NORTH AMERICAN ARCHAEOLOGIST, Vol. 24(4) 311-338, 2003

THE EFFECTS OF SAMPLE BIAS ON PALEOINDIAN FLUTED POINT RECOVERY IN THE UNITED STATES*

BRIGGS BUCHANAN
University of New Mexico

ABSTRACT

Paleoindian fluted point distribution in the contiguous United States is analyzed for possible sampling biases. Potential sample biases examined in this article include modern population, years since statehood, acreage in cultivation, and intensity of urban and residential development. Results suggest that at the state level Early Paleoindian fluted point samples are correlated with modern population, years since statehood, and acreage of urban development. A regression model developed to account for significant sampling biases demonstrates that regardless of sample biases several states in the Eastern United States have an overabundance of predicted Early Paleoindian fluted points. Sampling biases do not appear to affect fluted point varieties dating to the Middle Paleoindian period. Land cover characteristics are examined using principal components and cluster analyses in an effort to define biogeographical associations with fluted point types and possible environmental sampling constraints.

INTRODUCTION

Numerous studies have used the spatial distribution and abundance of Paleoindian fluted projectile points to imply the mode and tempo of colonization of the New World, the development of regional stylistic traditions, and the structure of prehistoric land use in the United States (Anderson, 1990; Anderson and Faught, 1998, 2000; Anderson et al., 1997; Anderson and Gillam, 2000; Blackmar, 2001; Brennan, 1982; Largent et al., 1991; Lepper and Meltzer, 1991; Martin, 1973, 1984; Mason, 1962; Meltzer and Bever, 1995). Because the abundance and

*An earlier version of this research was presented at the 39th annual meeting of the Southwestern Federation of Archeological Societies in Midland, Texas.

distribution of fluted points greatly influences interpretations made of the earliest well-documented occupation of the contiguous United States, it serves researchers employing this line of evidence that it be as nearly without sampling biases as possible in order to make reliable inferences from the data. The fluted point distributional data analyzed in this article were gathered from numerous sources and made available by Anderson and Faught (on Internet: http://www.anthro.fsu.edu/research/paleo/paleoind.html). Potential samples bias effects examined include recent population census data from the United States, number of years since inception of statehood, and area of land devoted to cultivation, and urban/residential use. Modern population, years since statehood, and urban development vary significantly with the abundance and distribution of Early Paleoindian points. In contrast, the potential sampling biases examined do not affect the occurrence of Middle Paleoindian fluted point varieties. Cultivation, a commonly suspected cause in increased fluted point recovery, is not significantly associated with any fluted point finds. Land cover characteristics, including acreage of shrubland, grassland, forest, and wetland, showing the distribution of general vegetation regimes are used to examine modern biogeographical associations with fluted point occurrences. Modern forested and wetland regions were most prominently associated with Early Paleoindian fluted points, whereas, fluted point varieties are associated with a variety of regionally specific environmental zones.

BACKGROUND

Paleoindian fluted point distributions play a critical role in the investigation of the initial settlement of North America. The issue of whether fluted points represent evidence of the First Americans or the fluorescence of an earlier population (Dillehay and Meltzer, 1991; Fiedel, 2000; Jablonski, 2002; Meltzer, 1995, 1997; Meltzer et al., 1997) remains embroiled in considerable controversy. Generally what is agreed upon, however, is that the first widespread archaeological signal of a Paleoindian presence in the contiguous United States is represented by Early Paleoindian fluted points and that the abundance and distribution of such artifacts provides evidence, albeit indirect, of the distribution of Paleoindians themselves (Shott, 2002). The logical link between the abundance of fluted points and Paleoindian population size and settlement patterns has not been explored fully as many confounding factors, such as the availability of raw materials, the spatial and temporal distribution of subsistence resources and water, are believed to influence this relationship. The potential sampling biases focused on in this article also influence the distribution and confound this interpretative link and should be made apparent.

Models of the Paleoindian colonization of North America have relied in part on the distribution of fluted points. Martin (1973, 1984; Mosimann and Martin, 1975) proposed a wave-of-advance model to explain, what appeared to researchers then, as not merely a coincidence of the arrival of First Americans and the extinction of megafauna. The wave-of-advance colonization model postulates that Clovis hunters spread rapidly across the United States without concern for environmental barriers and discarded fluted points evenly across the landscape. As the task of actually creating a nation-wide database of fluted point finds commenced, evidence began to emerge that indicated something different than Martin's proposed ubiquitous distribution of fluted points. The national fluted point database, recently compiled and published by Anderson and Faught (2000), revealed a patchy distribution of fluted points across the United States (shown in Anderson and Faught's [(2000:507)] fluted point distributional maps). The patchy distribution of fluted points, especially across most of Eastern North America, suggested a leapfrog pattern of migration where Paleoindian colonizers concentrated on particular staging areas that were unusually rich in critical resources (Anderson, 1990; Anderson and Gillam, 2000). It is imperative, therefore, to portray fluted point distributions as accurately as possible to distinguish among hypotheses and throw light on this remote period. Accuracy in inferring trends from fluted point numbers and distribution must take into account biases that potentially affect the sample analyzed.

The potential for sampling biases in fluted point distributions acknowledged by Anderson and Faught (2000:509) has not been examined at the sub-continental level that the data have been presented. Sampling biases have been documented in studies of fluted point distributions at the regional level and point to modern population and land use practices, most often cultivation, as correlative responses with fluted point counts (Lepper, 1983, 1985; Seeman and Prufer, 1984; Shott, 2002). These regional studies identified a correlation between increasing modern population and the occurrence of fluted points, suggesting that modern population potentially biases the number of fluted points recovered. Higher modern populations occurred with higher fluted point totals by area indicating that population may be a good proxy measure for intensity of artifact collection rates; whether that be through more archaeologists and artifact collectors finding fluted points, or the concomitant impacts to the land associated with higher populations resulting in the exposure of more fluted points. Mixed results, however, have been reported concerning the degree to which cultivation affects fluted point exposure and collection rates (compare Lepper, 1983 to Shott, 2002). In Shott's (2002) analysis of seven midwestern states the amount of cultivation was found not to co-vary with fluted biface counts. Lepper (1983), in contrast, found a significant relationship with cultivation and fluted point abundance.

METHODOLOGY

Considerable efforts have been made by Anderson and Faught (2000) in the compilation of fluted point data from published sources, state site records, and surveys. The fluted point data collected by Anderson and Faught constitute the database used in the following analysis. The database includes 12,791 fluted projectile points, 78% of which are Clovis and related unnamed types that comprise the Early Paleoindian (EP) fluted point sample (dating to ca. 11,500–11,000 BP). The remainder of the sample consists of fluted point varieties including: Folsom points (15%), Suwannee and Simpson points (4%), and Cumberland points (3%) (Anderson and Faught, 2000:509; fluted point varieties generally date to ca. 11,000–10,000 BP). The EP fluted point and the fluted point variety distributions are analyzed separately for indications of sampling biases as the research focus on the EP fluted point data is on colonization (Anderson and Gillam, 2000) and the focus on the Middle Paleoindian data is on the "settling in" process of regional adaptation (Meltzer, 2002).

Modern population figures used in the analysis are from published estimates made by the United States Census Bureau (2002). Population estimates are for July 2002 derived from the April 2000 population estimates base reflecting modifications of the Count Question Resolution Program. Estimates are made yearly after the decennial census and incorporate data from births, deaths, and domestic and international migration to update the census base counts.

Years since statehood is included in the analysis as a coarse-grain measure of the intensity of archaeological work. Years since statehood counts the number of years before a given state was incorporated into the United States. It is assumed that states with longer histories have fostered more formal archaeological work in terms of the establishment of colleges and universities and the formation of state archaeological offices and societies. Years since statehood probably reflects greater involvement of state and federal funds allotted to infrastructure building projects that increase the likelihood of ground disturbances.

Land cover characteristic data from satellite imagery made available by the United States Geologic Survey in the National Land Cover Database (NLCD) were used in the analysis. Land cover characteristics were derived using tasseled cap transformations (for explanation of standardization techniques see Huang et al., 2002) of Landsat 7 satellite imagery from early, peak, and late growing seasons (Homer et al., 2002; Vogelmann et al., 2001). Ancillary data used to develop land cover characterizations includes elevational data at 30m resolution acquired from the National Elevational Database, soil properties (available water capacity, organic carbon content, and quality) selected from the State Soil Geographic Data Base, image shape and texture determinations to distinguish land use types within similar land cover types, and image derivatives of percent imperviousness and percent tree canopy per pixel.

