When Hunting Meets Herding:

Multivariate Analysis of Cross-Cultural Datasets

Isaac I. Ullah[1]

[1] **Arizona State University, School of Human Evolution and Social Change**

DRAFT

DO NOT CITE IN ANY CONTEXT WITHOUT THE EXPRESS PERMISSION OF THE AUTHOR

Abstract

Keywords

Introduction

Pastoralism, as a productive subsistence strategy, originated in the Near East (goats and sheep) and North Africa (cattle) in the Neolithic, and subsequently spread from those areas to most of the Old World over the course of the late Holocene. In areas where farming preceded herding, domestic herd animals were incorporated into an established economic system that was based on delayed returns, food production, and that required sedentary habitation for at least part of the year. In other areas a hunting and gathering lifeway remained predominate, and pastoralism would have had to have been interfaced with an economic system based on the extraction of natural resources, instant returns, and that was inherently mobile. The adoption of pastoralism by hunter gatherers has been most widely discussed by archaeologists and anthropologists who work in southern Africa, but largely ignored elsewhere (with the exception, perhaps, of the adoption of Horses in the New World).

The debate in southern Africa has mainly centered around the particulars of specific cases using data particular to southern Africa ("The Bushman Debate", cf. [Kent 2002; Sadr 1997]), however, and few real archaeological models of hunter-herder interaction or the transition from hunting to herding have been produced. Southern Africa never-the-less provides one of the best opportunities to study herder-hunter interactions and the transition from hunting to herding. This is because these processes were active and widespread during the historic period, and to greater or lesser extent, are still ongoing at the present time. In order to move beyond debates about particulars, the pertinent questions that need to be asked are basic: Can we actually distinguish pastoralists, agriculturalists, and hunter-gatherers from the kinds of data likely to be recovered from archaeological survey and excavation? If so, can we then also distinguish transitional hunter/herders from regular hunters or regular herders? What types of variables are the best for making these determinations? In this paper, I use multivariate statistical analysis of a cross-cultural dataset of subsistence and mobility patterns derived for a purpositive sample of societies from the Standard Cross Cultural Sample (SCCS, [Murdock and White 2006]) to attempt to answer these questions. I then evaluate the solutions derived from the SCCS data with a separate smaller ethnographic dataset—newly coded by the author from primary ethnographic sources—that focuses on specific historic and modern southern African societies that at the time of ethnographic observation were in various stages of transition between a fully hunting and gathering lifestyle and a fully pastoral lifestyle.

Data and Sampling

The use of ethnographic data—especially from modern and historical hunter-gatherer groups—to create ethnoarchaeological models for interpreting archaeological patterns has become mainstream since the advent of the technique during the rise of "the new archaeology" in the 1960's and '70's (Yellen 1977). Arguably, "classic" ethnoarchaeology usually involves field work with a specific extant group (hunter-gatherers, pastoralists, small scale agriculturalists, etc.), and subsequently transforming data collected during the field work into a model designed to address a specific archaeological problem such as site formation, settlement patterns, or lithic usewear. Cross-cultural data, on the other hand, have been somewhat less used in ethnoarchaeological applications; for the most part they have been used to derive general models (Binford 1980) or simply just to search for patterns (Feinman and Neitzel 1984) rather than applied to specific archaeological problems. The SCCS is a unique database with over 2000 variables coded for 186 societies, and therefore is an ideal source for cross-cultural analysis of specific issues. The SCCS data are not without limitations; because the data for each group included in the database were collected by many different people at many different times, the creators of the SCCS had to code some variables at a fairly coarse resolution, which is a fact that must be accounted for during analysis.

For this study, the first issue was determining a sample of SCCS societies and variables that were appropriate for the questions at hand. Because subsistence strategy is the dependent variable in this study, I chose the presence of pastoralism (and score above 0 in SCCS variable 5) as my first culling factor. From the societies that remained, I then removed those that did not hunt (any score above 0 in SCCS variable 9). I then removed groups that secondary research show to be engaged in complex or state level systems and those based off of ancient historic accounts (eg. "Hebrews", "Romans"), and groups that relied highly on fishing as a main subsistence source (eg. Haida, Tlingit). In order to provide a spread of subsistence variability that is more appropriate for the distinguishing hunter-herders from pure hunters and pure pastoralists, I *post-facto* included two societies that would have otherwise been culled for being strict hunter gatherers that did not practice pastoralism (the Hadza and the pre-1970's !Kung), and one society that would have been culled for being strict pastoralists that did not hunt (the Masai). The name and SCCS reference number of the 60 SCCS societies that remained in my sample after the culling process are listed in Table 3.

I then chose a sample of appropriate SCCS variables that would allow me to explore patterns in the data that would be relevant to the issue of hunter-herder interaction/transition. Because this paper focuses on the *archaeological* visibility of this phenomenon, I had to ignore a large amount of SCCS variables that may be relevant to the issue but that are not likely to leave any archaeological traces. Of the remaining SCCS variables that were arguably archaeologically visible, I chose to focus on those SCCS variables related to subsistence strategy and settlement patterns as these are data that are familiar to and commonly recorded by archaeologists. Also, because of the afore-mentioned coarseness of some of the SCCS data, variables used in the analysis were coded differently. Some were fine-grained enough so that they could be used as actual percentages, while others were so coarse that they had to be "binned" into ordinal categories. Still other data were best represented as binary "presence/absence" categories. Table 1 lists the SCCS variables I chose for the multivariate analyses and how these data were recoded or manipulated to facilitate analysis.

The secondary "test" dataset was recorded from original ethnographic and historic sources, and focused on several southern African societies that were transitioning between hunting and herding during the time they were studied. In order to retain comparability with the results derived from the SCCS data, it was necessary for me to code them using the methods outlined in Table 1 In many cases, this meant I had to purposefully "bin" fine grained data into coarse ordinal categories (using the binning procedures outlined in Table 1). In other cases, where precise measurements of variables were unavailable, I had to estimate approximate rates or amounts from descriptive passages. Table 2 lists the groups included in this secondary dataset and the sources used to code the data.

