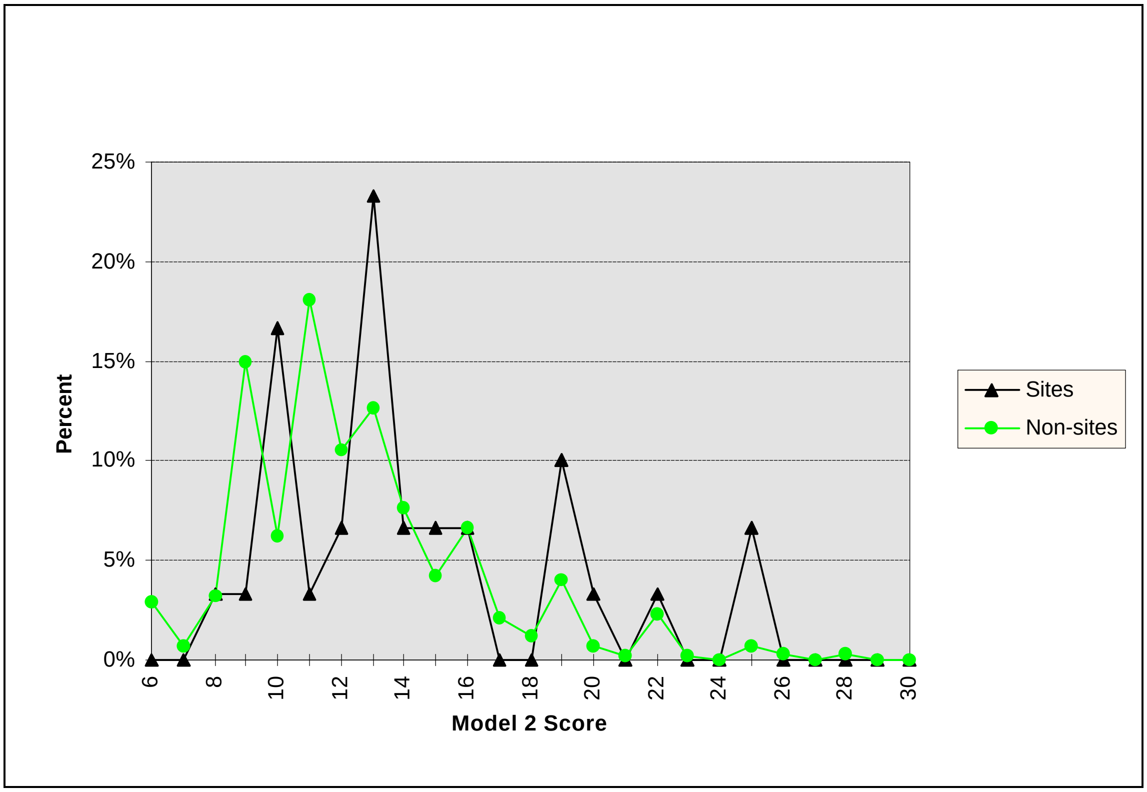


Figure 21. Percent distribution of Model 2 scores

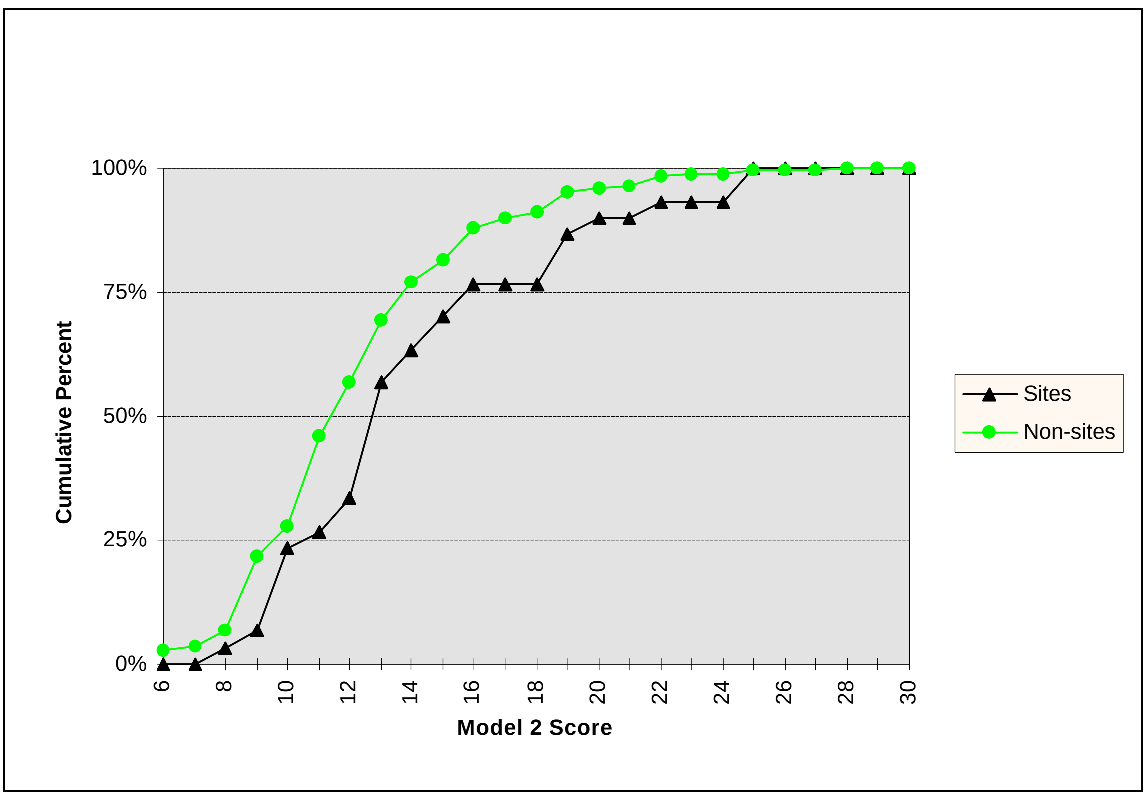


Figure 22. Cumulative percent distribution of Model 2 scores

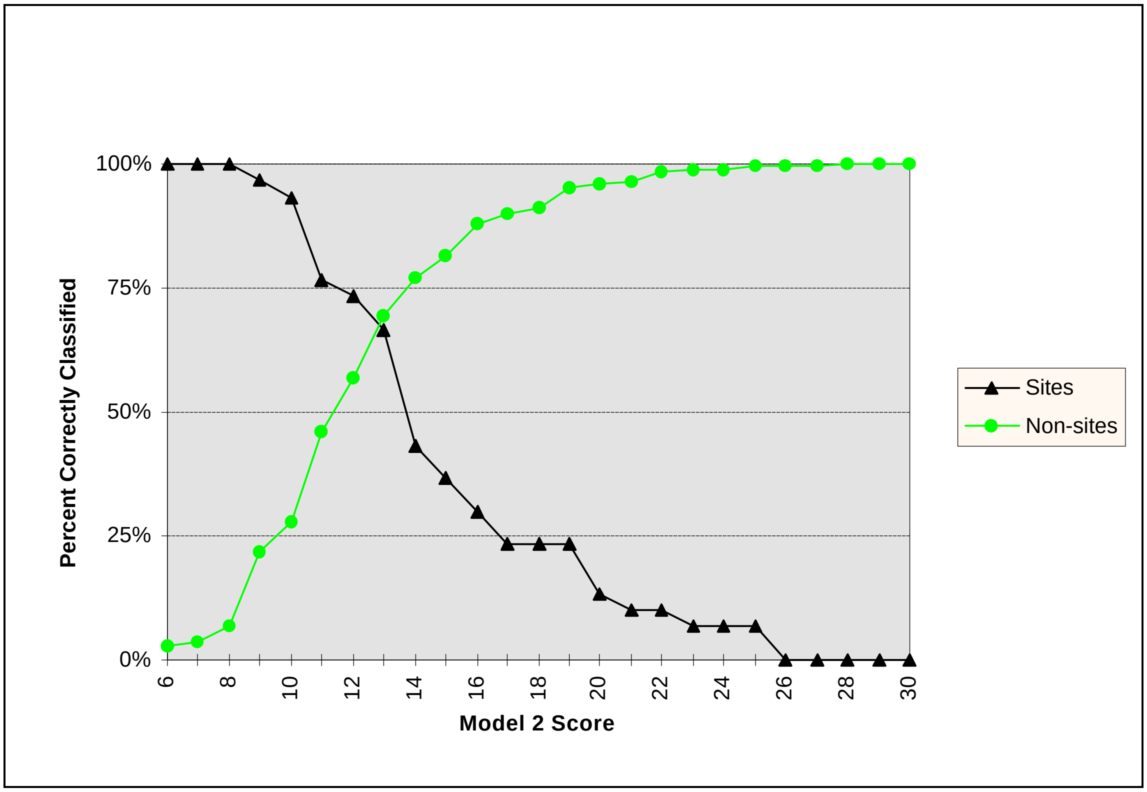


Figure 23. Model 2 areas correctly classified at each potential cutpoint score

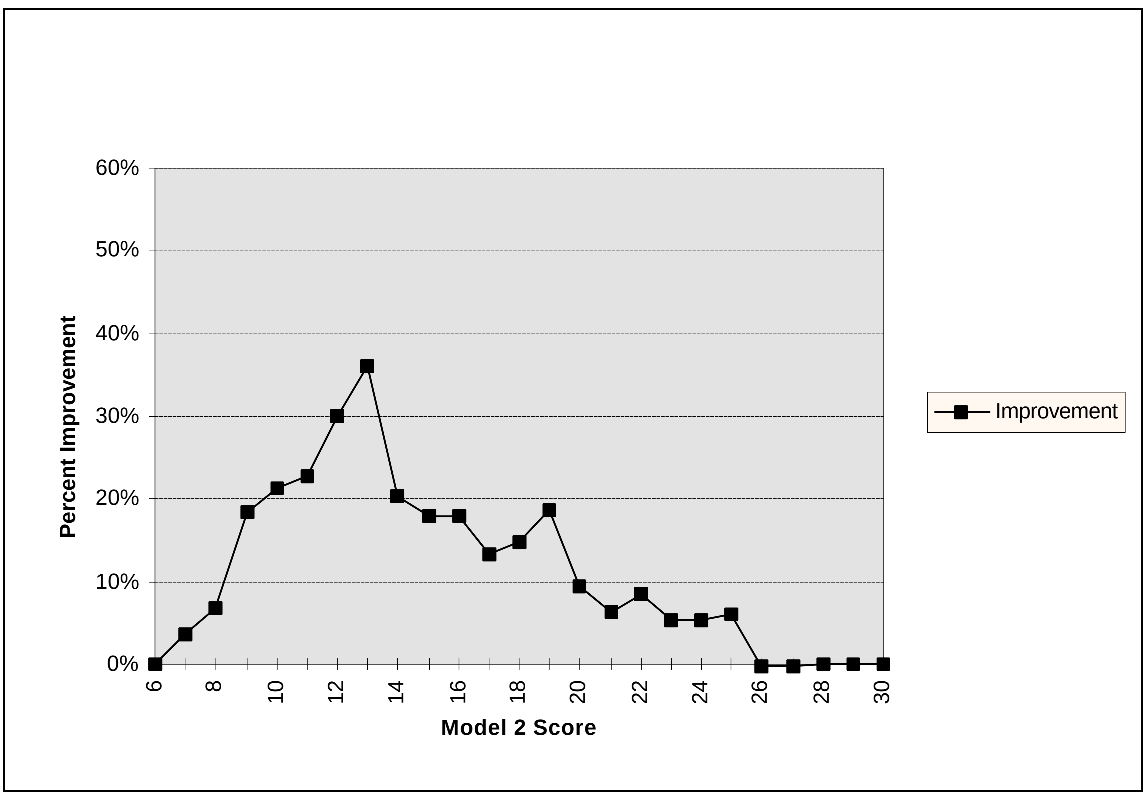


Figure 24. Model 2 improvement over chance

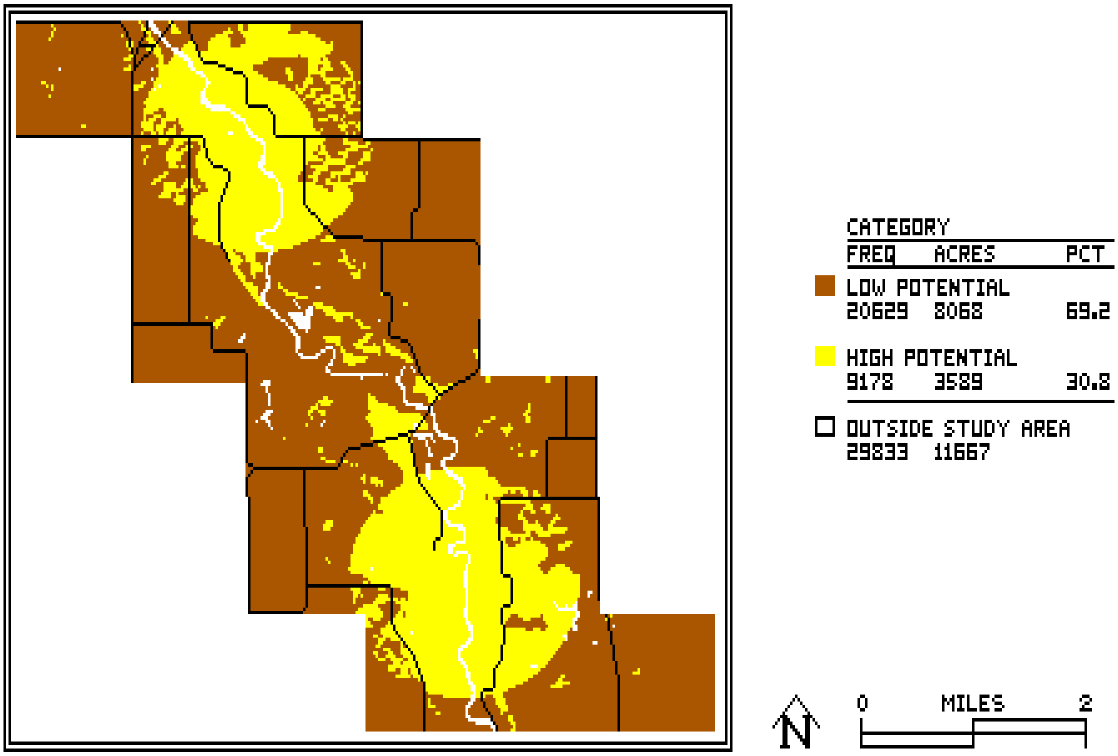


Figure 25. Model 2 cutpoint categories for the Minburn Unit

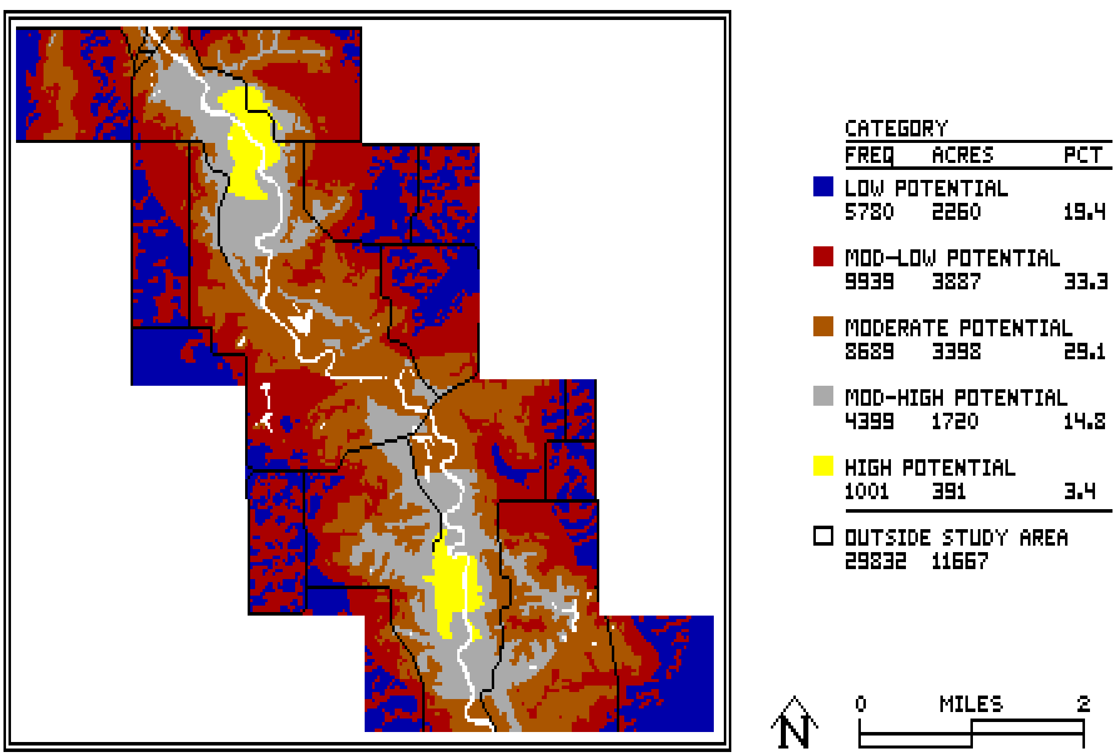


Figure 26. Model 3 for the Minburn Unit

Statistical distribution of Model 3 scores is shown in Figure 27. Again the trend was visible on the graph: non-sites scored predominantly low scores and known sites scored predominantly moderate to high scores. Another way to display the statistical distribution of Model 3 scores was by cumulative percentage (Figure 28). On this graph, the trend was also visible: non-sites scored predominantly low scores and known sites scored predominantly moderate to high scores. This graph was modified to create another graph showing percent correctly classified at each potential cutpoint score from 6 to 30 (Figure 29). To find the optimum, both percentages were combined at each potential cutpoint score from 6 to 30 (Figure 30).

For Model 3, the optimum cutpoint score was 11 (0.21 on a scale of 0 to 1), where the curve peaked on the graph (Figure 30). At a cutpoint score of 11, Model 3 provided an improvement of 31.8% over chance. This means that at a cutpoint score of 11, Model 3 correctly classified 96.7% of the known sites as high potential. At that cutpoint score, only 35.1% of the non-sites were correctly classified as low potential and, therefore, 64.9% were (incorrectly) classified as high potential. At a cutpoint score of 11, this would have produced a model of high potential which covered approximately 64.9% of the study area when mapped cell by cell.

Improvements over chance were 36.1% (Model 2) and 31.8% (Model 3). This resulted in a net loss in predictive power from Model 2 to Model 3. In addition, the improvement over chance was much lower for both models in the Phase 2 study area than in the Phase 1 study area.

Phase 1		**Phase 2**		
Model 2	**Model 3**	**Model 2**	**Model 3**	
16	19	13	11	Cutpoint score (on a scale of 6 to 30)