The mutually exclusive land cover characteristics used in this analysis combined several of the land cover classes published in the NLCD. Combined land cover categories of urban, forest, shrubland, grassland, cultivation, and wetland employed 12 NLCD land cover classes. The urban category includes acreage developed into high and low intensity residential and commercial/industrial/transportation use. Urban land cover consists of all area covered by paved roads, parking lots, and buildings. The forest land cover class includes acreage in deciduous, evergreen, and mixed forest. The cultivation land cover class includes acreage in row crops and small grains. The wetland land cover class includes woody and emergent herbaceous wetlands. Acreage of shrubland and grassland are land cover classes from the NLCD and used unchanged in this analysis. Total state area figures used in the analysis were from the NLCD as a number of discrepancies occurred with the state area figures reported in Anderson and Faught's database and the NLCD. Land cover characteristic data are used as proportions of the total state area.

Prior to analysis data were log-transformed to make differences in size relative rather than absolute (a constant was added prior to transformation of the EP fluted point and fluted point variety data due to the presence of zeros). All of the variables, except for acreage in forest, displayed distributions skewed to the right. Log-transformed variables in this study more closely approximated bivariate normality (e.g., Figure 1) and were used in the following correlation and regression analyses. Statistical analyses were run with Minitab version 11.2, except for the multiple regression backward elimination procedure that was run with SAS release 8. The non-parametric Spearman's rank procedure was used when transformed data clearly did not meet the assumptions of the regression models (normally distributed error terms and homoscedasticity (Sokal and Rohlf, 1997:456-457)). Spearman's coefficient is computed as the product-moment correlation coefficient on ranked data. Multiple linear regression was used to test the significance of a linear combination of various predictor variables to explain variation in fluted point occurrences. Null hypotheses in the regression analyses test if the slopes are equal to zero. Polarity and magnitude of regression coefficients in models are interpreted in relation to the number of fluted points.

Principal component analysis (PCA) was used to examine variation in the raw data between fluted point occurrences, area, population, and land cover characteristics. PCA is a multivariate technique for understanding variation and for summarizing measurement data potentially through variable reduction. PCA uses a rotation of the original coordinate axes to produce a new set of uncorrelated variables, or principal components, that are unit-length linear combinations of the original variables. Because PCA is exploratory, homoscedasticity and normality of the data are not required (Manly, 1994). Eigenvectors and eigenvalues produced through PCA demonstrate the polarity and magnitude of variation in each variable and the relative percentages of variation accounted for in each component respectively. Components are interpreted by grouping positive and negative loadings as a linear combination of weighted averages. Simplification in the interpretation of principal components is made by mentally eliminating features from the linear combination that have relatively small (in magnitude) loadings. A PCA on the correlation matrix employing standardized data was used in the analysis.

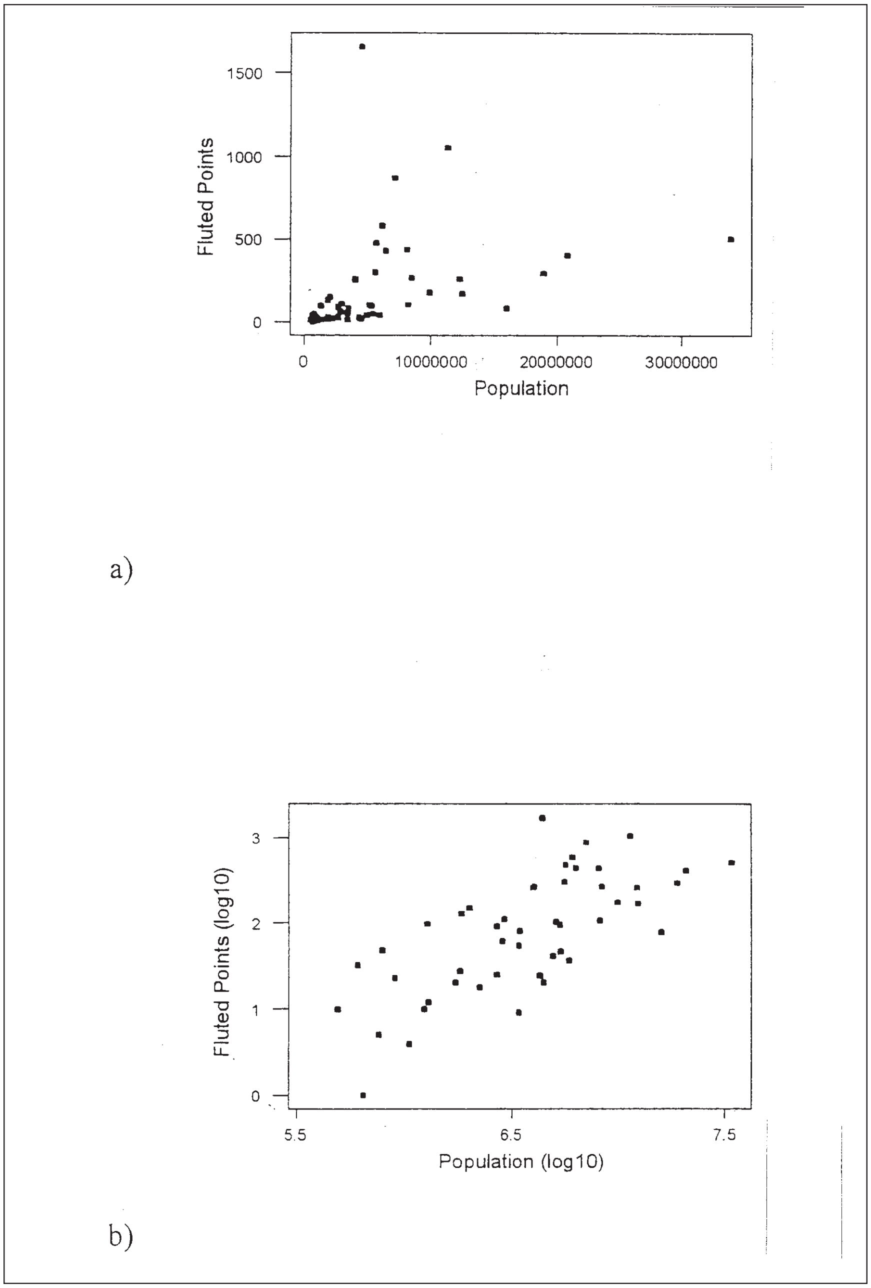

Figure 1. Bivariate plots of the number of fluted points and modern population by state. a) Raw data. b) Log-transformed data.

Cluster analyses were used to represent graphically the relationships between EP fluted points and fluted point varieties and the land cover characteristics described above. Variables are clustered using an agglomerative hierarchical clustering technique employing correlation as the distance measure.

RESULTS

Sample Biases by County

Attempts to analyze the statistical relationship between modern population and fluted point finds at the level of individual counties within states were hampered by the highly skewed nature of the data at this scale. Fluted point numbers, population, and area distributions by county remained significantly skewed to the right regardless of attempts to transform the data. The high proportion of counties without EP fluted points (60%) or fluted point varieties (86%) significantly contribute to the non-normal distribution of fluted point occurrences. Diagnostic plots of fluted points regressed against area and population reveal inconsistencies with the assumptions of the model. The residual plot should not exhibit systematic dependence of the sign or the magnitude of the residuals on the fitted values; however, a clear trend of positively arrayed outliers visibly violates the assumptions of the regression model for the EP fluted points (Figure 2a) and fluted point varieties (Figure 2b).

Correlation analyses of EP fluted points and fluted point varieties with area and population of county further illustrate the degree to which outliers influence the results of parametric tests. Comparison of the Pearson and Spearman's rho coefficients show large discrepancies suggesting the presence of influential outliers (Table 1). Analysis of sampling biases at the county level, therefore, is limited to the use and interpretation of the non-parametric Spearman's procedure. Examining the EP fluted point correlations with area and population demonstrates statistically significant correlations across all comparisons (after adjustment of the alpha level in concordance with the Bonferroni correction; Table 1). Partial correlation of the ranked variables demonstrates that after controlling for area, EP fluted point totals by county and population are significantly and positively correlated (adjusted $r^2 = 0.294$; $p = 0.000$). Fluted point varieties are positively correlated with area but, in contrast to EP fluted points, the data suggest fluted point varieties are not significantly correlated with population. Controlling for area, the correlation between fluted point variety totals by county and population remains negative and not significant (adjusted $r^2 = -0.0195$; $p = 0.280$). These results remain unaffected when point types are examined individually (i.e., Clovis, Folsom, Suwannee, Simpson, and Cumberland) against the variables.