Methods

All multivariate analysis was done in PAST, a robust and free statistical program developed for use with geological, palaeontological, and archaeological data (Hammer, et al. 2001). In order to better understand the nature of the data, the relative impact of the various input variables, and the quality of the data for making correlations between societies, initial exploration was conducted with k-means cluster analysis and hierarchical clustering. K-means clustering is an *n*-dimensional, non-hierarchical clustering technique that fits a multivariate dataset to a user defined number of clusters (cluster configuration level) as well as a variety of descriptive statistics that describe those cluster configurations. These statistics allow the analyst to make heuristic judgments as to the best scale of clustering and gives them an indication how well clustered the data is over all (Kintigh and Ammerman 1982). In addition, because k-means is a "supervised" clustering algorithm, various combinations of variables from the dataset can be input into the clustering routine and processed at various cluster configuration levels in order to determine an optimal set of variables for input into other types of multivariate analysis. Because of the randomness inherent to k-means analysis, slightly different cluster configurations can be achieve through multiple runs using identical data. To get around this, I performed a kind of "bootstrap" routine in which I normalized the output using the results of ten runs. All results of k-means clustering algorithms were then compared against SCCS variable 246 (subsistence economy) both to check the validity of the cluster configuration, and to see if the various combinations of input data variables supported the description in SCCS variable 246.

Hierarchical clustering is an unsupervised clustering routine explores neighbor linkages between types of multidimensional data based on a distance measure, and produces a dendrogram (similar to a cladistic tree) of associations where the length of branches is proportional to the amount of relatedness between types. Branches that are separated by a large number of forks and/or long sections are less related according to the distance measure used than branches that are close together. Therefore, unlike k-means, hierarchical clustering can only result in one clustering configuration per given distance measure. This dendrogram provides an excellent visual heuristic that is helpful for determining the degree of relatedness and the strength of correlation between types in the input data. In this analysis I use a distance measure based on Pearsons *r* that derives branch lengths based on the degree of correlation between variables in a multivariate linear regression, and which is appropriate for using on datasets of mixed frequency, ordinal, and binary data types. The results of hierarchical clustering were then compared to those returned by the k-means algorithm for additional confirmation of cluster types.

After the initial data exploration, the chosen variables for all the included societies were subjected to Correspondence Analysis (CA). CA is a robust and appropriate statistical method that, similar to Principle Components Analysis (PCA), is used to analyze variability in multivariate datasets, but that is specifically intended for abundance, ordinal and contingency (binary) data (ter Braak 1985). Much like PCA, CA produces hypothetical values (correspondence axes) that account for as much of the variability in the input dataset as possible. An eigenvalue matrix that explains the importance of each of these axes, and input types (input row headers. in this case the names of input societies) can be plotted in the resulting coordinate space created by any combination of correspondence axes. Additionally, the “weight” of variable types (input column headers. in this case, the names of measured mobility and subsistence variables) can also be plotted in the same coordinate space). Clusters of types will plot near the location the variables that are most responsible for explaining the variability within those types. In other words, the CA plot allows the analysts to better understand *why* some input types are more related to each other than they are to other input types.

Finally, I use ternary plots to gain a better understanding of the patterning within some of the more important sets of variables. Ternary plots are three axis graphs that commonly used in geology to plot the percentages of sand, silt, and clay present in soils, and are a more easily understood method of visualizing three dimensions of variability than pseudo-3D XYZ plots. Additionally, PAST allows for data points to be colored by a fourth variable (in fact, this is so for all graphical output in PAST, as will be seen below), so that one can explore *four* dimensions of data on one ternary plot.

Initial Results

Various combinations of SCCS variables were analyzed with k-means clustering. For the most part, these analyses resulted in two possible “natural” cluster configurations at 3 and 6 clusters respectively. Most combinations of variables successfully classified input societies according to SCCS variable 246 (subsistence economy) at the 3-cluster level. That the data cluster nicely at the 3 cluster configuration level is not surprising, given that most of the input societies really do focus more on one of the three major subsistence economies (hunting and gathering, pastoralism, and agriculture). Also, the strong correlation between the 3-cluster results and SCCS variable 246 indicates that the SCCS data is probably accurate and detailed enough for more rigorous multivariate analysis.

The 3-cluster results, however, are uninformative for differentiating economies that span two of the major subsistence types, so it is necessary to examine the 6 cluster configurations. Similar successful 6-cluster solutions were derived for runs that included all the variables and for runs that included just the 6 major subsistence variables (HGF, HERD, AG, hgf, herd, and ag in Table 1), but not for other combinations of input variables. This means that it is likely that the 6 main subsistence variables will weigh heavily in other multivariate analyses, but that the data do allow for finer distinguishing of economic base than simply "forager", "herder", or "agriculturalist".

Table 3 presents the 6-cluster results, and provides a brief explanation of the nature of each cluster. A non-statistical investigation of the similarities in subsistence activities of constituents of each cluster is enlightening. Cluster 1 groups societies that seem to spend most of their subsistence time engaged in hunting and gathering, and who also rely on hunted and gathered foods to a high degree. As described above, only two of these societies do not practice pastoralism at all, although most of the remaining groups are "horse" hunters who used their herd animals to aid in hunting activities. Cluster 2 groups societies that could be considered full time pastoralists who hunt wild game on occasion, but who seem to be most heavily dependant upon pastoral products. Cluster 3 defines groups who are almost pure agriculturalists. They may practice a small amount of animal husbandry, and they may also hunt or gather wild resources on occasion, but they do not rely on these activities to provide dietary staples. Cluster 4 is the largest cluster of societies, and includes groups that would fit the classic definition of "agropastoralists". They rely slightly more on farmed goods than on pastoral products, and do not rely on foraging for a large proportion of their diets.

The next two clusters are the most important, given the topic of this paper. Cluster 5 consists of societies that rely on a mix of small-scale agriculture or horticulture and hunting and gathering. These groups all rely on foraged resources for at least 40% or more of their diets, and only one group—the Huichol—relies on pastoral products to any significant degree (10%). Cluster 6 is perhaps even more important, as it groups societies loosely aligned with a pastoral way of life, but that only rely moderately on pastoral products in their diets. Of the three groups included in cluster 6, the Abkhaz rely the most heavily on a mixture of pastoral and agricultural products and do not rely much on foraged resources. The Tuareg also mainly rely on a mixture of pastoralism and agriculture, but also rely on foraged resources for about 20% of their diet. The Nama (Hottentot) are the most interesting case, as they rely on only pastoralism and hunting and gathering (50% each) and not at all on agriculture.

The results of hierarchical clustering are shown in Figure 1. SCCS societies are color coded by cluster for the 6-cluster the k-means configuration, which allows for the direct comparison between the result of the two clustering routines. In addition, societies from the "test" group are coded in magenta, and will be discussed below. The first four k-means clusters do indeed plot close to one another on the hierarchical dendrogram, but clusters 5 and 6 do not. This is interesting because clusters 5 and 6 are made up of societies that bridge agriculture and pastoralism with hunting and gathering, respectively. Furthermore, the k-means cluster of societies that practice both agriculture and pastoralism (cluster 4) *are* predicted by the unsupervised hierarchical clustering routine. This indicates that k-means clusters 5 and 6 are unique not in that they bridge two subsistence strategies, but that they bridge a *productive* system with an *extractive* system. Additionally, while these these types of societies can be detected with *n*-dimensional clustering, they are difficult to detect with simple neighbor linking algorithms.