0.44	0.56	0.29	0.21	Cutpoint score (on a scale of 0 to 1)
84.8%	89.7%	66.7%	96.7%	Known sites correctly classified as high potential
70.9%	61.7%	69.4%	35.1%	Non-sites correctly classified as low potential
29.1%	38.3%	30.6%	64.9%	Study area classified as high potential
55.7%	**51.4%**	**36.1%**	**31.8%**	**Improvement over chance**

Why did Model 2 and Model 3 in the Phase 2 study area have lower improvement over chance than in the Phase 1 study area? One potential reason was the difference in landscape characteristics between the South Raccoon River corridor studied in Phase 1 and the North Raccoon River corridor studied in Phase 2. As described in Section IV.A on descriptive modeling, the North Raccoon River valley is in a relatively young landscape (geologically), has a less well-developed drainage system, and therefore fewer confluences (Ruhe 1969; Prior 1991). Therefore, *proximity to confluences* was less useful as a variable in predictive modeling along the North Raccoon River corridor than along the South Raccoon River corridor.

Coefficient of Areal Correspondence (CAC) was used to measure the similarity spatial distribution of Model 2 and Model 3. In Phase 1 (along the South Raccoon River corridor) the CAC was 75 percent, indicating a relatively high agreement (overlap) between the results of Models 2 and 3. However, in Phase 2 along the North Raccoon River corridor (Minburn Unit), the CAC was 56 percent, indicating a moderate agreement (overlap) between the results of Models 2 and 3.

Model 4. Because of less agreement and lower predictive power of Models 2 and 3 in the Minburn Unit than in the Phase 1 study area, Model 4 was developed. Chi-square values resulting from descriptive modeling (see Section IV.A) showed that the three variables used in Model 3 could be used in Model 4, but should have different multipliers (weights):

Multiplier 3	Landscape position
Multiplier 2	Proximity to valley
Multiplier 1	Proximity to confluences

Points assigned to categories of each variable remained the same as in Model 3 (see Appendix C). Model 4 results (Figure 31) showed a 26.0 percent improvement over chance (Figure 32). This was a lower level of improvement than either Model 2 or Model 3. Clearly, *proximity to confluences* was not a useful variable in predictive models for the North Raccoon River corridor.

