APPLICATION OF GEOGRAPHIC INFORMATION SYSTEMS (GIS) TO SPATIAL SAMPLING AND DESIGN OF FARMERS' SURVEYS: THE CASE OF MAIZE IN KENYA

[1]R. M. Hassan, [2]P. Okoth, [1]J. Corbett, [2]P. Maingi

Abstract

GIS techniques were employed to design a geo-referenced survey of maize farmers in Kenya. Results from a cluster analysis of climate surfaces were used to define maize-specific adaptation zones. These were then combined with a maize area coverage, population maps, and a national sample frame to stratify maize production regions and develop the spatial scheme for allocating sampling densities and selection of sites and farmers to survey. This approach is appropriate for targeting distinct production situations and technology adaptation needs, and facilitates straightforward integration of survey, experimatal, and other data into the spatial database format for direct manipulation using GIS techniques. On the other hand, data sets compiled and processed plus methods applied to spatial stratification and design of the maize farmers' survey are relevant and easy to modify for other purposes, especially for planning commodity research and building spatially integrted information systems.

[1] International Maize and Wheat Improvement Centre (CIMMYT).

[2] Kenya Soil Survey, KARI.

Presented to the 3rd KARI Scientific Conference Nairobi, Kenya, November, 1992.

APPLICATION OF GEOGRAPHIC INFORMATION SYSTEMS (GIS) TO SPATIAL SAMPLING AND DESIGN OF FARMERS' SURVEYS: THE CASE OF MAIZE IN KENYA

1. Introduction

In Kenya, maize is produced under a wide range of agro-climates, soil types, and socio-economic conditions. Accordingly, farmers practices, productivity levels, patterns of input use, and adaptation of maize technologies vary significantly between locations. In order for maize researchers and policy makers to design appropriate intervention strategies for improved productivity in the maize sector, it is important to have a good understanding of the crop-environment interface in such diverse production situations. Any serious effort towards planning effective maize research in Kenya should therefore, start with characterizing the spatial variation in maize production systems, state of technology adoption, and constraints to increased productivity. Classification of maize production environments, is therefore needed to target broadly defined farming domains.

GIS techniques provide the tools for spatial classification and analysis. The present study utilized GIS techniques to design a geo-referenced survey of maize farmers across Kenya and develop procedures and methods for integrating survey and other data sets into the spatial base. Apart from the methods developed by the Maize Data Base (MDB) project, the spatial data sets compiled for this exercise will be relevant and very useful for other applications, especially in planning commodity research. Moreover, the classification technique used for developing maize-specific climate surfaces can easily be modified to develop different classification of production environments for other commodities.

Data sets and techniques of spatial analysis utilized in developing the sampling frame are discussed in section 2. Section 3 describes the resulting design of farmers survey and discusses the sampling strategy. A summary is presented in Section 4.

2. Data and Methods

In order to develop appropriate sampling clusters, it was decided that climatic differences would provide the best criteria for separating different regimes where maize is grown in Kenya. The existing agro-ecological zonation of Kenya (Sombroek et al., 1982, Jaetzold and Schmidt's, 1983) was considered too general in areas with flat terrain due to lack of definite criteria for separating the climatic boundaries. Another shortfall of these systems is their use of mean annual rainfall to classify the agro-ecological zones. This was considered inappropriate for crop specific zones because annual rainfall does not address the aspect of seasonality which is very important for most crops. Total annual rainfall may appear to be adequate, but when amounts of rainfall received during the growing season are computed a shortage is realized.

It was therefore decided that geostatistical interpolation methods be used to refine the climatic zones in areas where maize is grown. The refinement was based on a statistical cluster analysis method used by Pollak and Pham (1989). This method uses a digital elevation model (DEM), precipitation data, mean minimum and maximum temperatures, and mean monthly temperatures.

2.1 The base map

A digital elevation model (DEM) for Kenya was extracted from the DEM for the whole of Africa developed by Australia's Centre for Resource and Environmental Studies in 1992 (Corbet 1993). The DEM was originally constructed at a 1 km ground resolution. In order to reduce the volume of data, the DEM was contracted by selecting every fifth point so that the final DEM had a ground resolution of 5 km. This was further reduced to cover only areas where maize is grown in Kenya. A digital ARC-INFO map produced by the Department of Resource Surveys and Remote Sensing (DRSRS) of the Ministry of Planning and National Development in Kenya, showing spatial coverage of the maize growing areas in Kenya was used. The ARC-INFO map was first converted into an IDRISI raster image to create a mask from which all valid elevation points in the 5 km DEM could be extracted. The result was 7167 valid ground points. This formed the basic map used for creating the agro-climatic zones.

2.2. The climatic classification

For the construction of temperature surfaces, there were 81 stations for which monthly minimum and maximum temperature data were available. For precipitation surface interpolation, data from 979 stations were used (Jones, 1988). These datasets did not fill all the valid ground grids described above. Statistical interpolation was used to allocate values to the empty grid cells. The interpolation algorithm for the climatic classification followed a hierarchical cluster analysis using Ward's minimum variance clustering method (Ward, 1963). The variables analysed based on long-term monthly averages included mean, maximum, and minimum monthly air temperatures, mean monthly precipitation, and elevation. Similarity between two grid cells was evaluated by squared Euclidian distance (Kendall 1980).