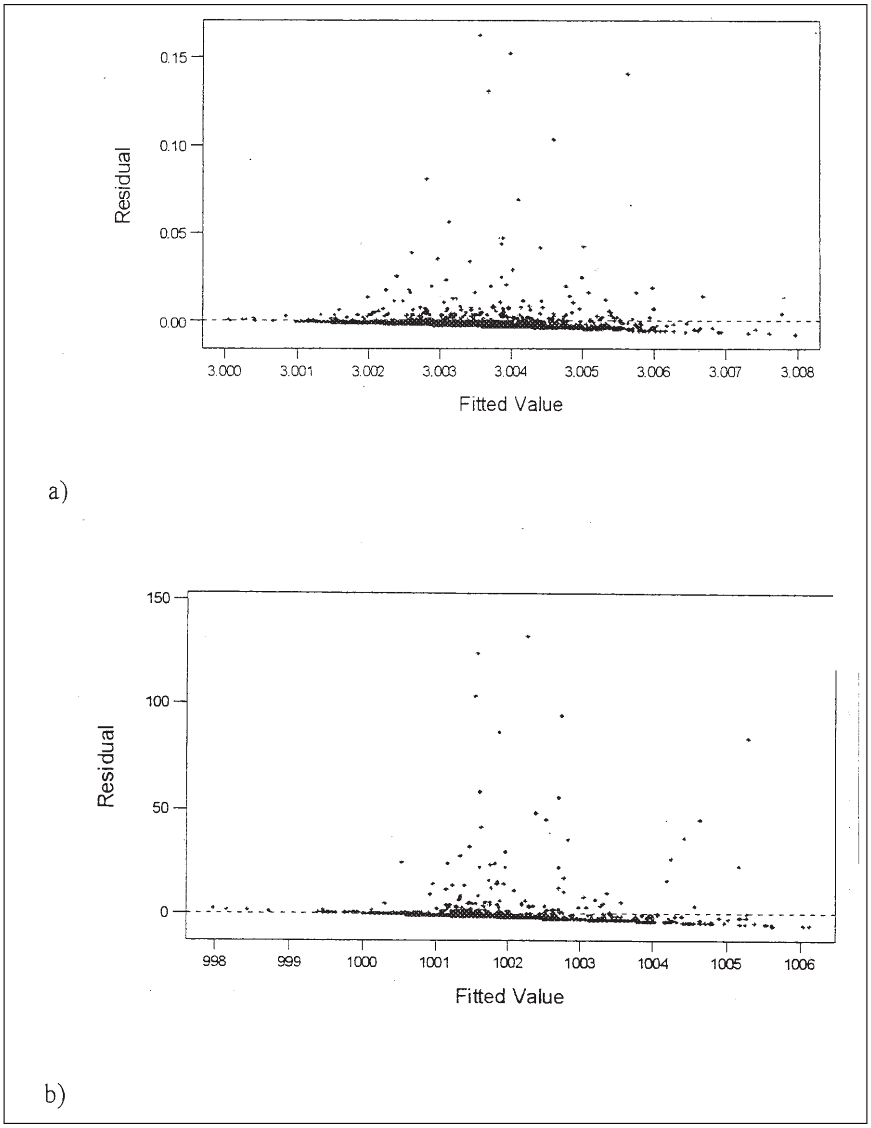

Figure 2. Plot of standardized residuals (constant variance of 1) against fitted values. a) Early Paleoindian fluted points by county regressed on area and modern population. b) Fluted point varieties by county regressed on area and modern population.

Table 1. Pearson and Spearman's Rho Correlations of Fluted Point Occurrences, Area, and Population at the County Level ($N = 3072$)

Correlation pair	Early Paleoindian fluted points[a] (Pearson)[c]	Early Paleoindian fluted points (Spearman's rho)[c]	Fluted point varieties[b] (Pearson)[c]	Fluted point varieties (Spearman's rho)[c]
Fluted-Population	.036 .043	.298* .000	−.006 .731	−.028 .123
Area-Population	.089* .000	−.055* .002	.089* .000	−.055* .002
Area-Fluted	.029 .109	−.115* .000	.081* .000	.155* .000

[a]Early Paleoindian points include points identified as Clovis and other unnamed variants.
[b]Fluted point varieties include Folsom, Cumberland, Suwannee, and Simpson points.
[c]Cell contents: top = correlation coefficient, bottom = p value.
*Correlation is significant at the Bonferroni adjusted $\alpha' = 0.008$ level (2-tailed).

Sample Biases by State

Multiple regression analysis of all fluted points (EP and fluted point varieties combined) on population and area at the scale of individual states within the contiguous United States exhibits significantly different slopes (rejection of null hypotheses) for population and area (Table 2). Examination of the plot of the residual against fitted values shows no clear dependence on polarity or magnitude of observations and therefore not in clear violation of the assumptions of the regression model (Figure 3). The data suggest that population is important in explaining the variation in all fluted points after area is taken into account and alternatively that area is important in explaining the variation in all fluted points after population is taken into account. Of interest here is the potential sample biasing effects of population; as area is factored out of distributional maps (Anderson and Faught, 2000) with the use of the ratio of fluted points divided by area. The positive population coefficient suggests predicted fluted point numbers increase as population increases within a given area.

A multiple regression backward elimination procedure was chosen to determine, as it was unknown prior to analysis, which variables were of most importance in a model. Regression of predictor variables solely on EP fluted point numbers permits investigation to focus on the period of the earliest widespread evidence of Paleoindian occupation. Variables used in the backward

Table 2. Multiple Regression Analysis Results of all Fluted Point Numbers per State Regressed on Area and Population of State (Regression Equation: Fluted points = –4.47 + 0.319 Area + 0.774 Population)

Predictor	Coefficient	Std. Dev.	*T*	*P*
Constant	–4.468	1.154	–4.73	0.000*
Area	0.3190	0.1516	2.10	0.041*
Population	0.7739	0.1663	4.65	0.000*

*Predictor is significant at $\alpha = 0.05$.

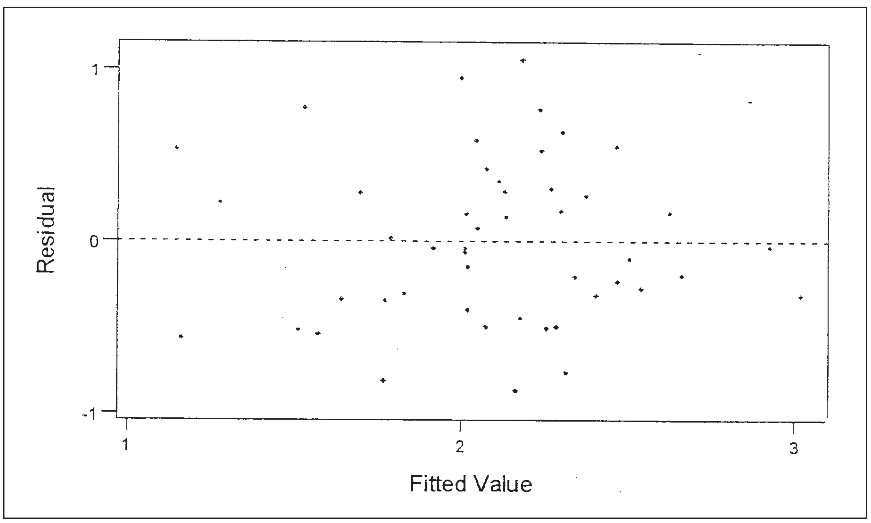

Figure 3. Plot of standardized residuals (constant variance of 1) against fitted values from all Paleoindian fluted points by state regressed on area and modern population.

elimination analysis include EP fluted point numbers and the potential sample biasing effects of area, years since statehood, population, cultivation, and urban development. The backward elimination procedure deletes unimportant variables, one at a time, starting from the full model (the full model and results are presented in Table 3). These variables when log-transformed approximate a bivariate normal distribution and all were included in the initial full multiple regression model (Figure 4). The full model returned an r^2 of 60.3% (Table 3). The first step removed acreage in cultivation as being of least importance in the model reducing

Table 3. Summary Results of the Backward Elimination Model Selection Procedure for Multiple Regression Analysis. Full model: Early Paleoindian Fluted Point = $\beta_0 + \beta_1$area + $\beta_2$population + $\beta_3$urban + $\beta_4$cultivation + $\beta_5$years since statehood + ε

Predictor	Coefficient	Std. Dev.	*F*	*P*
Constant	–12.71286	2.64512	23.10	< 0.001
Area	–0.646601	0.60600	1.14	0.2924
Population	1.15854	0.17270	45.01	< 0.001*
Urban	–0.53509	0.22329	5.74	0.0209*
Cultivation	–0.05976	0.13114	0.21	0.6509
Years	2.75612	0.87168	10.00	0.0028*

*Variables left in model are significant at $\alpha = 0.10$.