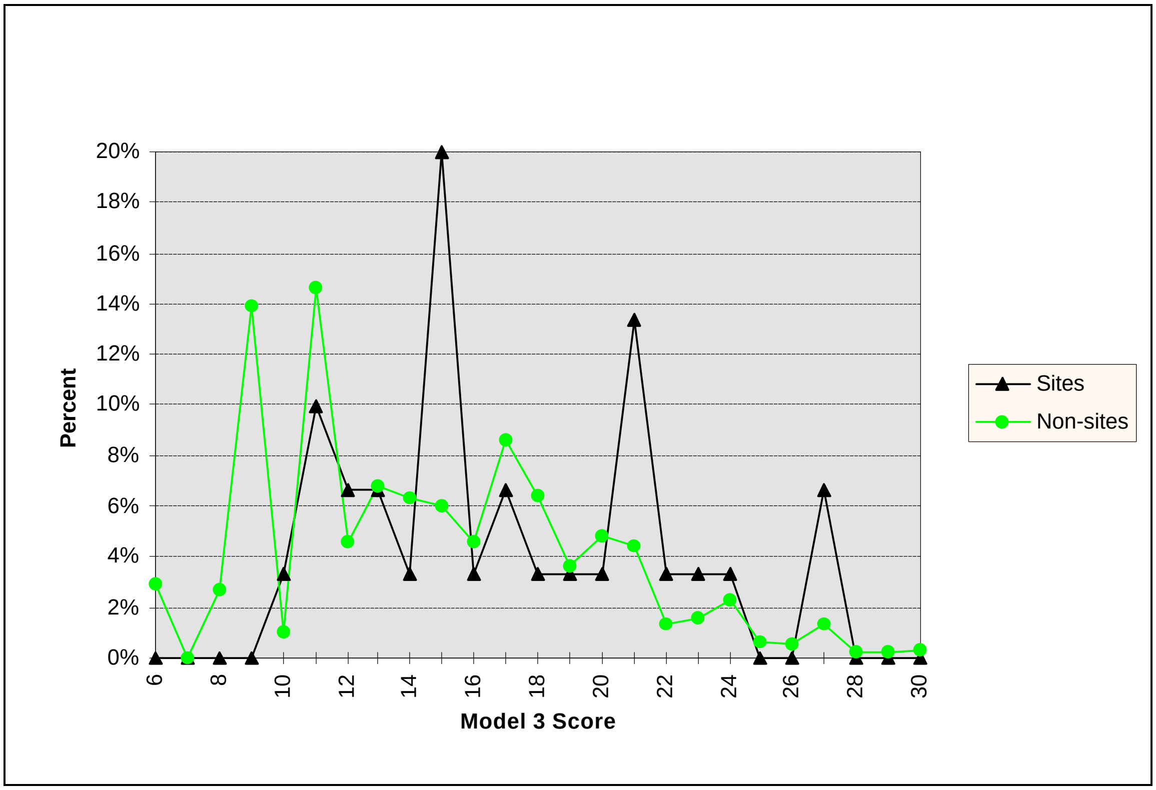


Figure 27. Percent distribution of Model 3 scores

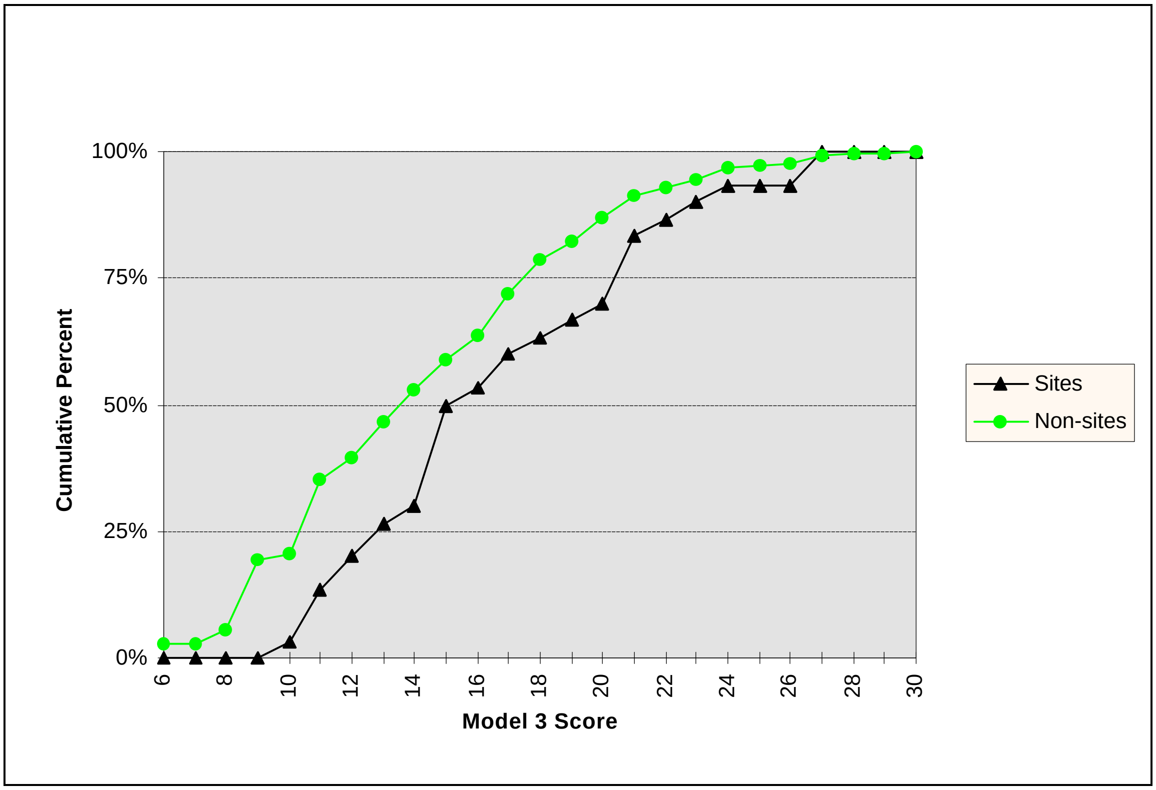


Figure 28. Cumulative percent distribution of Model 3 scores

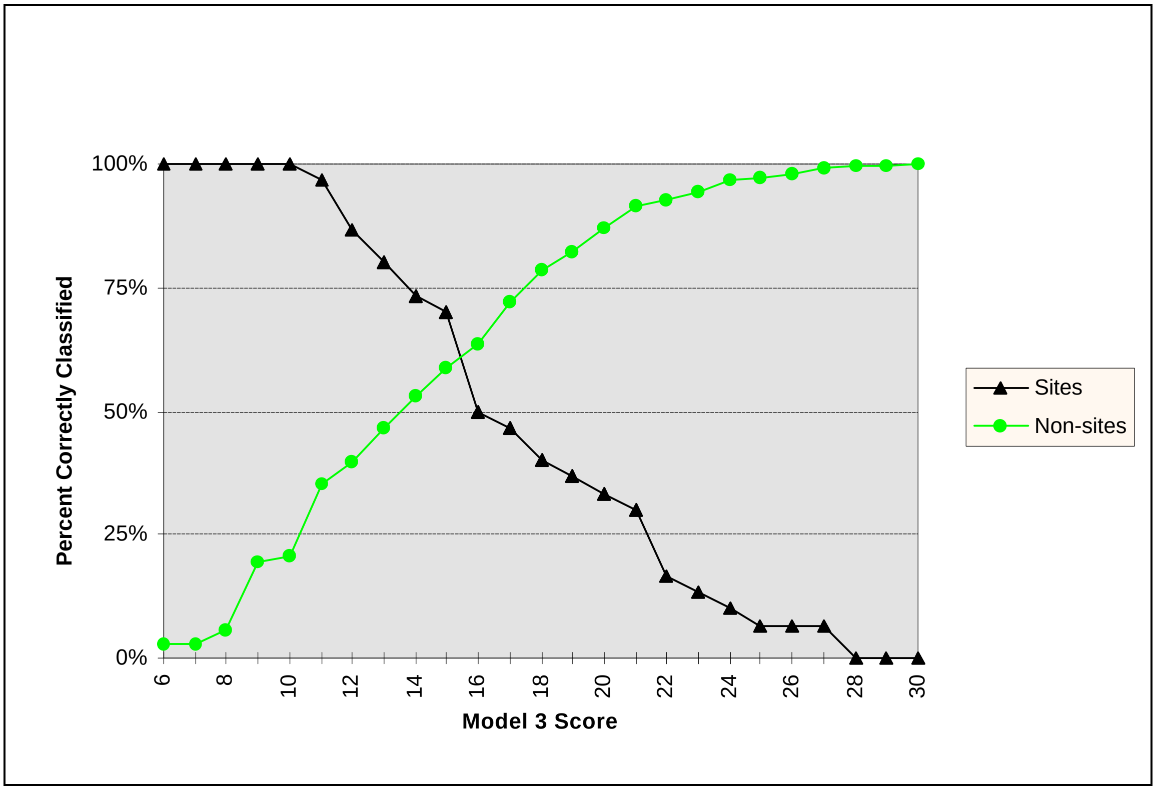


Figure 29. Model 3 areas correctly classified at each potential cutpoint score

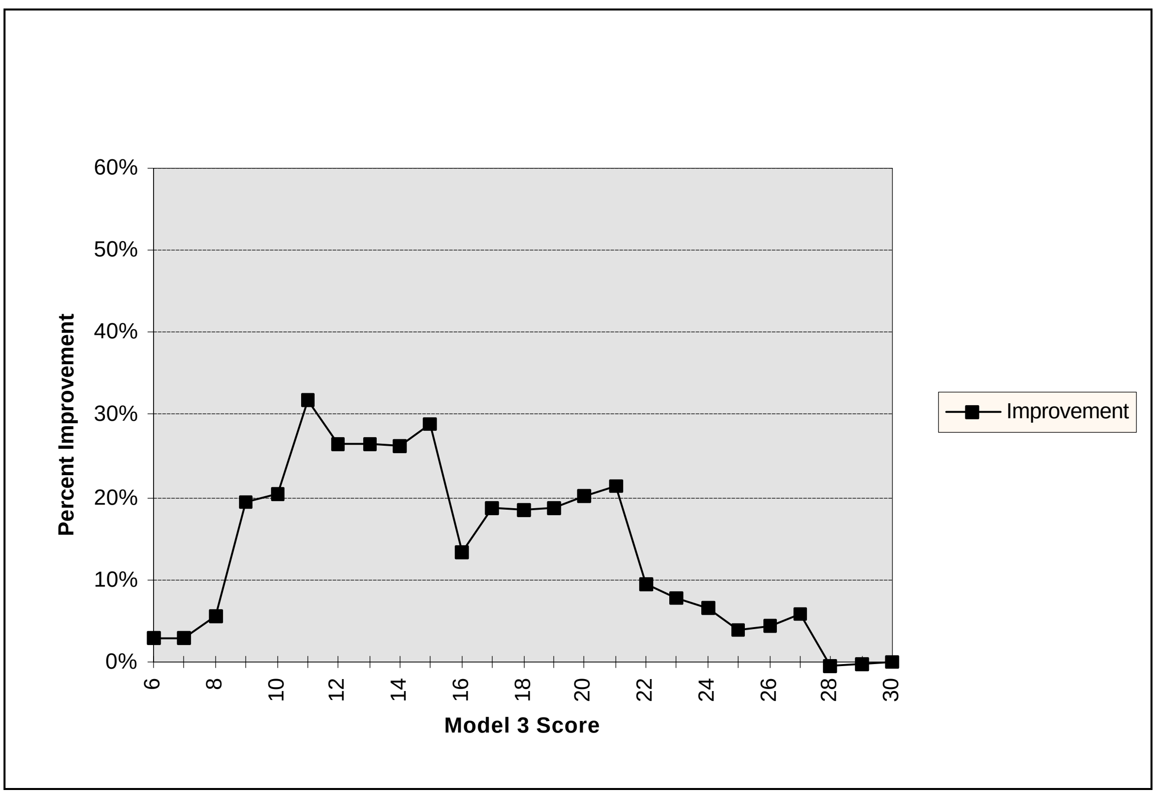


Figure 30. Model 3 improvement over chance

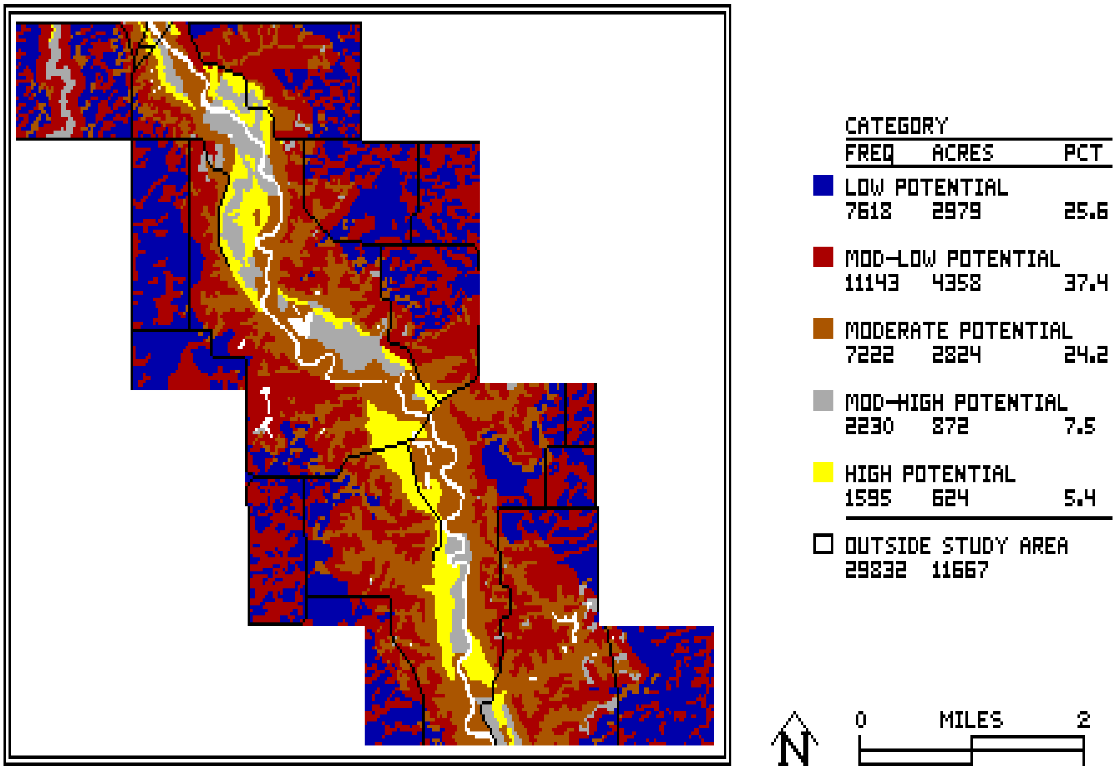


Figure 31. Model 4 for the Minburn Unit

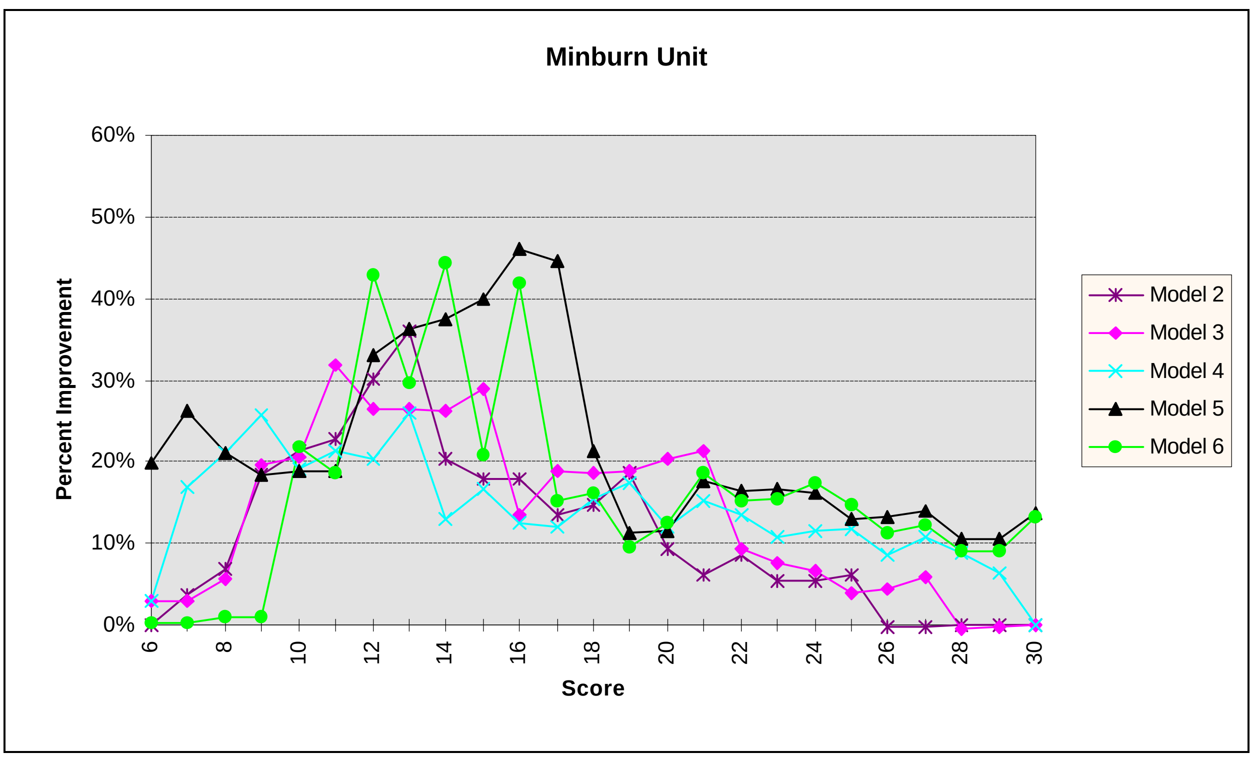


Figure 32. Improvement over chance (compare with chart on page 35)

Model 5. Model 5 was developed using the three variables with the highest Chi-square values (see descriptive modeling in Section IV.A):

Multiplier 3	Landscape position
Multiplier 2	GLO vegetation
Multiplier 1	Proximity to valley