CA provides a more detailed view of exactly how and why these clusters are distinguished. Figure 2 shows the SCCS societies plotted in the coordinate space defined by the first two CA axes, colored by cluster (see Table 3 for color codes) and in relation to "ideal" position of each input variable. Together, the first two CA axes account for nearly 80% of the total variability in the dataset, so spatial clustering of societies in this space indicate high degrees of similarity in terms of the variables that also plot in the same vicinity. For example, societies determined to be mostly agricultural or agropastoral by k-means analysis all plot in the lower right quadrant of the CA space. The types of input variables that also plot in this area of the Ca space indicate that these societies are related because they spend a large portion of the year sedentary, rely on agriculture for a large amount of their subsistence base, have high population densities, and have large communities. The few agropastoral societies that are more highly engaged in herding activities plot in the upper center of the CA space, a location that is influenced by a high degree of residential mobility. This indicates that these societies may be better classified as "transhumant" agropastoralists, and perhaps should be included in a separate group than more sedentary agropastoralists. Interestingly, all the agropastoralists plot near the nexus for ovicaprid pastoralism.

All the "horse" hunting societies plot in the lower left part of the CA space, near the variables that indicate a high reliance on equids, a high degree of residential mobility, and low population densities. The two non-horse hunters (Hadza, and pre-70's Kung) plot in the area of the graph that is between fully logistically mobile and residentially mobile. Both types of hunters plot on the side of the graph that is also dominated by a high reliance on hunting, a high time investment in hunting, and small community size.

The pastoral societies plot in two areas. Some plot very near the "horse" hunting societies, indicating that they have a high degree on equids or camelids, and that they are also highly residentially mobile. Other pastoral groups plot near to the portion of the CA plot that is highly influenced by the amount logistic mobility, and closer to the nexuses for bovid and ovicaprid pastoralism. Both types of pastoralists plot near variable that indicates a high reliance on pastoral products in the diet.

The groups included in k-means cluster 5 (hunter/agriculturalists) plot near the center of the CA space in area dominated neither by sendentism and agriculture, nor by residential mobility and hunting. This makes sense because these groups spend half the year living the first way, and half the year living the second way. Interestingly, groups included in k-means cluster 6 are widely spread in the CA space. The Nama plot almost directly between the "horse" hunters and the highly mobile pastoral groups. The Tuareg plot near the nexus for logistic mobility, near the transhumant agropastoral groups, but closer to the bovid nexus than the ovicaprid nexus. Finally, the Abkhaz plot very near to the sedentary agropastoralists, indicating that this group may have been misclassified by the cluster routine.

Because the CA plot seems to be mainly dominated by patterns of mobility and demographics, it does not allow us to understand the finer patterning of subsistence variables in our dataset. Figures 4 compares the relative percentages of foraged, pastoral, and agricultural items in the diets of the SCCS societies (4a, 4c), as well as the relative percentages of subsistence time spent engaged in foraging, pastoralism, and agriculture (4b, 4d). In addition, we can compare these patterns with the types of animals that are herded (4a, 4b) and the amount of food in the diet obtained through trade (4c, 4d). The following patterns are apparent: Both "Horse" hunters and "non-horse" hunters seem to spend most of their time foraging, and also rely highly on foraged foods. Camel pastoralists spend most of their time engaged in pastoral production, and also rely on a high amount of pastoral foods in their diets. People who herd Ovicaprid and Bovid seem to split most of their time between herding and agriculture, but vary widely in the degree to which they rely on the various food sources. Interestingly, herders that also rely to a high degree on pastoral products in their diets also rely to high degree on imported food items. Hunter-gatherers and agriculturalists rely on imported food items to a lesser degree.

Model Testing and Discussion

The results described above indicate there is definite patterning in the subsistence and mobility data from the SCCS societies. They also show that the SCCS societies can, for the most part, be classified correctly through multivariate analysis. But what to they add to our ability to identify transitional herder-hunters? To this end, it is useful to apply these techniques to our case study in southern Africa. The seven societies included in this test dataset were at varying stages of transition between hunting and herding. Of the seven, the post 1970's !Kung and Nharo were at the very early stages of incorporating pastoralism and agriculture into what was still a dominantly foraging lifeway (Guenther 1986; Yellen 1985). On the other extreme, the Bakwena in the early 20th century, being forced for political reasons to more arid regions in the southern Khalahari, were beginning to incorporate foraging into their mainly pastoral lifestyle (Okihiro 2000). The Babolaonge and the Bakgwatheng were Khalahari groups of mixed "Bushman" and Bantu ancestry that practiced systems of mixed pastoralism, foraging, and agriculture. These groups were displaced by the immigration of the Bakwena, and were semi-subjugated by them. As such were beginning to rely more foraging than they had previously, although the Bakgwathenge relied more on small garden plots than did the Babolaonge (Okihiro 2000). The Damara are a relatively numerous group of hunter-gatherers living in the Namib desert who seem to have had high levels of interaction with neighboring pastoralists. In fact, they speak the Nama language, which is a non-Kung Khoisan language spoken by the pastoral Nama peoples. At the time of ethnographic observation, the Damara were practicing a very laissez-faire fashion of goat herding whereby flocks would be allowed to wander freely in the bush, and milking was sporadic if practiced at all (Barnard 1992). The Cape KhoiKhoi are arguably descended from the original wave of pastoral people that entered into South Africa around 2000 YBP (Barnard 1992). The KhoiKhoi were mainly pastoralists, but still relied to fairly high degree on hunted and gathered foods.

In order to compare these societies with the SCCS results, it is necessary to add them to the SCCS dataset, and re-run all the multivariate analyses with the combined dataset. These societies were then labeled in output charts so that they could be assessed according to the dimensional framework defined by the variability in the SCCS data. In hindsight, it is apparent that addition of the seven test societies skews the results of the multivariate tests. It would have been more accurate to add each of the seven societies to the total dataset individually, and re-run the multivariate analyses in a "jackknife" type of rotation. Never-the-less, the results as shown below are interesting.

K-means analysis with the extended dataset resulted in a slightly different 6-cluster configuration than with just the SCCS data alone. As such, the k-means results for the expanded dataset are not directly comparable with the original results. Hierarchical clustering on the expanded dataset was almost completely identical, however, and because the SCCS data are color coded by their original k-means cluster, we can compare how the test societies fit into the various cluster configurations on the dendrogram (Figure 1). Perhaps unsurprisingly, the Bakwena and the Bakgwatheng plot with the main cluster of agropastoralists, while the KhoiKhoi plot with the main group of pastoralists. Interestingly, the Damara and the Babolaonge plot with the Tuareg who are pastoralists that also hunt. The modern Nharo and !Kung have their own branch in the dendrogram, showing that they do not closely correlate with any of the clusters of SCCS groups. This is especially telling, because the pre-1970's !Kung are part of the original SCCS input data, but they plot very far away on the dendrogram from the post-1970's !Kung.