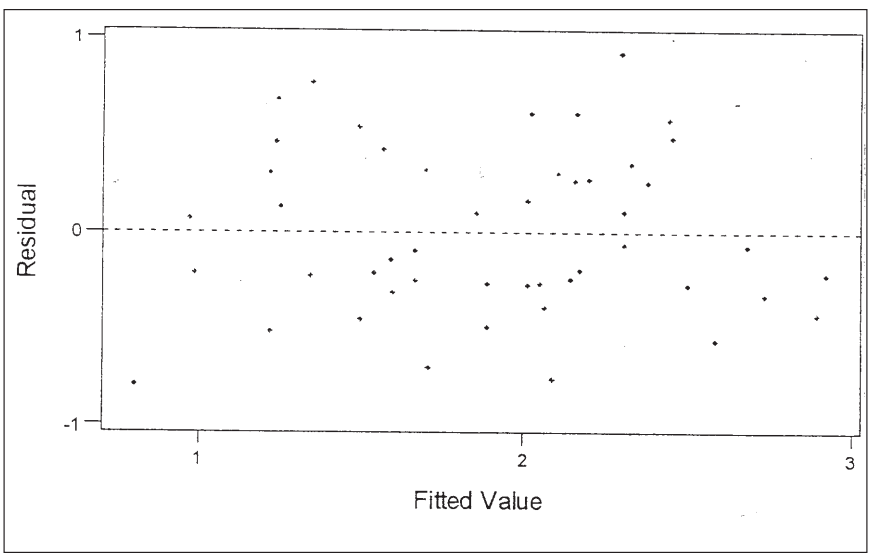

Figure 4. Plot of standardized residuals (constant variance of 1) against fitted values for the regression model including Early Paleoindian fluted points by state on area, years since statehood, population, acreage in cultivation, and urban development.

the r^2 to 60.1%. The second step removed area from the model and reduced the r^2 to 59%. The final model, including only the variables significant at the $\alpha = 0.10$ level, consists of population, years since statehood, and urban development.

The final r^2 of 59% from the backward elimination procedure is improved only slightly (60.1%) by the removal of four states (Alabama, New Mexico, North Dakota, and Rhode Island) with large standardized residuals and therefore were kept in the analysis. States with large standardized residuals are of interest, however, as they provide information on anomalous fluted point occurrences. Alabama has the largest number of EP fluted points (ranked 1st) and relatively large area (ranked 26th). New Mexico also has a relatively large number of EP fluted points (ranked 18th) and large area (ranked 3rd). In contrast, North Dakota, although being one of the largest states (ranked 15th) has no fluted points listed in the Anderson and Faught survey (ranked 48th); and Rhode Island being the smallest state (ranked 48th) also has a small number of fluted points (ranked 47th).

The results of the regression analysis indicate that population, years since statehood, and urban development are important in explaining variation in the abundance and distribution of EP fluted points. Population and years since statehood have positive coefficients indicating that predicted EP fluted point numbers increase as these variables increase. Urban development has a negative coefficient indicating a decrease in acreage in urban development with increasing EP fluted points. Fitted values are calculated for each state given the regression model (EP fluted point = $\beta_0 + \beta_{1population} + \beta_{2urban} + \beta_{3years} + \varepsilon$) and when compared with actual values, after the coefficients are transformed back into units expressed in the raw data (using antilog), show the states that have more or fewer than predicted fluted points (Figure 5). States with significantly more EP fluted points than expected are located primarily in the Eastern United States.

Examination of predictor variables in regard to fluted point varieties focuses on the settling in process of the Middle Paleoindian period. Regression of fluted point varieties on area, years since statehood, population, cultivation, and urban land use resulted in no significant relationships (area, $p = 0.330$; years since statehood, $p = 0.149$; population, $p = 0.563$; cultivation, $p = 0.321$; urban, $p = 0.613$). The reliability of these results, however, is doubtful based on the spread of residuals against fitted values (Figure 6). Discrepancies between Pearson and Spearman's rho coefficients also highlight the presence of potentially influential outliers (Table 4). Therefore, analysis is limited to the less robust, but more appropriate non-parametric Spearman's rank results. Statistically significant positive correlations were found with fluted point varieties and urban development. Partial correlation of the ranked variables, however, did not yield significant differences. Fluted point varieties are not significantly correlated with urban development when area is controlled for (adjusted $r^2 = -0.1339$; $p = 0.370$). Population and acreage in cultivation also are not significantly associated with fluted point varieties. The relationship between fluted point varieties and

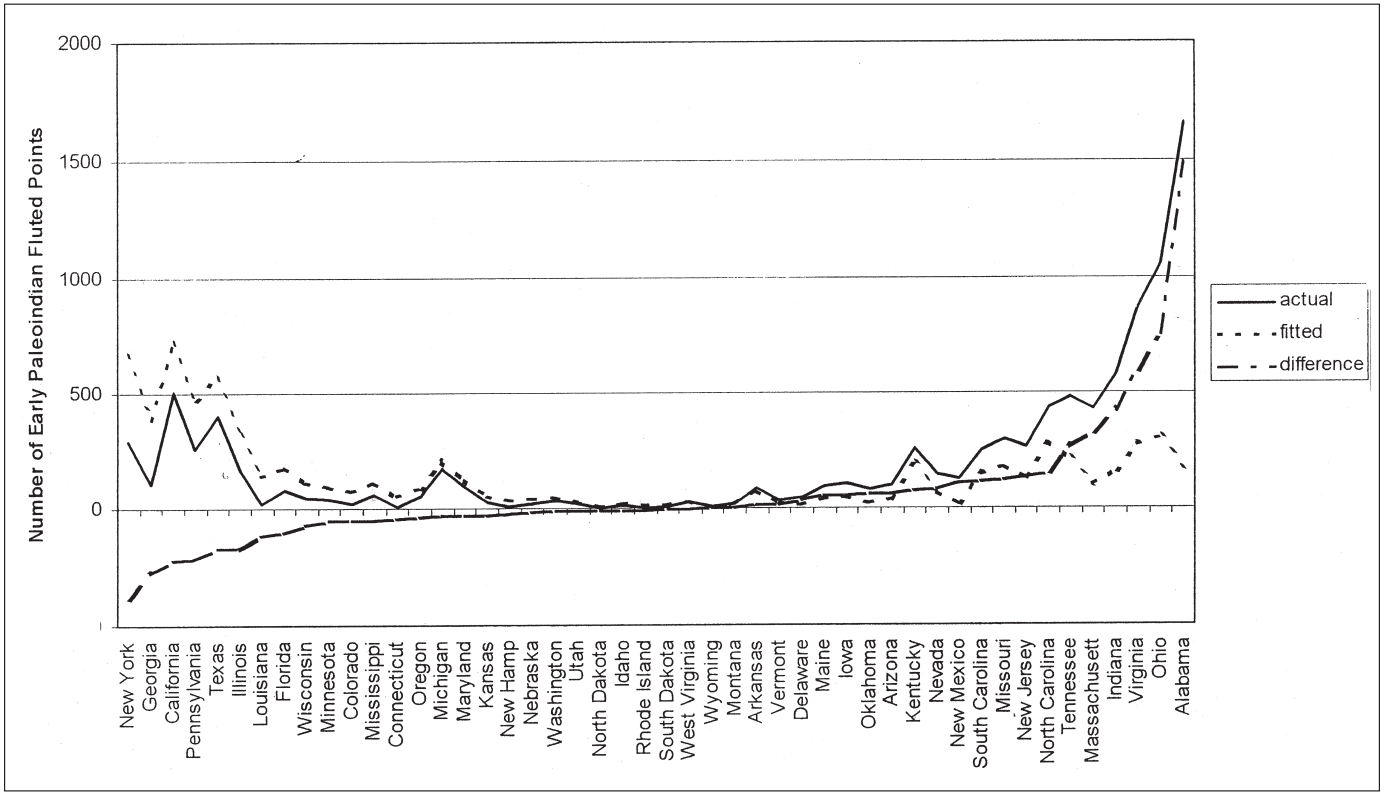

Figure 5. Line chart showing actual, fitted, and difference between actual and fitted Early Paleoindian fluted point numbers by state. Fitted Early Paleoindian fluted point numbers are derived from multiple regression model with population, years since statehood, and urban development as predictors.

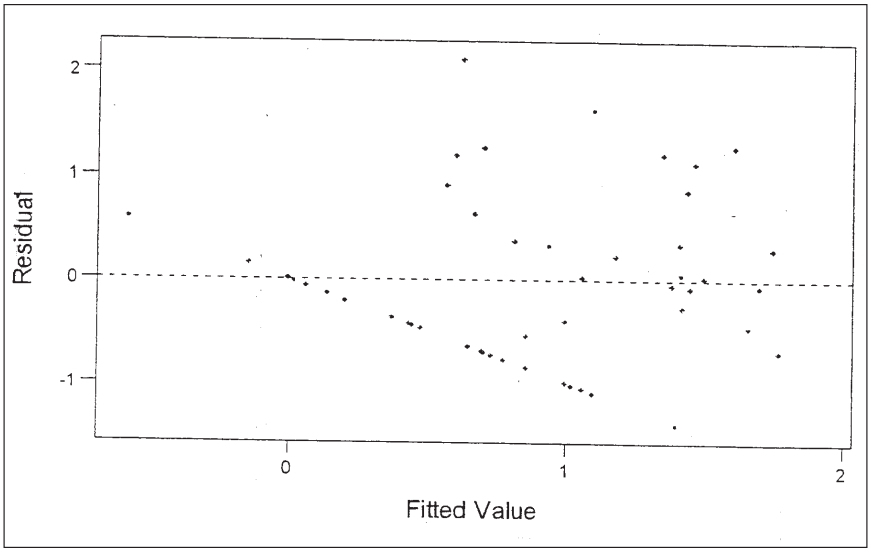

Figure 6. Plot of standardized residuals (constant variance of 1) against fitted values for the regression model including fluted point varieties on area, years since statehood, population, acreage in cultivation, and urban development.

population does not change after controlling for area (adjusted $r^2 = -0.0951$; $p = 0.525$) or between fluted point varieties and cultivation after controlling for area (adjusted $r^2 = -0.1686$; $p = 0.257$).