Chi-square measures for these three variables ranged from 74 to 97, while Chi-square measure for the other three variables ranged from 21 to 49. Though the spatial distribution and Chi-square values for *GLO vegetation* and *proximity to valley* were similar, the CAC was 49 percent (Figure 33), indicating a moderate overlap (see Section IV.A).

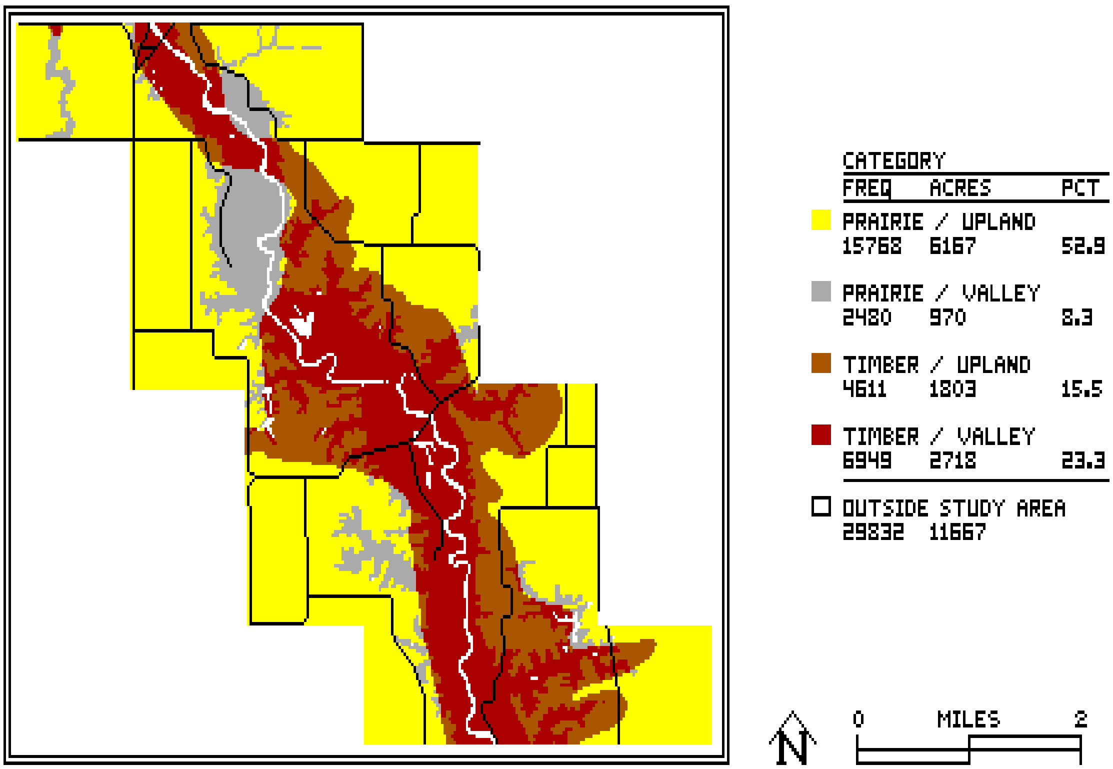


Figure 33. Spatial intersection of valley and GLO timber (CAC = 49%)

Points assigned to categories of *landscape position* and *proximity to valley* remained the same as in Model 3 (see Appendix C). Points assigned to *GLO vegetation* categories were based on frequencies computed during descriptive modeling:

5 pts.	Timber
1 pt.	Prairie

Model 5 results (Figure 34) showed a 46.0 percent improvement over chance. This was a higher level of improvement than Models 2, 3, or 4 (Figure 32). The substitution of *GLO vegetation* for *proximity to confluences* was responsible for the increased level of improvement over chance.