The CA results are even more illustrative, and uphold the differentiation among the test societies described by the hierarchical cluster results (Figures 2 and 3). The KhoiKhoi plot on the middle left of the coordinate space, indicating that they most closely resemble the logistically mobile pastoral groups. The Bakwena and the Bakgwatheng plot squarely in the middle right of the coordinate space, indicating that their economy is more like a form of semi-sedentary agropastoralism. The Damara and the Babolaonge plot very close to the Tuareg high in the center of coordinate space, indicating that they are logistically mobile herder-hunters. Finally, the post 1970's !Kung and Nharo plot in the low center of the CA space, showing that they have a subsistence and mobility profile that is unique amongst the spread of other societies. This location reflects the fact that they split their time between being sedentary at water holes and residentially mobile in the bush, and that while foraging is most important to their diet and takes up most of their subsistence time, herding is also important.

Placing the test groups on the subsistence plots in figure 4 provides more detail about how they differ. The post 1970's Nharo and !Kung plot high in their reliance on hunted and gathered food items in their diets, but split their subsistence time almost equally between the three activities. The KhoiKhoi and the Damara, followed closely by the !Kung and Nharo, are the least reliant on agricultural items in their diets, but all four societies are highly reliant on imported food items. The two test societies that herd cattle—the Bakwena and the KhoiKhoi—are dissimilar in their dietary profiles, but are the two societies among those of the test set that spend the largest amount of time engaging in pastoral production.

Conclusion

Multivariate analysis of cross-cultural data seems to be a promising and appropriate technique to address the questions posed at the beginning of this paper. Given the coarseness inherent to the SCCS data, the methods used in this paper were surprisingly accurate for distinguishing societies according subsistence strategy, and combining and comparing the results of multiple multivariate analyses provides more robust results. More importantly, the results show that the kinds of data that are wide available to archaeologists—mainly data types that related to subsistence, demography, and mobility—are probably accurate enough to successfully situate archaeological cultures in the context of ethnographically derived data.

Although the analyses did distinguish SCCS groups that spanned more than one subsistence niche, these results were somewhat less convincing that for those SCCS groups engaged overwhelmingly in one lifeway. Never-the-less, the classification of the test societies was fairly successful, especially with regards to the post 1970's !Kung and Nharo. These two groups consistently plotted as significantly different from all other groups analyzed here, but in a way that made sense in relation to the variables being measured. In this sense, the methods explored in this paper are extremely successful for distinguishing societies at the very initial stages of transitioning from hunting to herding.

With some modification (e.g., utilization of finer grained input data, "jackknife" rotation of test cases through the multivariate analyses), the techniques presented here should provide a valuable tool for the archaeological study of the spread of pastoralism into areas occupied by hunter-gatherers. Future analysis of archaeological data should further prove the validity of this approach. Hopefully, the investigation of complex archaeological problems through verifiable models built from multivariate ethnographic data can help archaeologists move beyond tedious, sometimes nasty, and ultimately useless debates about particulars, and into more interesting and fruitful avenues of research.

Tables

Variable	SCCS variable #	Variable Graphing Code	Bin Amount	Ordinal Code	Bin Explanation	Transformation to Binary Codes During Analysis?
Average amount of food production/extraction time spent engaged in Hunting, Gathering, and Fishing	Hunting: 9 Gathering: 11 Fishing: 7	HGF,	0 < 10% 10-50% > 50%	0 1 2 3	None/Non-Food Low Moderate High	N N N N
Average amount of food production/extraction time spent engaged in Pastoralism	5	HERD,	0 < 10% 10-50% > 50%	0 1 2 3	None/Non-Food Low Moderate High	N N N N
Average amount of food production/extraction time spent engaged in Agriculture	3	AG	0 < 10% 10-50% > 50%	0 1 2 3	None/Non-Food Low Moderate High	N N N N
% of diet from Hunted, Gathered, and Fished foods	Hunting: 203 Gathering: 204 Fishing: 205	hgf	Not binned	1-100	Percentage	N
% of diet from Pastoral foods	206	herd	Not binned	1-100	Percentage	N
% of diet from Agricultural foods	207	ag	Not binned	1-100	Percentage	N
Amount of food obtained by Trade	2	trade	0 < 10% 10-50% > 50%	0 1 2 3	None/Non-Food Low Moderate High	N N N N
Main herd animal species (more than one possible)	6	ovicap bov eq/cam	Collected as Presence / Absence	0/1 0/1 0/1	Ovicaprids Bovids Equids or Camelids	Y Y Y
Size of main prey animals	10	largeanim smallanim	Collected as Presence / Absence	0/1 0/1	Large Animals Small Animals	Y Y
% of year Residentially Mobile	66, 234	%resmob	Not binned	1-100	Percentage	N
% of year Logistically Mobile	66, 234	%logmob	Not binned	1-100	Percentage	N
% of year Sedentary	66, 234	%sed	Not binned	1-100	Percentage	N
Population Density	64	sparse msparse mdense dense	< 1 person/mi 1-5 person/mi 5-25 person/mi >25 person/mi	1 2 3 4	sparse moderately sparse moderately dense dense	Y Y Y Y
Community Size	63, 235	smallcom medcom largecom	< 50 persons 50-99 persons > 100 persons	1 2 3	Small Medium Large	Y Y Y

Table 1: Variables included for analysis and explanation of how they were coded and any transformations that were done during analysis

Society Name	**Reason for inclusion**	**Sources**
Post 70's Kung	A “Bushman” group at the very early stages of goat herding and a very little agriculture, but still mainly hunter gatherers	(Biesele, et al. 1989; Solway and Lee 1990; Yellen 1985)
Post 70's Nharo	A “Bushman” group at the very early stages of goat herding and a very little agriculture, but still mainly hunter gatherers	(Barnard 1986, 1992; Barnard and Widlok 1996; Biesele, et al. 1989; Guenther 1986)
Bakwena	A non-“Bushman” cattle pastoral group who was forced into the Kalahari, and began to rely more on hunting and gathering of wild resources.	(Okihiro 2000)
Bakgwatheng	A goat herding and small scale farming group of mixed “Bushman” decent in the Kalahari that also relied to a high degree on hunting and gathering wild resources	(Okihiro 2000)
Babolaonge/ Bashaga	A Kalahari “Bushman” group that mainly relied on hunting and gathering, but that also practiced goat pastoralsim	(Okihiro 2000)
Damara	A group of “Bushmen” hunter gathers in the Namib that also practiced a small amount of goat herding and were “patrons” of the Nama, a cattle pastoralist group.	(Barnard 1992)
Cape KhoiKhoi/ Korana	A non-“Bushman” “click” speaking group that were mainly pastoralists at first contact, but who also hunted and gathered wild resources to a large extent.	(Barnard 1992; Denbow and Wilmsen 1986; Guelke and Shell 1992; Ross 1980; Smith 1983; Smith 1992; Stapleton 1991)

Table 2: Groups included in the secondary “test” dataset, the reason for including them in the dataset, and the main sources of ethnographic data used to code the dataset.