Biogeographical Associations with Fluted Points

PCA of EP fluted points and the defined variables produced nine principal components (PC), the first four of which account for more than 79% of the variation and are interpreted here (the remaining five PCs have less than 10% of the variation each; Table 5). The first PC, accounting for 36.5% of the variation, shows a weighted average of EP fluted points, population, years since statehood, forest, and urban development juxtaposed with shrubland and grassland (Figure 7a). Cultivation can be ignored in PC1 due to low variability. PC2 accounts for 17.4% of the variation and with PC1 makes up over half of the overall variation. In PC2, EP fluted points are associated with population, forest, and shrubland contrasted with a linear combination of cultivation, grassland, and wetland (Figure 7b). Years since statehood and urban can be ignored in PC2. PC3 and PC4 account for a minor amount of the variation (14.3% and 11.1%, respectively) and show some residual variation inconsistent with PC1 and PC2.

Table 4. Pearson and Spearman's Rho Correlations of Fluted Point Variety Occurrences, Area, Years since Statehood, Population, Acreage of Urban Development, and Cultivation at the State Level ($N = 48$)

Correlation pair	Fluted point varieties (Pearson)[a]	Fluted point varieties (Spearman's rho)[a]
Fluted-Area	.364 .011	.518* .000
Fluted-Years	−.266 .068	−.494* .000
Fluted-Population	.104 .481	−0.84 .572
Area-Years	−.613* .000	−805* .000
Area-Population	.368 .010	−.004 .977
Fluted-Urban	−.110 .459	−.456* .001
Fluted-Cultivation	−.179 .225	.138 .349
Urban-Cultivation	−.182 .215	.070 .636
Area-Urban	−.505* .000	−.730* .000
Area-Cultivation	−.060 .684	−.012 .936

[a]Cell contents: top = correlation coefficient, bottom = p value.
*Correlation is significant at the Bonferroni adjusted $\alpha' = 0.005$ level (2-tailed).

Table 5. The First Four Principal Components (PC) from the PCA

Variable	PC1	PC2	PC3	PC4
Eigenvalue	*3.2834*	*1.5680*	*1.2826*	*1.0006*
Proportion	*0.365*	*0.174*	*0.143*	*0.111*
EP Fluted	0.204	–0.137	–0.650	–0.274
Population	0.126	–0.133	–0.621	0.521
Years	0.507	0.062	–0.037	–0.078
Urban	0.350	0.004	0.188	0.416
Cultivation	–0.071	0.705	–0.271	–0.166
Forest	0.417	–0.338	0.198	–0.351
Shrubland	–0.339	–0.479	–0.042	0.280
Grassland	–0.442	0.109	0.017	0.048
Wetland	0.273	0.327	0.202	0.494

PC3 has a weighted average of EP fluted points with population, cultivation, and forest and PC4 shows just a weighted average of EP fluted points with forest and cultivation (Table 5).

Fluted point variants (Folsom, Cumberland, and combined Suwannee and Simpson) are examined separately in the PCA. Eleven PCs are distinguished in the analysis of fluted point variants and other variables. The first three PCs contain a majority of the variation (61%) and are interpreted here (Table 6). PC1 exhibits a linear combination of Folsom points, grassland, and shrubland in contrast to years since statehood, wetland, forest, and urban development associated to a limited degree with Suwannee/Simpson points; ignoring Cumberland points, population, and cultivation variables (Figure 8a). In PC2, a strong relationship between Suwannee/Simpson points, wetland, cultivation, and population is contrasted with forest, shrubland, and Cumberland points (Figure 8b); ignoring Folsom, years since statehood, and urban development variables. PC3 displays a weighted average of Suwannee/Simpson, population, shrubland, wetland, and urban development.

A dendrogram showing the results of the cluster analysis of the EP fluted points and land cover characteristic variables displays EP fluted points most closely linked to population (Figure 9a). EP fluted points and population are then linked to a nested arrangement of years since statehood, forest, urban development, and wetland. The second cluster contains shrubland, grassland, and acreage in cultivation. This dendrogram reflects the results of the PCA and the variation accounted for in PC1. Two clusters are evident in the dendrogram showing the

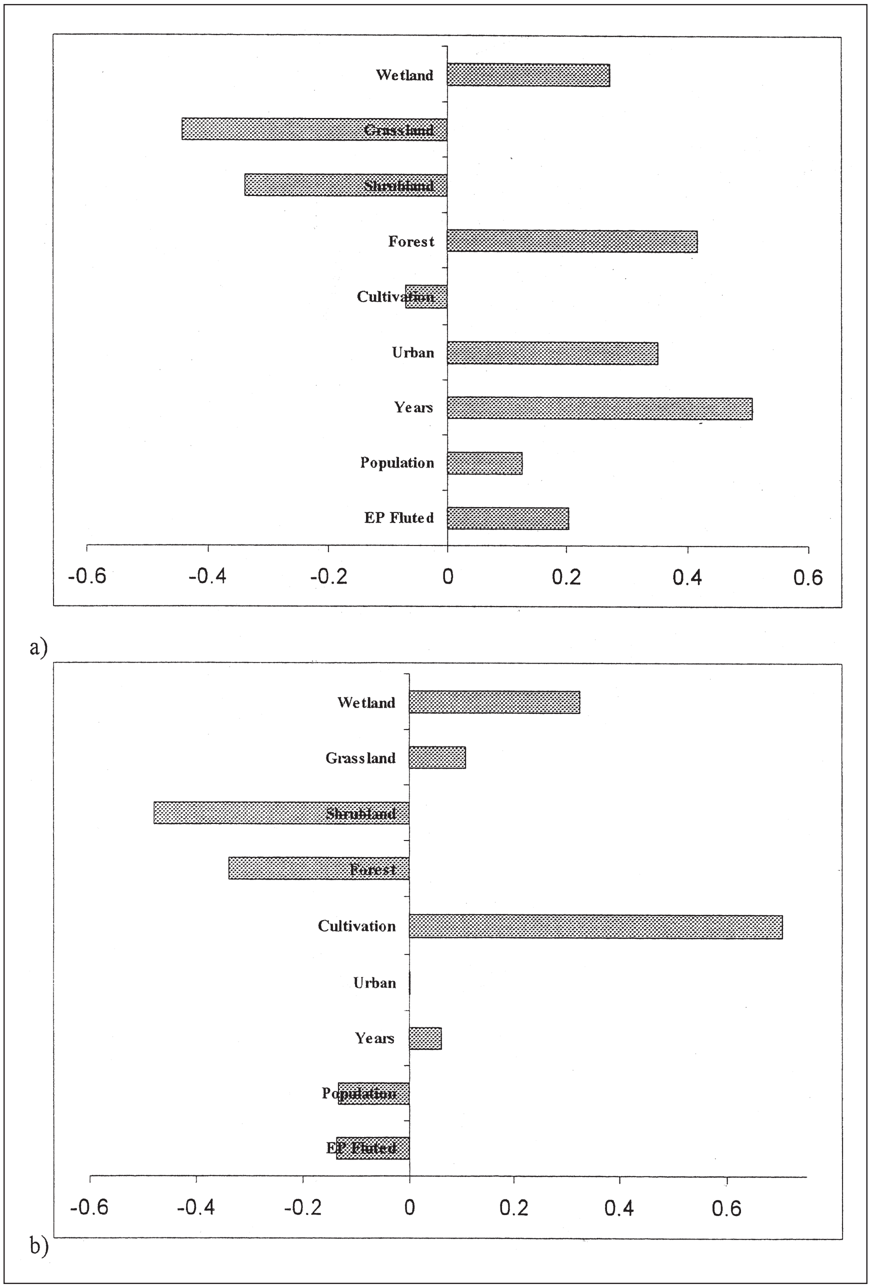

Figure 7. Principal component loadings from the eigen analysis of the correlation matrix. a) Bar graph of PC1 loadings. b) Bar graph of PC2 loadings.