Model 6. One of the recommendations at the conclusion of Phase 1 research was to base predictive modeling criteria on the results of descriptive models which compare characteristics of known sites to *known non-sites* (rather than to a random sample of non-sites). The hypothesis was that these predictive models would have greater predictive power (as measured by the logit modeling improvement over chance) than predictive models based on the results of descriptive models involving a random sample of non-sites.

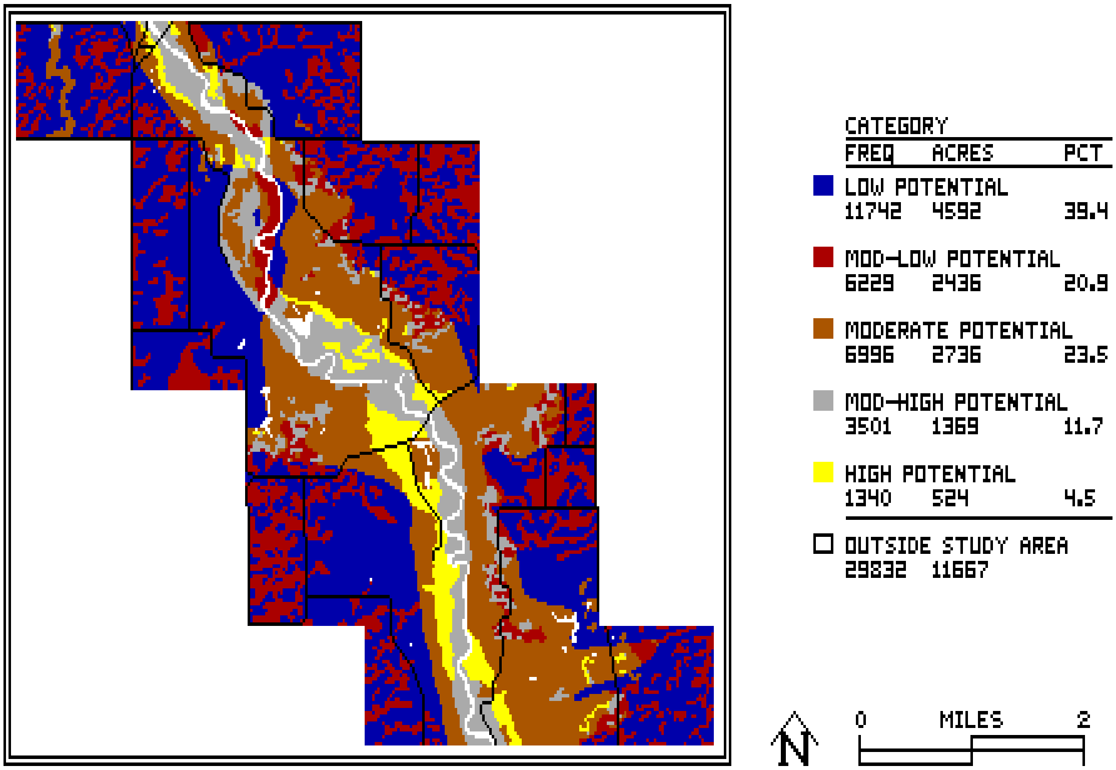


Figure 34. Model 5 for the Minburn Unit

As described in Section III.B, the field team mapped known non-sites when surveying the Minburn Unit. The location and extent of these known non-sites was digitized on a new data layer. Descriptive modeling then compared the characteristics of 21 known sites with 37 known non-sites. New Chi-square values were computed and compared to Chi-square values computed earlier:

Phase 2 Variable Data layer	**New Chi-square Cells**	**New Chi-square Normalized**	**Earlier Chi-square Cells**	**Earlier Chi-square Normalized**
Landscape position	285.150	100	97.340	100
Proximity to valley	222.126	78	73.654	76
Native vegetation	81.556	29	48.986	50
Proximity to confluences	72.404	25	35.127	36
GLO vegetation	37.417	13	81.412	84
Land cover	29.068	10	21.217	22

The variable *landscape position* was first on both lists. *Proximity to valley* moved from third to second highest, but its normalized Chi-square value (relative to *landscape position)* was almost the same (78 versus 76). The largest change in the two lists involved *native vegetation* and *GLO vegetation*. *GLO vegetation* moved from second to fifth on the list; its normalized Chi-square value decreased from 84 to 13. Though *native vegetation* moved from fourth to third on the list, its normalized Chi-square value decreased (from 50 to 29), slightly above the normalized value for *proximity to confluences* (25).

On this basis, Model 6 was developed using the three variables with the highest new Chi-square values:

Multiplier 3	Landscape position
Multiplier 2	Proximity to valley
Multiplier 1	Native vegetation

Points assigned to categories of *landscape position* and *proximity to valley* remained the same as in Model 3 (see Appendix C). Points assigned to *native vegetation* categories were based on frequencies computed during descriptive modeling:

5 pts.	Grass & trees, mesic prairie
3 pts.	Trees
1 pt.	Dry prairie, wet prairie, water

Model 6 results (Figure 35) showed a 44.3 percent improvement over chance. This was a higher level of improvement than for Models 2, 3, or 4, but lower than for Model 5 (Figure 32). The substitution of *native vegetation* for *GLO vegetation* was responsible for the slight decrease in the level of improvement over chance.

As expected, the mean score for known non-sites was lower than the mean score for known sites:

15.2	Mean score for known non-sites (by cells)
17.9	Mean score for known sites (by sub-sites)
22.3	Mean score for known sites (by cells)
14.9	Mean score for sample of non-sites (by cells)

However, the mean score for the sample of non-sites was the lowest of the four values, slightly lower than the mean score for known non-sites (lower by a difference of 0.3). This may be explained by the fact that the field survey team examined areas they felt had at least some potential (rather than little or no potential) for archaeological sites.

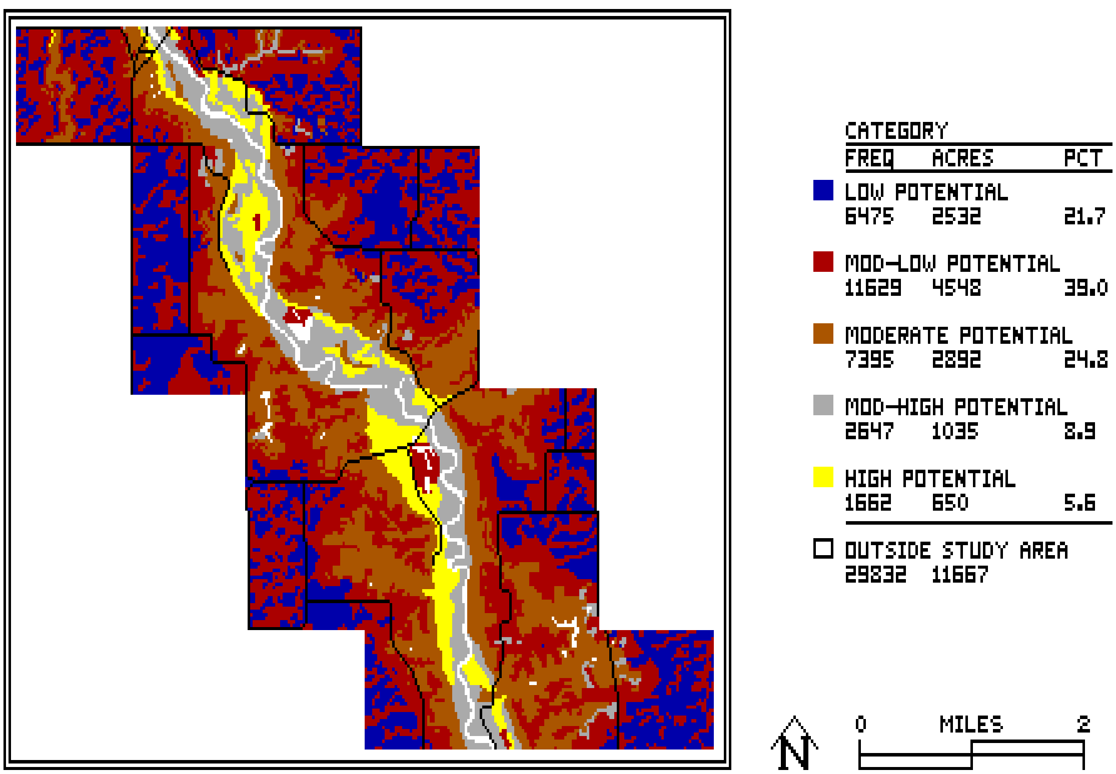


Figure 35. Model 6 for the Minburn Unit

V. Conclusions and recommendations

A. Predictive modeling

The predictive model with the greatest predictive power (as measured by logit modeling improvement over chance) was Model 5:

Model 2	Model 3	Model 4	Model 5	Model 6	
13	11	13	16	14	Cutpoint score (on a scale of 6 to 30)
0.29	0.21	0.29	0.42	0.33	Cutpoint score (on a scale of 0 to 1)
66.7%	96.7%	70.0%	79.3%	83.3%	Known sites correctly classified as high potential
69.4%	35.1%	56.0%	66.7%	60.9%	Non-sites correctly classified as low potential
30.6%	64.9%	44.0%	33.3%	39.1%	Study area classified as high potential
36.1%	**31.8%**	**26.0%**	**46.0%**	**44.3%**	**Improvement over chance**

Model 5 used modeling criteria (variables, multiplier, weights, points) based on descriptive modeling Chi-square values which compared characteristics of 21 known sites with a random sample of non-sites in the Minburn Unit. The improvement over chance (46.0%) was only slightly higher than for Model 6 (44.3%). Model 6 used modeling criteria based on descriptive modeling Chi-square values which compared characteristics of 21 known sites with 37 known non-sites in the Minburn Unit.