SCCS #	Society Name	Main Subsistence (sccs V246)	Interpretation of Cluster	K-means 6 cluster result	Correspondence Analysis Chart Symbol
2	Kung Bushmen	Gathering contributes most	These groups are mainly hunter-gatherers who either use animals (eg. horses) to help them hunt, practise only a very small amount of pastoralism, or do not practise pastoralism at all. They also do not rely on nor practice much agriculture.	1	Open Trinagle
9	Hadza	Gathering contributes most		1	
137	Paiute (North.)	Gathering contributes most		1	
140	Gros Ventre	Hunting contributes most		1	
147	Comanche	Hunting contributes most		1	
148	Chiricahua	Gathering contributes most		1	
182	Lengua	Hunting contributes most		1	
183	Abipon	Hunting contributes most		1	
185	Tehuelche	Hunting contributes most		1	
25	Pastoral Fulani	Pastoral contributes most	These groups are pastoralists who spend most of their time engaged in pastoral activities, and also rely to a high degree on pastoral products They may hunt, but don't rely on hunting to a large degree. They rely to varying degrees on agricultural products in their diets.	2	Closed circle
34	Masai	Pastoral contributes most		2	
36	Somali	Pastoral contributes most		2	
38	Bogo	Pastoral contributes most		2	
46	Rwala Bedouin	Pastoral contributes most		2	
58	Basseri	Pastoral contributes most		2	
65	Kazak	Pastoral contributes most		2	
66	Khalka Mongols	Pastoral contributes most		2	
159	Goajiro	Pastoral contributes most		2	
10	Luguru	Extensive agriculture contributes most	These groups are small scale agriculturalists who may also practice hunting, gathering, and/or pastoralism to some degree, but who rely mainly on agricultural products for their basic subsistence.	3	Open Circle
12	Ganda	Intensive agriculture contributes most		3	
19	Ashanti	Extensive agriculture contributes most		3	
35	Konso	Intensive agriculture contributes most		3	
62	Santal	Intensive agriculture contributes most		3	
70	Lakher	Extensive agriculture contributes most		3	
149	Zuni	Intensive agriculture contributes most		3	
3	Thonga	Extensive agriculture contributes most	These are groups are mainly small scale agriculturalists, but who also rely to large degree on pastoral products in their diets, and who mainly split their subsistence activity time between these two activities. They may also rely to varying degrees on hunted and gathered resources, but they do not spend much time on those activities.	4	Open Square
4	Lozi	Intensive agriculture contributes most		4	
5	Mbundu	Extensive agriculture contributes most		4	
8	Nyakyusa	Intensive agriculture contributes most		4	
11	Kikuyu	Intensive agriculture contributes most		4	
21	Wolof	Extensive agriculture contributes most		4	
22	Bambara	Intensive agriculture contributes most		4	
24	Songhai	Intensive agriculture contributes most		4	
26	Hausa	Intensive agriculture contributes most		4	
27	Massa (Masa)	Intensive agriculture contributes most		4	
29	Fur (Darfur)	Intensive agriculture contributes most		4	
30	Otoro Nuba	Intensive agriculture contributes most		4	
31	Shilluk	Extensive agriculture contributes most		4	
33	Kaffa (Kafa)	Intensive agriculture contributes most		4	
37	Amhara	Intensive agriculture contributes most		4	
39	Kenuzi Nubians	Intensive agriculture contributes most		4	
40	Teda	Intensive agriculture contributes most		4	
42	Riffians	Intensive agriculture contributes most		4	
48	Gheg Albanians	Intensive agriculture contributes most		4	
57	Kurd	Intensive agriculture contributes most		4	
60	Gond	Extensive agriculture contributes most		4	
67	Lolo	Intensive agriculture contributes most		4	
68	Lepcha	Intensive agriculture contributes most		4	
72	Lamet	Extensive agriculture contributes most		4	
141	Hidatsa	Intensive agriculture contributes most		4	
172	Aymara	Extensive agriculture contributes most		4	
184	Mapuche	Intensive agriculture contributes most		4	
142	Pawnee	Extensive agriculture contributes most	These groups are small scale agriculturalists who also seem to depend on hunting and gathering for a large portion of their diets. They do not spend much time on pastoral activites.	5	X Sign
143	Omaha	Two or more sources contribute equally		5	
150	Havasupai	Intensive agriculture contributes most		5	
151	Papago	Intensive agriculture contributes most		5	
152	Huichol	Extensive agriculture contributes most		5	
1	Nama Hottentot	Pastoral contributes most	These groups are pastoralists who only rely on pastoral products to a moderate degree, regardless of how much time they spend on pastoral activities.	6	Plus sign
41	Tuareg	Pastoral contributes most		6	
55	Abkhaz	Pastoral contributes most		6	

Table 3: List of SCCS societies used in the analysis organized by k-means 6-cluster results. The table also shows the colors and symbols used to differentiate these clusters on the hierarchical cluster dendrogram and on the CA plot in Figures 1 and 2 respectively.