Table 6. The First Three Principal Components (PC) from the PCA

Variable	PC1	PC2	PC3
Eigenvalue	*3.5300*	*1.6929*	*1.4943*
Proportion	*0.321*	*0.154*	*0.136*
Folsom	–0.321	0.044	–0.238
Cumberland	0.099	0.261	0.145
Suwannee/Simpson	0.102	–0.461	–0.455
Population	0.090	–0.136	–0.373
Years	0.475	0.064	0.085
Urban	0.328	–0.062	–0.213
Cultivation	–0.044	–0.473	0.582
Forest	0.386	0.426	–0.066
Shrubland	–0.336	0.232	–0.381
Grassland	–0.431	–0.148	0.006
Wetland	0.292	–0.458	–0.189

fluted point varieties and land cover characteristics (Figure 9b). The first cluster includes Folsom points linked to grassland, shrubland, and then cultivation. A cord links the first cluster to the second cluster consisting of Suwannee/Simpson points linked most closely to wetland, followed by years since statehood, forest, and urban. Cumberland points and population are linked to the second clusters as outliers.

DISCUSSION

The large proportion of counties without fluted points hampered the analyses of sampling biases at the county level. Non-parametric tests found EP fluted points significantly correlated with modern population after controlling for area. This correlation undoubtedly has several root causes among which some reasonable suggestions include increased numbers of archaeologists, artifact collectors, and the concomitant increase in impacts to the land that can result in more artifact exposure that goes along with higher populations. Another possible cause of the correlation is a connection between modern and EP population settlement decisions. Areas with high modern populations today are situated in the same general environmental settings that EP populations may have preferred as well, such as forested areas and along watercourses. In contrast to the EP fluted point findings, fluted point varieties dating to the Middle Paleoindian period were not

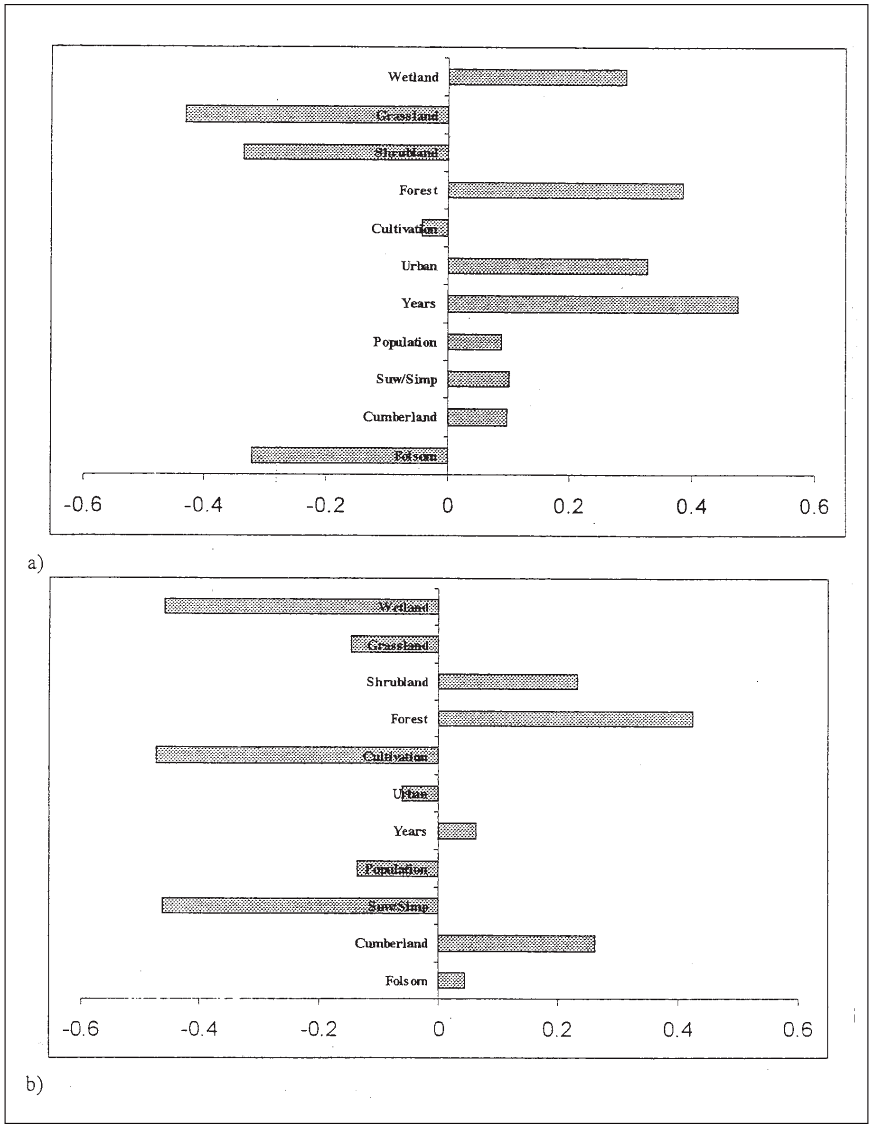

Figure 8. Principal component loadings from the eigen analysis of the correlation matrix. a) Bar graph of PC1 loadings. b) Bar graph of PC2 loadings.

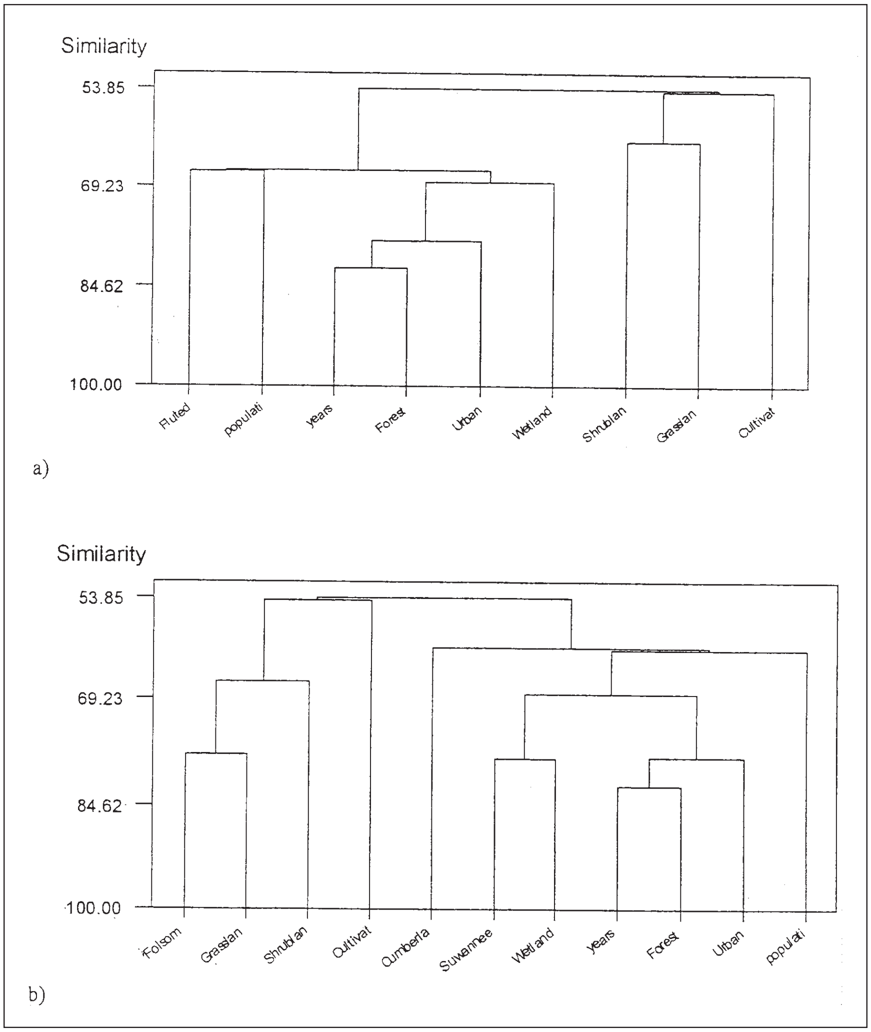

Figure 9. Dendrogram for single-linkage cluster analysis on the correlation matrix. a) Early Paleoindian fluted points in relation to population, years since statehood, and land cover characteristics. b) Fluted point varieties in relation to population, years since statehood, and land cover characteristics.

significantly associated with modern population at the county level after controlling for area.