Both of these values for improvement over chance was less than those for Models 2 and 3 in the Phase 1 study area along the South Raccoon River corridor. There, Model 2 had a 55.7% improvement over chance and Model 3 had a 51.4% improvement over chance. However, Models 5 and 6 in the Minburn Unit had higher values for improvement over chance than some similar archaeological site modeling studies. Warren (1990b, p. 208) reported a 35% improvement (which he called "a moderate advantage over a chance classification"); Carmichael (1990, p. 221) reported a 27% improvement; and Kvamme (1992) reported improvements of 25%, 30%, and 48%. About the first, Kvamme (1992, p. 31) said, "comparatively speaking, this model was not very powerful."

In contrast, Models 5 and 6 in the Minburn Unit had lower values for improvement over chance than site modeling studies which used the same logit modeling procedure, but involved different resources. In predictive modeling studies for red squirrel distribution, Pereira and Itami (1991, p. 1482) reported improvements for three models: environmental (63%), trend surface (57%), and baysian (63%). In predictive modeling studies for wetland restoration in the Deep Loess region of southwest Iowa, Anderson (1996b, p. 9) reported improvements for two models: 82.1% and 83.0%.

This comparison suggests that Model 5 and Model 6 (improvement over chance 46.0% and 44.3%, respectively) are a useful and appropriate tool for predicting high-probability locations of archaeological resources. However, low to moderate levels of improvement over chance in all these studies of archaeological resources may be due to difficulties in studying past cultures and the archaeological evidence they left behind. For the most part, archaeological resources are hidden, inaccessible, unmapped, unstudied, and unprotected (Schiffer 1987). They also involve evidence of people who left no written record. In selecting locations for habitation, agricultural fields, storage, ceremonies, burial, and other activities, they used decision criteria involving some variables which are not easily mapped or which we may never discover.

Additional archaeological research is needed on past cultures in Iowa and on the nature and distribution of their archaeological evidence. GIS descriptive and predictive modeling techniques (similar to those used in this study) could be applied to new knowledge resulting from the research. Likewise, these descriptive and predictive modeling techniques could be applied to existing site records over an area larger than Dallas County, perhaps even the entire state (as is being done in other states, such as Arkansas, Mississippi, and Minnesota).

B. Hypotheses

As described earlier in Section I.B, hypotheses for Phase 2 of the research included the following:

1. Descriptive models which compare characteristics of known sites to *known non-sites* (rather than to a random sample of non-sites) result in predictive models which have greater predictive power (as measured by the logit modeling improvement over chance)
2. Predictive models customized specifically for the Minburn Unit (in the North Raccoon portion of the Greenbelt) have greater predictive power than those developed during Phase 1 for the South Raccoon portion of the Greenbelt

The first hypothesis was not supported by the research results. Model 6, which was based on known non-sites, had an lower improvement over chance (44.3%). Model 5, which was based on a sample of non-sites, had a higher improvement over chance (46.0%). However, the two values was quite similar (a difference of 1.7%, which may not be statistically significant). Even though the values may not be significantly different, the hypothesis was not supported because Model 6 does not have greater predictive power than Model 5 (as measured by improvement over chance). Because the two values for improvement over chance (46.0% and 44.3%) was similar, additional research is warranted to test this hypothesis in other study areas.

The second hypothesis was supported by the research results. Two of the three models developed specifically for the Minburn Unit (Models 5 and 6) had improvement over chance values (46.0% and 44.3%, respectively) that were higher than for Models 2 and 3 (36.1% and 31.8%, respectively). This measure of greater predictive power of Models 5 and 6 was due primarily to selection of variables that are appropriate for the local landscape.

This also explains why Model 4 had an improvement over chance (26.0%) that was lower than for Models 2 and 3. Model 4 used the same variables as Models 2 and 3, but assigned different multipliers (according to the Chi-square values from descriptive modeling). Therefore, the selection of variables had a greater effect in achieving high improvement over chance than the selection of multipliers.

These results emphasize the importance of developing predictive models for specific study areas. Differences in the landscape, natural resources, and use patterns by past cultures among different study areas warrant descriptive modeling to customize predictive models for each study area or each part of a study area.

C. Research objectives

Objectives of this research were described earlier in Section I.B:

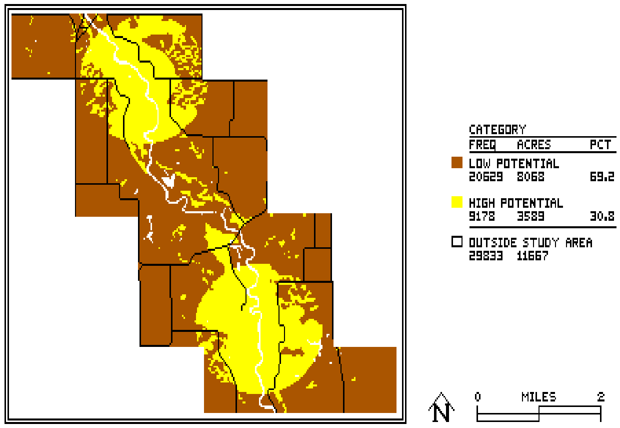
1. Document landscape characteristics of known archaeological sites
2. Help locate additional archaeological sites through predictive modeling
3. Provide context for landscape planning and management decisions

Modeling used GIS technology and statistical techniques to provide a means for meeting these objectives. Through descriptive modeling, 21 known sites in the Minburn Unit were described in terms of their proximity to stream confluences, landscape position, proximity to river valley, 1990 land cover, GLO vegetation, and native vegetation. To better understand the 21 known sites, characteristics of a random sample of non-sites were also described. Through frequencies and Chi-square measures, differences between known sites and non-sites were documented. Landscape position, proximity to valley, GLO vegetation, and native vegetation had the greatest differences. The use of Chi-square measures on the random sample of non-sites and the 37 known non-sites were helpful in descriptive modeling and as a basis for criteria in predictive modeling.

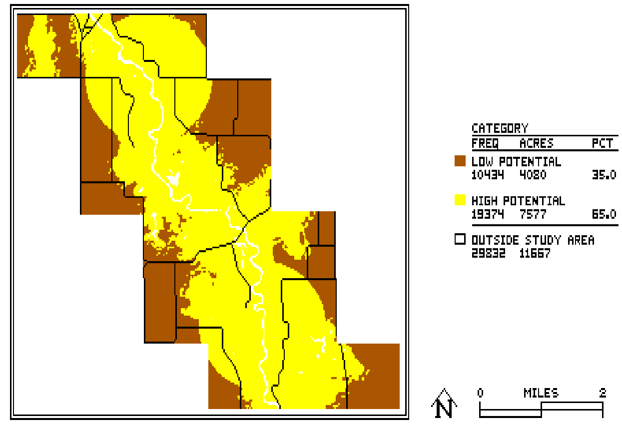
Predictive models were based on descriptive models, landscape archaeology principles, professional experiences of archaeological consultants, and statistical probabilities. Models 5 and 6 (which was based on all four) appear to be more successful in the Minburn Unit than earlier predictive models (which were based primarily on the last three).

Predictive models help locate additional archaeological sites and aid landscape planning and management because predictive models are both deterministic and empirical. They are deterministic because they use maps to show the potential at each location in the study area. They are empirical because the potential for additional archaeological sites is expressed in terms of numerical score which is based on a combination of three variables (data layers).

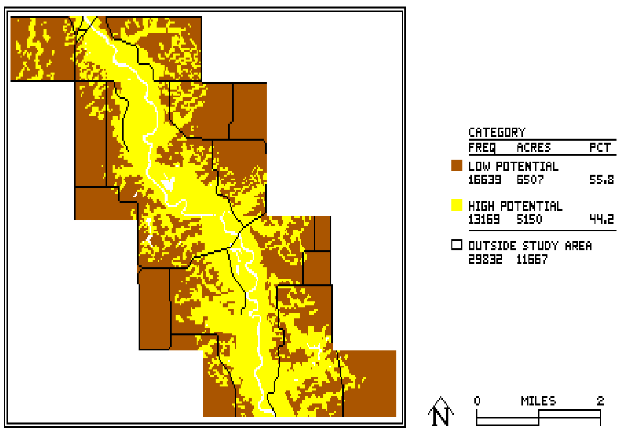
As described above in Section V.A, Model 5 resulted in the greatest improvement over chance of any of the models applied to the Minburn Unit. However, there are ways of using results from all the models. Figure 37 shows an overlay map of the cutpoint high potential areas resulting from Models 2, 3, 4, 5, and 6 (Figure 36). The intersection (overlap) area yields a CAC of 23 percent, indicating low agreement between the models. This intersection area amounted to approximately 11 percent of the Minburn Unit. However, the union of areas (high potential in at least one of the models) amounted to approximately 51 percent of the Minburn Unit. Even though logit modeling uses the concept of binary classification, this overly technique can help make further distinctions. The danger in this, of course, is that not all of the models being combined through overlays are uniform in quality or are equally appropriate for the study area.



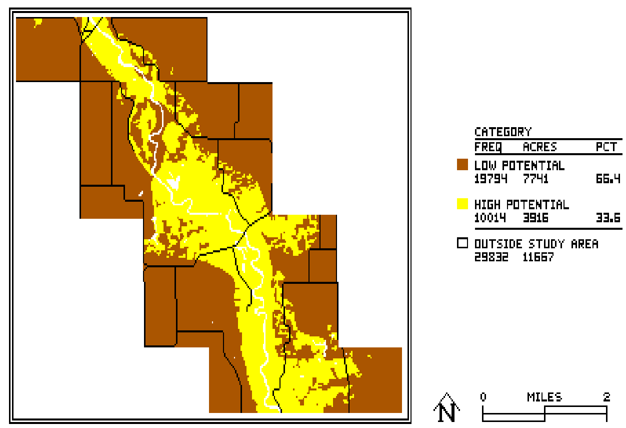
Model 2



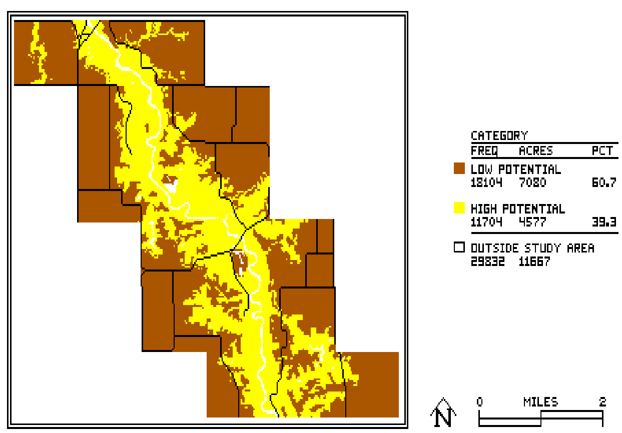
Model 3



Model 4



Model 5



Model 6

Figure 36. Cutpoint maps for Models 2, 3, 4, 5, and 6.