Figures

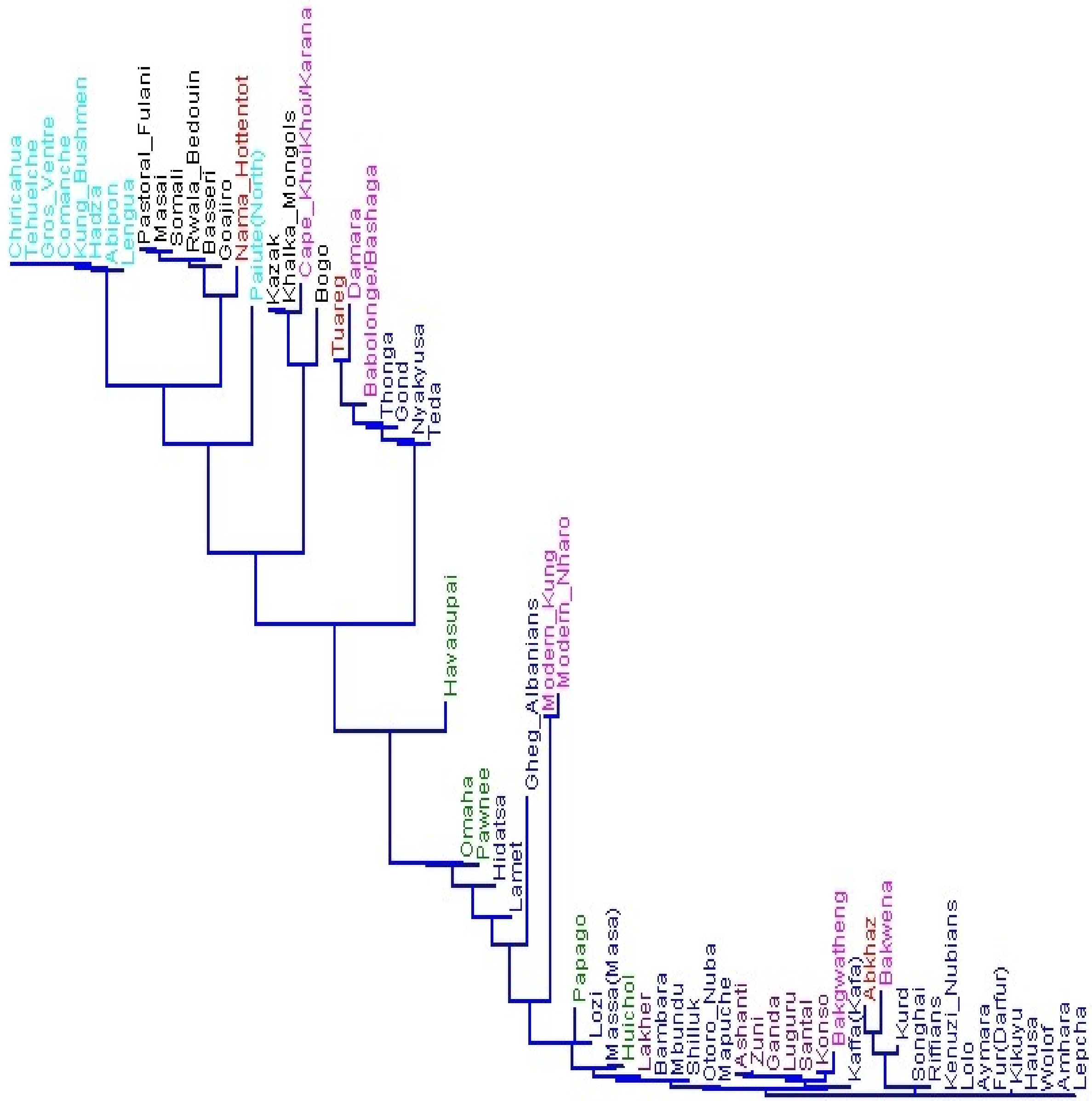

Figure 1: Hierarchical clustering dendrogram showing relative linkages between all societies included for analysis. All SCCS societies are color coded according to the k-means 6-cluster results (Table 3), and all the test societies (Table 2) are colored magenta.

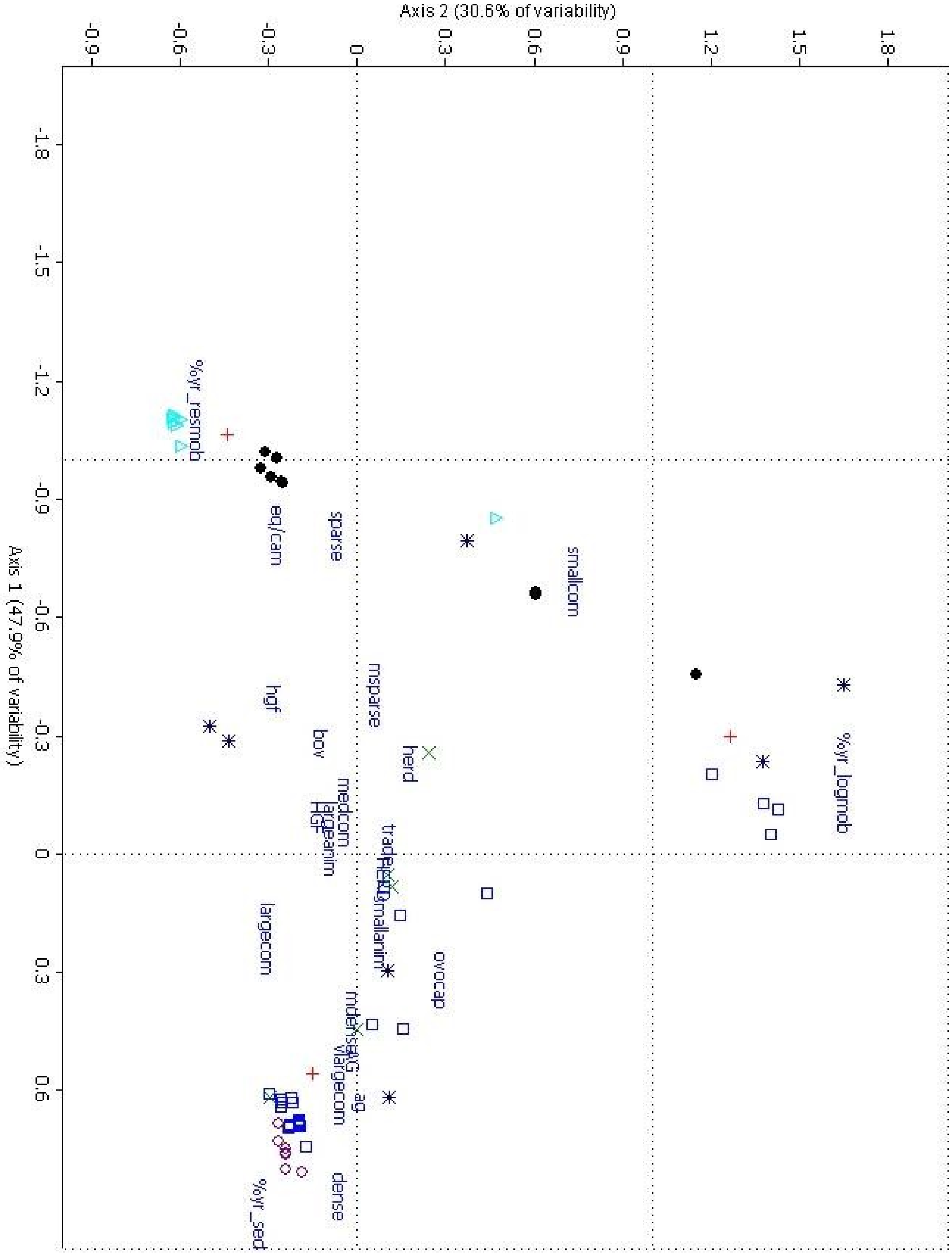

Figure 2: Plot of the 60 input SCCS societies in the coordinate space defined by CA axes 1 and 2, and color coded by the 6-cluster k-means results. Labels show the centers of the spheres of influence for each variable type in the same coordinate space. Test data points are shown as asterisks.