The backward elimination multiple regression procedure selected population, years since statehood, and urban development as the most important variables in explaining variation in EP fluted points at the state level. Predicted EP fluted numbers using the coefficients from the regression equation showed that Alabama, Ohio, Virginia, Indiana, Massachusetts, and Tennessee have significantly more EP fluted points than predicted accounting for differences in population, years since statehood, and acreage of urban development. Survey effort may in part explain the abundance of EP fluted points in Ohio and Virginia, states that Anderson (1990:166) indicated had among the oldest and best fluted point surveys in the United States (McCary, 1984; Seeman and Prufer, 1982). Excavation data from the exceptionally productive site of Bull Brook I influences the density of points (ca. 400 points as reported in Anderson, 1990:175) in Massachusetts. Because excavation and surface collected data do not represent qualitatively different information (Dunnell and Dancey, 1983; although the differences in sampling effort between excavation and survey potentially are significant), the overabundance of predicted EP fluted points in these states biased by intensive work at productive sites should not be entirely discounted.

States with fewer than predicted EP fluted points include the two largest states in the contiguous United States, California and Texas, and several Eastern states including New York, Pennsylvania, and Georgia. An explanation for the lack of predicted EP fluted points in these states may be due in part to the dense urban populations in some of these states. As of 2000, New York had the largest city (New York City) in the country, California had two of the largest cities (Los Angeles–2nd and San Diego–7th) in the top ten, Texas had three of the largest cities (Houston–4th, Dallas–8th, and San Antonio–9th) in the top ten, and Pennsylvania had the fifth largest city (Philadelphia) in the country (U.S. Census Bureau, 2002). Concentrated populations in urban centers inflate overall state population figures and consequently the estimates made of predicted fluted point numbers taking population into account. The urban factor particularly seems to affect large states with concentrated populations where vast areas of the state may be relatively sparsely populated. Georgia, however, remains aberrant to the pattern of states with deficient EP fluted point counts, as Georgia does not have a pronounced urban population.

PCA and cluster analysis results also indicated an association of EP fluted points with high populations. Modern forested and wetland regions were most prominently associated with EP fluted points. The suggestion that fluted point samples have been biased toward more open landscapes opposed to wooded areas (Lepper, 1983) does not appear to be substantiated with these data. EP fluted points primarily co-occurred with forested areas. At a coarse-scale it appears that EP fluted points are found in significant numbers in the forested, and to a lesser degree wetland, interfluvial landforms of the Eastern U.S., a geomorphological setting conducive to surface finds (Goodyear, 1999:433). Limited, but steady, sedimentation in Prairie and Plains setting appears to work against surface exposure of artifacts, except in rare period of erosion such as the Dust Bowl (Seebach, 2000). The shrublands of the Western United States are to a lesser degree associated with EP fluted point finds but the overall relative deficiency in points may be related to a lower density of game animals in the similarly limited biomass environments during the Late Pleistocene and the consequently lower EP population.

Fluted point varieties (Folsom, Suwannee, Simpson, and Cumberland) were more difficult to examine due to the skewed and leptokurtic nature of the data. Fluted point varieties undoubtedly represent the settling in process of Paleoindians adapting to culturally and environmentally circumscribed regions. Visual inspection of fluted point variety distribution maps clearly shows separate clusters formed by these point styles (Anderson and Faught, 2000). Fluted point varieties, in contrast to EP fluted points, exhibited no significant relationships with modern population, urban development, or cultivation. Multivariate analyses of fluted point varieties and land cover characteristics showed a clear association of Folsom points with grassland, shrubland, and to a lesser degree cultivation. Folsom sites primarily are distributed across the Great Plains and the shrubland land areas to the west and south of this physiographic region. Suwannee and Simpson points were associated with areas of forest, wetland, and urban development in the Southwest, whereas Cumberland points predominantly were associated with the forested regions of the South.

The discrepancy between the results of the EP fluted point analyses, suggesting sampling biases toward modern population, years since statehood, and urban development, and the results of the fluted point variety analyses showing no significant relationships primarily may be a result of the spatially circumscribed nature of the fluted point variety distributions. The overall regression indicates population was a biasing factor in the sample amassed to date, the fluted point varieties, however, represent approximately a quarter of the total fluted point sample distributed in spatially discrete areas of relatively low population (states in the Great Plains and Southeast). In fact, EP fluted point occurrences by state are not significantly correlated with fluted point variety finds (Spearman's $r = -0.048$, $p = 0.744$) suggesting that different processes account for the distribution of Early and Middle Paleoindian fluted points. Population growth and cultural and environmental change at the regional level with potentially increasing territoriality into the Middle Paleoindian period may provide a partial explanation for the increased quantity of fluted point varieties in particular areas.

Counter-intuitively, EP fluted points and fluted point varieties are not correlated with acreage devoted to cultivation. Cultivation, often suspected as a process that contributes to the exposure and recovery of the numerous fluted point finds in the Eastern United States (Lepper, 1983, 1985; Seeman and Prufer, 1984), did not significantly correlate with fluted point numbers in this article. The results presented in this article corroborate Shott's (2002) finding that cultivation did not bias fluted point abundance in the Midwestern United States, and Anderson's (1990:171) qualitative assessment that ". . . the low incidence of Early Paleoindian artifacts in areas that are both heavily farmed and collected, as in large areas of the Gulf coastal plain, suggest(ing) that prehistoric rather than contemporary phenomenon are represented."

Another unexpected finding was the negative relationship found between EP fluted point numbers and acreage of urban development. Urban development initially was added as a possible factor in exposure and recovery of more fluted points due to ground disturbing impacts. This statistical relationship undoubtedly is the result of several factors that acted to deter fluted point finds in urban areas; among which may be the comparatively recent regulations imposing archaeological work for federally funded urban development and the potential for EP occupation, if not exposed on Pleistocene landforms, to be deeply buried and potentially well below impact depths associated with urban development construction activities (Ferring, 1994).

The states with more than predicted EP fluted points based on the regression model encompass the hypothesized Paleoindian staging areas of Anderson and Faught (Anderson, 1990; Anderson and Faught, 2000:509; Anderson et al., 1997) centered on the Tennessee, Ohio, and Cumberland drainage systems and across several other Eastern states. The uneven pattern of EP fluted point distribution across the United States, the impetus for the leap frog model of Paleoindian technological spread (Anderson and Faught, 2000; Anderson and Gillam, 2000), is supported, at least at the coarse-scale, after sampling biases were accounted for in the analysis. The fitted fluted point data do not support the wave-of-advance model and the hypothesized ubiquitous distribution of EP fluted points.

SUMMARY

The results of statistical analyses presented in this article indicate that at the county and state level Early Paleoindian fluted point samples correlate with modern population. Other sampling biases identified at the state level that significantly correlate with the abundance of Early Paleoindian fluted points are years since statehood and acreage of urban development. Predicted Early Paleoindian point numbers derived from the regression model developed to account for significant sampling biases demonstrate that the primary results gleaned from the fluted point distributional maps (Anderson and Faught, 2000) are not severely affected. Several Eastern states contain an overabundance of Early Paleoindian fluted points after taking into account sampling biases. The fluted point distribution data do not represent an evenly dispersed distribution that would support models of widespread coverage of the contiguous United States by Early Paleoindian fluted point makers. The proposed staging areas of Anderson and Faught (20900; Anderson, 1990) occur in the states with an overabundance of predicted points, including Alabama, Ohio, Virginia, Massachusetts, and Tennessee among others, evidence that provides support for their model. When compared to other states with fewer than predicted fluted points in the Southeast, including Georgia, Louisiana, and Florida, and the Northeast, including New York and Pennsylvania, situated in the same region as states with an overabundance of fluted points, the juxtaposition suggests that areas or environmental zones within certain states were favored by Early Paleoindians. Discrepancies in fluted point finds between states may relate to any number of factors including time depth, intensity of occupation, land use practices, and technological organization characteristics, however no inferential link between these factors and the pattern of fluted point occurrences has been developed in detail (Futato, 1982; Purdy, 1982; Storck, 1982).

This article relied on state level data to test for the presence of sampling biases. This coarse-level of analysis undoubtedly glosses over differences in sample biases at finer scales. Results from regional studies have shown that similar processes affecting artifact exposure and collection rates have occurred (Lepper, 1983, 1985; Seeman and Prufer, 1982, 1984; Shott, 2002) and provide support for the general trends described in this article. Although regional differences certainly exist, further analyses directed toward dissecting regional patterns to determine the consequences of potential differences are needed. Other potential sampling biases discovered in this article will be developed in future work. The possible affects of urban population density and intensity of state archaeological survey programs should be factored into future models to discount the potential effects of large cities and to better account for intensity of archaeological work. Meltzer and Bever's (1995) method of using the number of recorded sites for all time periods by county to gauge intensity of archaeological investigation may provide a better measure of archaeological work than years since statehood used in this article. These measures should improve the analysis of intra-state level sampling biases, but it remains, that after accounting for three statistically significant biasing factors a pattern of patchy fluted point distribution is evident in the record of fluted point distributions.

ACKNOWLEDGMENTS

I would like to thank Richard Strauss and Roger Moeller for their helpful comments and advice. This research was made possible through the tremendous efforts of David Anderson and Michael Faught in compiling and making accessible the fluted point database.