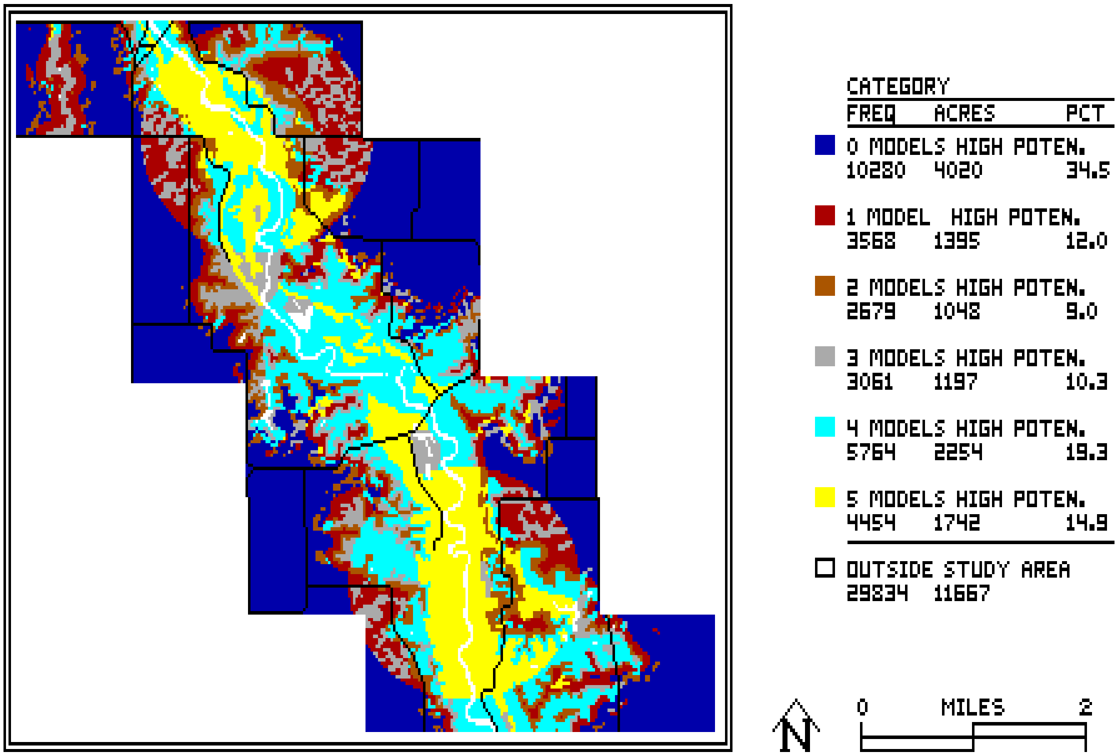


Figure 37. Spatial intersection of Models 2, 3, 4, 5, and 6

D. Additional refinements to predictive models

Predictive models for the Minburn Unit could be refined in several ways. First, multipliers (weights for variables) could be based more closely on the relative (normalized) Chi-square values of selected variables. The GIS software used for this study allows only integer numbers to be used for multipliers. Other GIS software may allow used of real numbers (in addition to integer numbers) for multipliers.

Second, descriptive modeling could explore the potential of more variables than the six used in this study. Warren (1990), for example, included 26 variables in descriptive modeling for a portion of Shawnee National Forest in southern Illinois. Anderson (1996b) included 24 variables in descriptive modeling of 19 wetland sites in Fremont County, Iowa.

Third, more than three variables (or, perhaps, fewer than three variables) could be included in predictive models for the Minburn Unit. However, one pitfall to be avoided is spatial autocorrelation, in which two or more variables have the effect of canceling the effectiveness and influence of each other (Unwin 1981, p. 134). In mapping approaches (such as in this study) that use map arithmetic, spatial autocorrelation can potentially reduce the effectiveness of predictive models because of central tendency (all portions of the study area with similar scores through “averaging”). To avoid the possibility of autocorrelation in this research for the Minburn Unit, CAC measures were used to evaluate the similarity in amount and spatial distribution between pairs of variables. This approach also helped avoid duplication and additional expense of digitizing and using similar variables.

E. Additional field surveys and modeling

The northern-most portions of the Raccoon River Greenbelt have not had archaeological surveys and GIS modeling similar to that already completed for the remainder of the Greenbelt. Areas needing field surveys and modeling include the Perry Unit and Dawson Unit, which together total approximately 26 square miles (approximately 13 percent of the complete Greenbelt area). Fortunately, this area is under the least development pressure and potential for negative impacts, primarily because this area is further from the Des Moines metropolitan area than the remainder of the Greenbelt.

VI. Bibliography

Anderson, Paul F. 1980. *Regional Landscape Analysis*. Environmental Design Press, Reston, Virginia, 248 p.

Anderson, Paul F. 1992. *Primer for Geodesy: GIS teaching software for landscape planning*. Department of Landscape Architecture, Department of Agronomy, Land Use Analysis Laboratory, Iowa State University, Ames, 10 p.

Anderson, Paul F. 1996a. *GIS research to digitize maps of Iowa 1832-1859 vegetation from Government Land Office township plat maps*. Iowa Department of Natural Resources and Department of Landscape Architecture, Iowa State University, Ames, 256 p.

Anderson, Paul F. 1996b. *GIS modeling of floodplain land uses and resource management alternatives in the Deep Loess region of western Iowa*. Golden Hills RC&D, Oakland, Iowa, and Department of Landscape Architecture, Iowa State University, Ames, 16 p.

Beavers, Glenn H. 1977. *A land information support program: MSDAMP Multi-Scale Data Analysis and Mapping Package*. Land Use Analysis Laboratory, Iowa State University, Ames, 124 p.

Bednarz, James C. 1979. *Status, habitat utilization, and management of Red-shouldered Hawks in Iowa*. MS thesis, Department of Animal Ecology, Iowa State University, Ames, 105 p.

Bettis, E. Arthur. 1987. History of the upper Mississippi valley. *Iowa Geology* 12, p. 12-15. Iowa Department of Natural Resources, Geological Survey Bureau, Des Moines. (QE111.I61x)

Bettis, E. Arthur. 1988. The role of geology in shaping the archaeological record. *Iowa Geology* 13, p. 12-15. Iowa Department of Natural Resources, Geological Survey Bureau, Des Moines. (QE111.I61x)

Butzer, Karl W. 1982. *Archaeology as human ecology: method and theory for a contextual approach.* Cambridge University Press, New York. 364 p. (CC81.B87.1982)

Carmichael, David L. 1990. GIS predictive modelling of prehistoric site distributions in central Montana. In *Interpreting space: GIS and archaeology*, Kathleen M.S. Allen, Stanton W. Green, and Ezra B.W. Zubrow, eds. Taylor and Francis, London, p. 216-225. (CC83.I58)

Clark, W.A.V. and P.L. Hosking. 1986. *Statistical methods for geographers.* John Wiley and Sons, New York. 518 p. (G70.3.C55)

Dillworth, Mary E., Jerry L. Whistler, and James W. Merchant. 1994. Measuring landscape structure using geographic and geometric windows. *Photogrammetric Engineering and Remote Sensing* 60: 10, p. 1215-1224.

Ebdon, David. 1985. *Statistics in geography.* Basil Blackwell, Oxford. 232 p. (G70.3.E23)

Fruhling, Larry. 1994. *Vital flood data not relayed.* Des Moines Register. p. 1M and 5M.

Gradwohl, David M. 1978. The Native American experience in Iowa: an archaeological perspective. In *The worlds between two rivers: perspectives on American Indians in Iowa* by Gretchen M. Bataille, David M. Gradwohl, and Charles L.P. Silet. Iowa State University Press, Ames, 148 p. (E78.I6.W67)

Hodder, I. 1979. Spatial patterns of the past: problems and potentials. In *Statistical applications in the spatial sciences*, Neil Wrigley, ed. Pion Limited, London, p. 189-202. (G70.3.S7)

Kleinbaum, David G. 1994. *Logistic regression: a self-learning text.* Springer-Verlag, New York. 282 p. (R853.S7.K54)

Krist, Frank J. and Daniel G. Brown. 1994. GIS modeling of Paleo-Indian period caribou migrations and viewsheds in northeastern Lower Michigan. *Photogrammetric Engineering and Remote Sensing* 60: 9, p. 1129-1137.

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*, W.J. Judge and L. Sebastian, eds. U.S. Bureau of Land Management: Denver, p. 325-428.

Kvamme, Kenneth L. 1990. GIS algorithms and their effects on regional archaeological analysis. In *Interpreting space: GIS and archaeology*, Kathleen M.S. Allen, Stanton W. Green, and Ezra B.W. Zubrow, eds. Taylor and Francis, London, p. 112-125. (CC83.I58)

Kvamme, Kenneth L. 1992. A predictive site location model on the high plains: an example with an independent test. *Plains Anthropologist* 37: 138, p. 19-40.

Minnick, R. F. 1964. A method for the measurement of areal correspondence. *Papers of the Michigan Academy of Science, Arts and Letters* 49, p. 333-344.

Moore, David S. and George P. McCabe. 1989. *Introduction to the practice of statistics.* W.H. Freeman, New York. 790 p. (QA176.12.M65)

Pazner, Micha, K. Chris Kirby, and Nancy Thies. 1989. *Formation Map II map processor: a geographic information system for the Macintosh.* John Wiley and Sons, New York. 165 p.

Pereira, Jose and Robert M. Itami. 1991. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham Red Squirrel. *Photogrammetric Engineering and Remote Sensing* 57: 11, p. 1475-1486.

Prior, Jean C. 1991. *Landforms of Iowa.* University of Iowa Press, Iowa City. 153 p.

Ruhe, Robert V. 1969. *Quaternary landscapes in Iowa.* Iowa State University Press, Ames. 255 p.

Schiffer, Michael B. 1987. *Formation processes of the archaeological record.* University of New Mexico Press, Albuquerque. 428 p. (CC80.S335)

Unwin, David. 1981. *Introductory spatial analysis.* Methuen, London. 212 p. (GA23.U59x)

U.S. Department of Agriculture. 1983. *Soil survey of Dallas County, Iowa.* Soil Conservation Service, Washington, DC, map scale 1:15,840, 168 p.