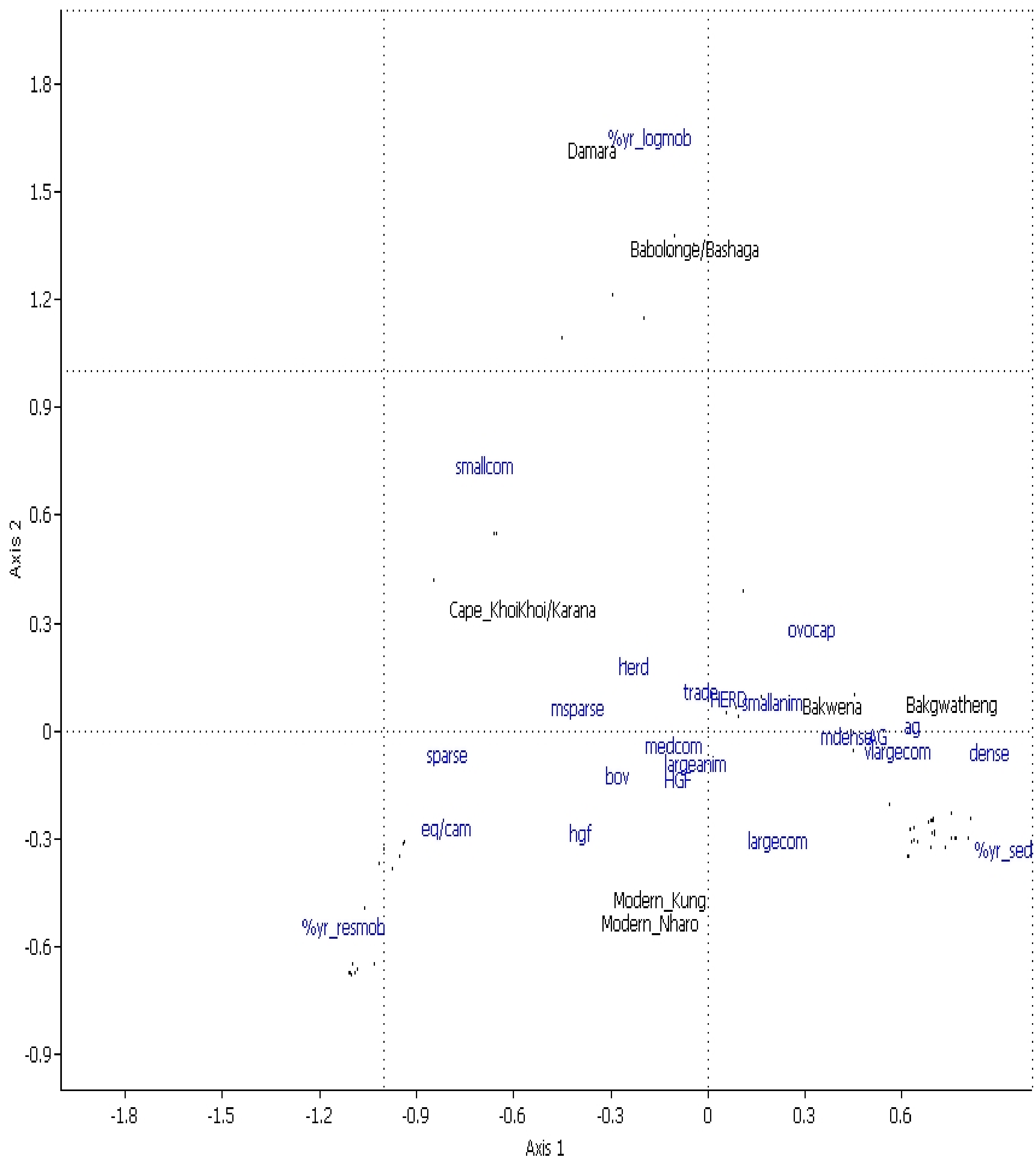

Figure 3: CA plot showing the location of the test groups in relation to the centers of the spheres of influence for the input variables