REFERENCES

ANDERSON, DAVID G.
1990 The Paleoindian Colonization of Eastern North America: A View from the Southeastern United States. In *Early Paleoindian Economies of Eastern North America* K. B. Tankersley and B. L. Isaac (eds.), Research in Economic Anthropology, *5*, pp. 163-216.

ANDERSON, DAVID G. and MICHAEL K. FAUGHT
1998 The Distribution of Fluted Paleoindian Projectile Points: Update 1998, *Archaeology of Eastern North America*, *26*, pp. 163-187.
2000 Palaeoindian Artifact Distributions Evidence and Implications, *Antiquity*, *74*, pp. 507-513.

ANDERSON, DAVID G., MICHAEL K. FAUGHT, and J. CHRISTOPHER GILLAM
1997 *Paleoindian Site/Artifact Distributions Viewed from a Very Large Scale: Evidence and Implications*. Paper presented at the 1997 Annual Society for American Archaeology Meeting, Nashville. (on Internet: http://www.anthro.fsu.edu/research/paleo/saa98.html).

ANDERSON, DAVID G. and J. CHRISTOPHER GILLAM
2000 Paleoindian Colonization of the Americas: Implications from an Examination of Physiology, Demography, and Artifact Distribution, *American Antiquity*, *65*:1, pp. 43-66.

BLACKMAR, JEANNETTE M.
2001 Regional Variability in Clovis, Folsom, and Cody Land Use, *Plains Anthropologist*, *46*, pp. 65-94.

BRENNAN, LOUIS
1982 A Compilation of Fluted Points of Eastern North America by Count and Distribution: An AENA Project, *Archaeology of Eastern North America*, *10*, pp. 27-46.

DILLEHAY, TOM D. and DAVID J. MELTZER (eds.)
1991 *The First Americans: Search and Research*, CRC Press, Boca Raton.

DUNNELL, ROBERT C. and WILLIAM S. DANCEY
1983 The Siteless Survey: A Regional Scale Data Collection Strategy, *Advances in Archaeological Method and Theory*, *6*, pp. 267-287.

FERRING, C. REID
1994 The Role of Geoarchaeology in Paleoindian Research. In *Method and Theory for Investigating the Peopling of the Americas*, R. Bonnichsen and D. G. Steele (eds.), Center for the Study of First Americans, Corvallis, Oregon, pp. 57-72.

FIEDEL, STUART J.
2000 The Peopling of the New World: Present Evidence, New Theories, and Future Directions, *Journal of Archaeological Research*, *8*:1, pp. 39-103.

FUTATO, EUGENE M.
1982 Comments on the *AENA* Fluted Point Survey, *Archaeology of Eastern North America*, *11*, p. 20.

GOODYEAR, ALBERT C.
1999 The Early Holocene Occupation of the Southeastern United States: A Geoarchaeological Summary. In *Ice Age Peoples of North America. Environments, Origins, and Adaptations of the First Americans*, R. Bonnichsen and K. L. Turnmire (eds.), Oregon State University Press for the Center for the Study of the First Americans, Corvallis, Oregon, pp. 432-481.

HOMER, C. G., C. HUANG, L. YANG, and B. WYLIE
2002 *Development of a Circa 2000 Landcover Database for the United States*, ASPRS Proceedings, Washington, D.C.

HUANG, C., B. WYLIE, C. HOMER, L. YANG, and G. ZYLSTRA
2002 Derivation of the Tasseled Cap Transformations based on Landsat 7 At-Satellite Reflectance, *International Journal of Remote Sensing*, *23*:8, pp. 1741-1748.

JABLONSKI, NINA G. (ed.)
2002 *The First Americans: The Pleistocene Colonization of the New World*, Memoirs of the California Academy of Sciences Number 27.

LARGENT, F. B., JR., M. R. WATERS, and D. L. CARLSON
1991 The Spatiotemporal Distribution and Characteristics of Folsom Projectile Points in Texas, *Plains Anthropologist*, *36*:137, pp. 323-341.

LEPPER, BRADLEY T.
1983 Fluted Point Distributional Patterns in the Eastern United States: A Contemporary Phenomenon, *Midcontinental Journal of Archaeology*, *8*, pp. 269-285.
1985 The Effects of Cultivation and Collecting on Ohio Fluted Point Finds: A Reply to Seeman and Prufer, *Midcontinental Journal of Archaeology*, *10*, pp. 241-250.

LEPPER, BRADLEY T. and DAVID J. MELTZER
1991 Late Pleistocene Human Occupation of the Eastern United States. In *Clovis Origins and Adaptations*, R. Bonnichsen and K. L. Turnmire (eds.), Oregon State University Press for the Center for the Study of the First Americans, Corvallis, Oregon, pp. 175-184.

MARTIN, PAUL S.
1973 The Discovery of America, *Science*, *179*, pp. 969-974.
1984 Prehistoric Extinctions: The Global Model. In *Quaternary Extinctions: A Prehistoric Revolution*, P. S. Martin and R. G. Klein (eds.), University of Arizona Press, Tucson, pp. 354-403.

MANLY, BRYAN F. J.
1994 *Multivariate Statistical Methods: A Primer*, 2nd Edition, Chapman and Hall, New York.

MASON, RONALD J.
1962 The Paleo-Indian Tradition in Eastern North America, *Current Anthropology*, *3*, pp. 227-278.

MCCARY, BEN C.
1984 *Survey of Virginia Fluted Points*, Archaeological Survey of Virginia, Special Publication, p. 12.
MELTZER, DAVID J.
1995 Clocking the First Americans, *Annual Review of Anthropology*, *24*, pp. 21-45.
1997 Monte Verde and the Pleistocene Peopling of the Americas, *Science*, *276*, pp. 754-755.
2002 What Do You Do When No One's Been There Before? Thoughts on the Exploration and Colonization of New Lands. In *The First Americans, The Pleistocene Colonization of the New World*, N. G. Jablonski (ed.), Memoirs of the California Academy of Sciences Number 27, San Francisco, California, pp. 27-58.
MELTZER, DAVID J. and MICHAEL R. BEVER
1995 Paleoindians of Texas: An Update on the Texas Clovis Fluted Point Survey, *Bulletin of the Texas Archeological Society*, *66*, pp. 7-81.
MELTZER, D. J., D. K. GRAYSON, G. ARDILA, A. W. BARKER, D. F. DINCAUZE, C. V. HAYNES, F. MENA, L. NUNEZ, and D. J. STANFORD
1997 On the Pleistocene Antiquity of Monte Verde, Southern Chile, *American Antiquity*, *62*:4, pp. 659-663.
MOSIMANN, J. E. and P. S. MARTIN
1975 Simulating Overkill by Paleoindians, *American Scientist*, *63*, pp. 304-313.
PURDY, BARBARA A.
1982 Comments on "A Compilation of Fluted Points of Eastern North America by Count and Distribution: An *AENA* Project," *Archaeology of Eastern North America*, *11*, p. 29.
SEEBACH, JOHN D.
2000 *Drought or Development? Patterns of Paleoindian Site Discovery in the Western United States*. Paper presented at the 65th Annual meeting of the Society for American Archaeology, Philadelphia.
SEEMAN, MARK F. and OLAF H. PRUDER
1982 An Updated Discussion of Ohio Fluted Points, *Midcontinental Journal of Archaeology*, *9*, pp. 227-233.
1984 The Effects of Cultivation and Collecting on Ohio Fluted Point Finds: A Cautionary Note, *Midcontinental Journal of Archaeology*, *9*, pp. 227-233.
SHOTT, MICHAEL J.
2002 Sample Bias in the Distribution and Abundance of Midwestern Fluted Bifaces, *Midcontinental Journal of Archaeology*, *27*:1, pp. 89-123.
SOKAL, ROBERT R. and F. JAMES ROHLF
1997 *Biometry, The Principles and Practice of Statistics in Biological Research*, 3rd Edition, W. H. Freeman and Company, New York.
STORCK, PETER
1982 Commentary on *AENA*'s Compilation of Fluted Points in Eastern North America, *Archaeology of Eastern North America*, *11*, pp. 34-36.
U.S. CENSUS BUREAU
2002 *Table ST-EST2002-01*, State Population Estimates, Population Division, U.S. Census Bureau (on Internet: http://eire.census.gov/popest/estimates/php).

VOGELMANN, J. E., S. M. HOWARD, L. YANG, C. R. LARSON, B. K. WYLIE, AND N. VAN DRIEL
2001 Completion of the 1990s National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources, *Photogrammetric Engineering and Remote Sensing*, *67*, pp. 650-652.

Direct reprint requests to:

Briggs Buchanan
102-6385 Hawthorn Ln.
Vancouver, BC, Canada
V6T 1Z4
e-mail: briggs@unm.edu