Warren, Robert E. 1990a. Predictive modelling in archaeology: a primer. In *Interpreting space: GIS and archaeology*, Kathleen M.S. Allen, Stanton W. Green, and Ezra B.W. Zubrow, eds. Taylor and Francis, London, p. 90-111. (CC83.I58)

Warren, Robert E. 1990b. Predictive modelling of archaeological site location: a case study in the Midwest. In *Interpreting space: GIS and archaeology*, Kathleen M.S. Allen, Stanton W. Green, and Ezra B.W. Zubrow, eds. Taylor and Francis, London, p. 201-215. (CC83.I58)

Wrigley, Neil. 1976. *Introduction to the use of logit models in geography.* Institute of British Geographers, London. 33 p. (G70.23.W74)

Wrigley, Neil. 1977. *Probability surface mapping: a new approach to trend surface mapping.* Transactions of the Institute of British Geographers New Series 2, London. p. 129-140.

VII. Appendices

Appendix A. Data sources for GIS data layers

Appendix B. Attribute data for known archaeological sites (example)

Appendix C. Summary of predictive modeling criteria

Appendix A. Data sources for GIS data layers

Soil types, slope classes, erosion classes
Dallas County Soil Survey, Soil Conservation Service, US Dept. of Agriculture, October 1983, 4"/mile

Land cover/land use, transportation, utilities
Aerial photographs, color infrared, US Geological Survey, 5-15-83, 1.1"/mile
Aerial photographs, color infrared, Iowa Dept. of Natural Resources, 1983, approx. 2.64"/mile
Aerial photographs, black and white, ASCS, US Dept. Agriculture, Summer 1983, 8"/mile
Dallas County Soil Survey, Soil Conservation Service, US Dept. of Agriculture, October 1983, 4"/mile
Topographic quadrangle maps, US Geological Survey, 1985, 2.64"/mile
County transportation map, Iowa Dept. of Transportation, 1986, 0.5"/mile

Zoning
Dallas County Planning and Zoning map, 1974, updated 1990, 1”/mile

Elevation and slope aspect
Digital terrain tape, US Geological Survey, 1972

Rare plant and animal species and communities
Iowa Natural Areas Inventory, Iowa Dept. of Natural Resources, October 1989, listing

Archaeological sites
Iowa Office of the State Archaeologist, October 1989, listing
Iowa Office of the State Archaeologist, field mapping by Cindy Peterson, 1994-1996

Historic sites
Iowa Office of Historic Preservation, October 1989, listing

Forest, wetland, and prairie reserve areas
Dallas County Assessor and Iowa Dept. of Natural Resources, December 1989, listing

CRP acres
Soil Conservation Service and ASCS, US Dept. of Agriculture, December 1989, listing

Historic sites, archaeological sites, geologic sites
Public lands, valuable habitat areas
Dallas County Conservation Department staff, December 1989, listings, interviews, map annotations
Dept. of Natural Resources management biologist, December 1989, interview, map annotations

Geology, soil associations, watersheds, streams, drainage order
General land use, counties and boundaries
Iowa MSDAMP Database, ISU Land Use Analysis Lab, 1970-1975, 0.03"/mile

General Land Office (GLO) historic vegetation
GLO township plat maps, US Bureau of Land Management, 1847-1851, 2"/mile (see Anderson 1996a)

Appendix B. Attribute data for known archaeological sites in the Minburn Unit

Unit #	*Site ID*	*GV N*	*GLO vegetation*	*LC N*	*1990 Land cover*	*SM N*	*Soil Mapping Unit*	*Landscape Position*	*Native vegetation*	*DV N*	*Distance to Valley*	*DC N*	*Distance to Confluence*
Minb1	30	9	Timber	6	Old field	8	16821 Hayden	Upland ridge	Trees	4	0-200 ft.	9	1-2 mi.
				3	Sc. trees	1	73621 Lester	Upland ridge	Grass & trees	5	200-500 ft.		
Minb2	308	3	Timber	2	Pasture	2	73621 Lester	Upland ridge	Grass & trees	1	Valley	3	1-2 mi.
				1	Woodland	1	35671 Hayden-Storden	Valley wall	Trees	2	0-200 ft.		
Minb3	309	2	Timber	2	Woodland	2	56621 Moingona	Terrace	Grass & trees	2	Valley	2	1-2 mi.
Minb4	310	2	Timber	2	Cropland	2	56621 Moingona	Terrace	Grass & trees	2	Valley	2	0.25-0.5 mi.
Minb5	311	3	Timber	3	Pasture	3	56621 Moingona	Terrace	Grass & trees	3	Valley	3	0.25-0.5 mi.
Minb6	312	3	Timber	1	Sc. trees	3	16861 Hayden	Valley rim	Trees	3	Valley	3	>2 mi.
				2	Woodland								
Minb7	313	4	Timber	3	Sc. trees	4	73621 Lester	Upland ridge	Grass & trees	4	0-200 ft.	4	1-2 mi.
				1	Pasture								
Minb8	314	4	Timber	2	Woodland	4	16821 Hayden	Upland ridge	Trees	4	0-200 ft.	4	0.25-0.5 mi.
				2	Sc. trees								
Minb9	316	25	Timber	20	Cropland	20	48511 Spillville	PDFloodplain	Mesic prairie	25	Valley	19	0.5-1 mi.
				5	Road	4	15859 Spillville-Coland	PDFloodplain	Mesic prairie			6	1-2 mi.
						1	13149 Hanlon-Spillville	MPFloodplain	Mesic prairie				
Minb10	317	60	Timber	60	Cropland	15	53611 Hanlon	MWFloodplain	Mesic prairie	60	Valley	60	0.5-1 mi.
						31	56621 Moingona	Terrace	Grass & trees				
						14	16851 Hayden	Valley rim	Trees				
Minb11	318	2	Timber	2	Pasture	2	73621 Lester	Upland ridge	Grass & trees	2	0-200 ft.	2	1-2 mi.
Minb12	319	1	Timber	1	Woodland	1	73621 Lester	Upland ridge	Grass & trees	1	0-200 ft.	1	0.25-0.5 mi.
Minb13	320	1	Timber	1	Woodland	1	16831 Hayden	Upland sideslope	Trees	1	0-200 ft.	1	1-2 mi.
Minb14	321	2	Timber	2	Cropland	2	73621 Lester	Upland ridge	Grass & trees	2	0-200 ft.	2	0.5-1 mi.
Minb15	322	2	Timber	2	Cropland	2	16821 Hayden	Upland ridge	Trees	2	0-200 ft.	2	>2 mi.
Minb16	323	2	Prairie	2	Cropland	1	5511 Nicollet	Upland sideslope	Mesic prairie	1	0-200 ft.	2	>2 mi.
						1	73621 Lester	Upland ridge	Grass & trees	1	200-500 ft.		
Minb17	324	9	Prairie	9	Cropland	2	5511 Nicollet	Upland sideslope	Mesic prairie	3	200-500 ft.	9	1-2 mi.
						7	13821 Clarion	Upland ridge	Mesic prairie	6	500-1000 ft.		
Minb18	325	3	Prairie	3	Cropland	3	13821 Clarion	Upland ridge	Mesic prairie	3	500-1000 ft.	3	1-2 mi.
Minb19	326	2	Timber	2	Cropland	1	10711 Webster	Upland flats	Wet prairie	2	0-200 ft.	2	1-2 mi.
						1	73621 Lester	Upland ridge	Grass & trees				
Minb20	327	2	Prairie	2	Cropland	2	13821 Clarion	Upland ridge	Mesic prairie	2	500-1000 ft.	2	1-2 mi.
Minb21	328	2	Timber	2	Cropland	1	13821 Clarion	Upland ridge	Mesic prairie	2	200-500 ft.	2	0.5-1 mi.
						1	6242 Storden	Valley rim	Xeric prairie				

Appendix C. Summary of predictive modeling criteria

Model 1

3 wt. **Proximity to stream confluences**
- 5 pts. 0.01-0.25 miles
- 4 pts. 0.26-0.50 miles
- 3 pts. 0.51-1 mile
- 2 pts. 1.01-2 miles
- 1 pt. >2 miles

2 wt. **Landscape position**
- 5 pts. Terraces--elevated floodplain ridges
- 4 pts. Floodplain ridges--moderately well-drained first bottoms
- 3 pts. Alluvial fans--elevated deposits of local alluvium
- 2 pts. Wetlands--poorly-drained floodplain and uplands
- 2 pts. Mixture--moderately well-drained and poorly drained first bottoms
- 1 pt. Other landscape positions

1 wt. **Proximity to valley**
- 5 pts. 0.01-200 feet
- 4 pts. 201-500 feet
- 3 pts. Valley
- 2 pts. 501-1000 feet
- 1 pt. >1000 feet

Model 2

3 wt. **Proximity to stream confluences**

2 wt. **Landscape position**
- 4 pts. Alluvial fans--elevated deposits of local alluvium (note: was 3 pts. in Model 1)

1 wt. **Proximity to valley**

Model 3

3 wt. **Proximity to stream confluences**

2 wt. **Proximity to valley** (*note: was 1 in Model 2*)
- 5 pts. Valley (note: was 3 pts. in Model 2)
- 4 pts. 0.01-200 feet (note: was 5 pts. in Model 2)
- 3 pts. 201-500 feet (note: was 4 pts. in Model 2)

1 wt. **Landscape position** (*note: was 2 in Model 2*)
- 4 pts. Valley rim--upland bluff ridges, valley shoulder (note: not included in Model 2)
- 3 pts. Wetlands--poorly-drained floodplain and uplands (note: was 2 pts. in Model 2)

Model 4

3 wt. **Landscape position** (*note: was 2 in Model 2; was 1 in Model 3*)

2 wt. **Proximity to valley** (*note: was 1 in Model 2; was 2 in Model 3*)

1 wt. **Proximity to stream confluences** (*note: was 3 in Model 2 and Model 3*)

Model 5

3 wt. **Landscape position** (*note: was 2 in Model 2; was 1 in Model 3; was 3 in Model 4*)

2 wt. **GLO historic vegetation** (*note: not included in previous models*)
- 5 pts. Timber
- 1 pt. Prairie

1 wt. **Proximity to valley** (*note: was 1 in Model 2; was 2 in Model 3 and Model 4*)

Model 6

3 wt. **Landscape position** (*note: was 2 in Model 2; was 1 in Model 3; was 3 in Model 4 and Model 5*)

2 wt. **Proximity to valley** (*note: was 1 in Model 2; was 2 in Model 3 and Model 4, was 1 in Model 5*)

1 wt. **Native vegetation from soils** (*note: not included in previous models*)
- 5 pts. Grass & trees, mesic prairie
- 3 pts. Trees
- 1 pt. Dry prairie, wet prairie, water