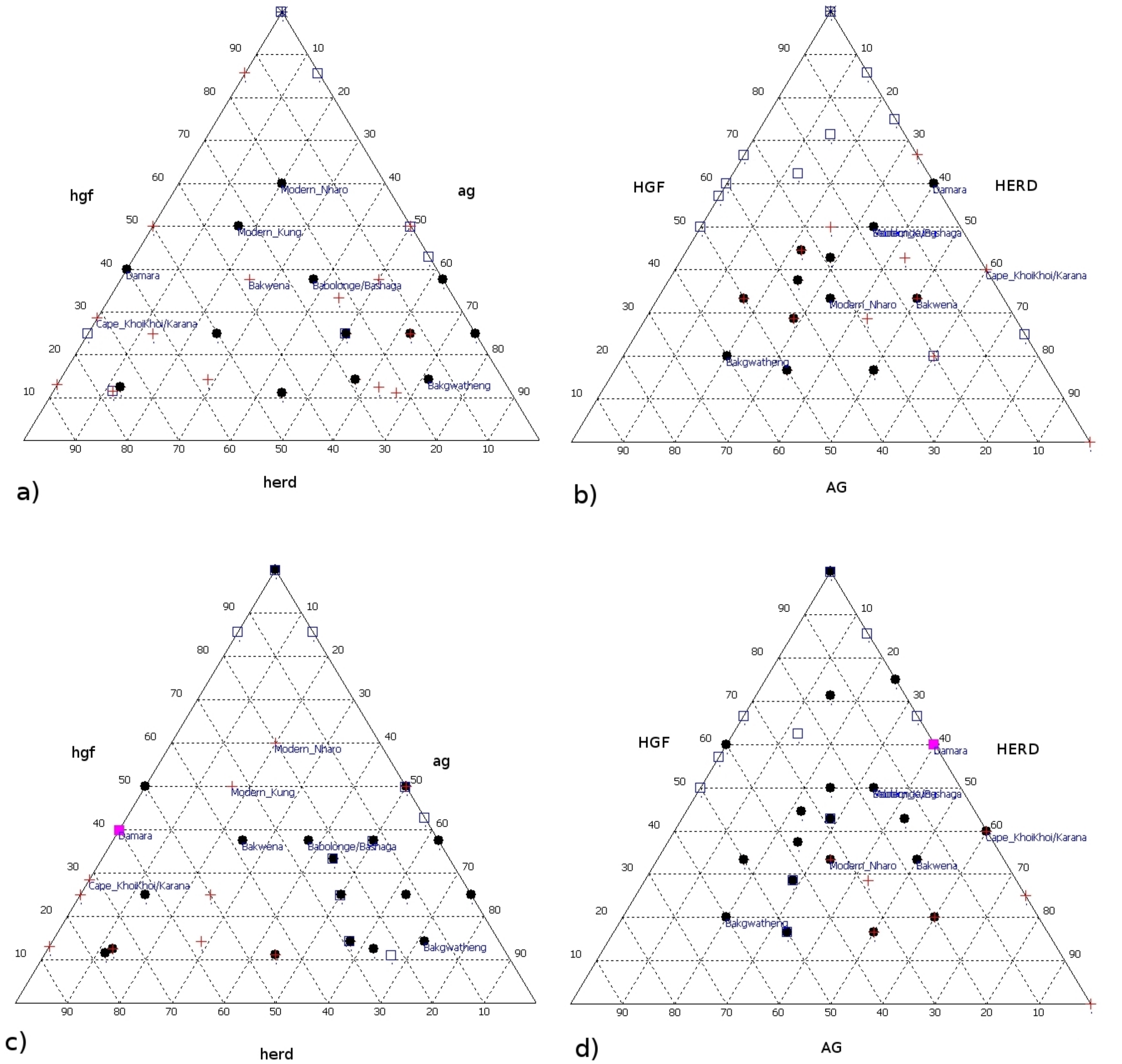

Figure 4: Ternary diagrams relating to subsistence. a) percentage of diet from foraged foods (hgf), pastoral foods (herd), and agricultural foods (ag) colored by amount of imported food items. b) percentage of subsistence time spent foraging (HGF), engaged in pastoralism (HERD), and engaged in agriculture (AG) colored by amount of imported food items. c) percentage of diet from foraged foods (hgf), pastoral foods (herd), and agricultural foods (ag) colored by main herd animal. b) percentage of subsistence time spent foraging (HGF), engaged in pastoralism (HERD), and engaged in agriculture (AG) colored by main herd animal. The names of "test" societies are indicated next to their symbol.

Symbol key: Herd Animals—None=*, Ovicaprids=●, Bovids=+, Equids/Camilids=□

Amount of imported food—None=□, Low=●, Moderate=+, High=■

References

Barnard, A.

1986 Some aspects of Nharo ethnobotany. In *Contemporary studies on Khoisan : in honour of Oswin Kohler on the occasion of his 75th birthday*, edited by R. Vossen and K. Keuthmann, pp. 55-82. vol. 1. 2 vols. H. Buske, G. Routledge & Kegan Paul, ltd., Hamburg, London.

1992 *Hunters and herders of Southern Africa : a comparative ethnography of the Khoisan peoples.* Cambridge University Press, Cambridge; New York.

Barnard, A. and T. Widlok

1996 Nharo and Hai//om settlement patterns in comparative perspective. In *Cultural diversity among Twentieth-Century foragers: An African perspective*, edited by S. Kent, pp. 87-107. Cambridge University Press, Cambridge.

Biesele, M., M. Guenther, R. Hitchcock, R. Lee and J. MacGregor

1989 Hunter, Clients, and Ssquatters: The contemporary socioeconomic status of Botswana Basarwa. *African Study Monographs* 9(3):109-151.

Binford, L. R.

1980 Willow's smoke and dogs' tails: Hunter-gatherer settlement systems and archaeological site formation. *American Antiquity* 45:4-20.

Denbow, J. R. and E. N. Wilmsen

1986 Advent and Course of Pastoralism in the Kalahari. *Science* 234(4783):1509-1515.

Feinman, G. and J. E. Neitzel

1984 Too Many Types: An Overview of Sedentary Prestate Societies in the Americas. *Advances in Archaeological Method and Theory* 7:39-102.

Guelke, L. and R. Shell

1992 Landscape of Conquest: Frontier Water Alienation and Khoikhoi Strategies of Survival, 1652-1780. *Journal of Southern African Studies* 18(4):803-824.

Guenther, M. G.

1986 *The Nharo Bushmen of Botswana : tradition and change.* H. Buske, Hamburg.

Hammer, Ø., D. A. T. Harper and P. D. Ryan

2001 PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* 4(1):9.

Kent, S.

2002 Autonomy or serfdom? Relations between prehistoric neighboring hunter-gatherers and farmer/pastoralists in southern Africa. In *Ethnicity, Hunter-Gatherers and the Other: Association or Assimilation in Africa*, edited by S. Kent, pp. 48–92. Smithsonian Institution Press, Washington, DC.

Kintigh, K. W. and A. Ammerman

1982 Heuristic approaches to spatial analysis in archaeology. *American Antiquity* 47(1):31-63.

Murdock, G. P. and D. R. White
2006 Standard Cross-Cultural Sample: online edition. In *Social Dynamics and Complexity*. Working Paper Series. Institute for Mathematical Behavioral Sciences, University of California Irvine, Irvine.

Okihiro, G. Y.
2000 *A social history of the Bakwena and peoples of the Kalahari of southern Africa, 19th century*. African studies. Edwin Mellen Press, Lewiston, N.Y.

Ross, R.
1980 Ethnic Identity, Demographic Crises and Xhosa-Khoikhoi Interaction. *History in Africa* 7:259-271.

Sadr, K.
1997 Kalahari archaeology and the Bushman debate. *Current Anthropology* 38(1).

Smith, A. B.
1983 Prehistoric Pastoralism in the Southwestern Cape, South Africa. *World Archaeology* 15(1):79-89.

1992 *Pastoralism in Africa*. Hurst & Company, London.

Solway, J. S. and R. B. Lee
1990 Foragers: Genuine or Spurious? *Current Anthropology* 31:109-146.

Stapleton, T. J.
1991 "They No Longer Care for Their Chiefs": Another Look at the Xhosa Cattle-Killing of 1856-1857. *The International Journal of African Historical Studies* 24(2):383-392.

ter Braak, C. J. F.
1985 Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics* 41(4):859-873.

Yellen, J. E.
1977 *Archaeological Approaches to the Present: Models for Reconstructing the Past*. Academic Press, New York.

1985 The integration of herding into prehistoric hunting and gathering economies. In *Frontiers: Southern African Archaeology Today*, edited by M. Hall, G. Averay, D. M. Avery, M. L. Wilson and A. J. B. Humphreys. BAR International Series, Cambridge Monographs in African Archaeology. BAR, Oxford.