The cluster analysis was stopped at 50 clusters which were then reduced to 12 based on maize adaptation classification as noted by Pollak and Corbett (1993). For precision and applicability to a cropping season, data for only six months, March to August (Long rains season) were used. This was found to represent the main maize production season and differences in them could easily indicate different maize growing conditions. The result of the cluster analysis were interpreted as maize adaptation zones (Figure 1 and Table 1).[3]

2.3 Population data

Population density formed the second clustering component. The pre-1989 Central Bureau of Statistics (CBS) population census was used to determine the population density within the maize growing areas. First, the number of households per sublocation were obtained from an informal survey preceding the actual population census. This information was originally stored in a data base (DBASE) and later combined with a spatial digital coverage of the administrative units within the maize growing region. In order to have manageable population strata, it was decided to group population density into the following four classes:

---

[3] Detailed discussion of the cluster analysis procedure used for Kenya is found in Corbett 1993.

(a) Areas with less than 100 persons/km^2

(b) Areas with 100-400 persons/km^2

(c) Areas with 400-800 persons/km^2

(d) Areas with more than 800 persons/km^2

The resulting population classification is depicted in Figure 2.

2.4 Intensity of Maize Cultivation

Maize coverage formed the third layer of information in the spatial sampling frame. This was considered important for locating samples for farmer interviews. The DRSRS produces a map showing maize cultivation intensity in Kenya every year. The map is produced using data collected by aerial photography and ground verification surveys (Ottichilo 1991). The intensity is interpreted on a district basis using units generated from 5 x 2.5 km transects. The map is converted to a digital format by digitizing the intensity boundaries on ARC-INFO.

DRSRS classified the country into 5 distinct maize production intensities as follows:

(a) Less than 0.5 percent cover - None to very low

(b) 0.5 to 5 percent cover - Low

(c) 5 to 10 percent cover - Medium

(d) 10 to 20 percent cover - High

(e) 20 to 40 percent cover - Very High

Figure 3 shows the distribution of the maize production intensities in Kenya. This information was used to convert the maize intensity zones to actual maize area in hectares by multiplying the median of the percent coverage with the area of each production class.

2.5 The National Sample Frame (NSF):

The National Sample Surveys and Evaluation Program III (NASSEP III) of the Central Bureau of Statistics (CBS) in Kenya has developed a NSF based on population census data. The NASSEP III has 1048 rural and 324 urban clusters (Table 2), spread in about 40 percent of the total sub-locations in the country. For several reasons, elected to adopt the NASSEP III clusters to sample maize farmers. The wide coverage and distribution of NASSEP III clusters in the maize production zone depicted in Figure 4, shows its adequacy for spatial sampling. Also, the NSF provides the most up-to-date listing of Enumeration Areas, plus the name and occupation of the household head, size of holding, as well as other information about households within each cluster.[4] Each cluster comprises over 100 households. Moreover, NASSEP III clusters data were available in ARC-INFO and hence ready to input with the other data sets described above. Most important was the saving of the considerable time and resources that would have been required in absence of the NASSEP III clusters, to develop a country wide frame for selecting survey sites using alternative procedures such as the spatial grid search.

2.6 Data Manipulation

The climate surfaces were originally formed in a raster based Geographical Information System (GIS) software package known as IDRISI. The maize area, population census, and NASSEP III information were in a vector based GIS (ARC-INFO). In order for the basic information sets to be manipulated together, the IDRISI

---

4 Updated for 1990-94 using the 1989 population census data (Akach 1990).

climatic zonation data were converted to an ARC-INFO data format for compatibility with the other three.

3. Spatial Classification and the Sampling Strategy:

The ARC-INFO spatial analysis module was used to overlay the climate surfaces map, the maize area, and population data to create a union of the three layers of information in one database. The union command ensured that no information was lost from the three data sets. From this data base tabular information contained in the DBASE of ARC-INFO was exported to LOTUS and manipulated to give the results shown in Table 3 which were used in the sampling design.

3.1 The sample size

The statistical theory of sampling provides objective criteria for choosing an optimal sample size. As only part of the population is selected in the sample, surveys are therefore, designed to achieve the highest possible efficiency in approximating attributes of the underlying population using the sample data. Sampling precission can be increased by either reducing variability between sampling units or by increasing the size of the sample. With limited resources, gains in efficiency from larger samples are weighted against costs associated with sampling extra units. Due to lack of pre-survey data on the degree of variability within the population of maize farmers with regard to relevant variables, this study could not determine an optimal sample size statistically. However, maize farmers were grouped into homogenous strata using climate and population density variables (Table 3). Therefore, the sample size of 1200 farmers was determined on basis of budgetary resources and time available for the survey.

3.2 Allocation of Sampling Fractions

The distribution of area under maize in 1990 between the various Climate-Population (CP) strata is shown in Table 3. More than 50 percent of the total area under maize fall in the Mid Altitude Tropical (MAT) zone (dry, moist and wet) followd by the High Altitude Tropical (HAT) zone (38% in its dry, moist, wet, and cold parts). About 96 percent of the estimated total area under maize fall within the twelve Agro-climate zones (ACZs) of Table 3 which occupy about 15 percent of Kenya's total land area. The rest of the maize (4 percent) is grown in the remaining 85 percent of the total land. This meant that, there is one hectare of maize in every 1000 ha of land in the region classified as zero-to-less than 0.5 percent maize cover zone (Figure 3). This zone was accordingly, considered a no-maize zone and dropped from the sample.

In addition to their influence on farmers practices and choice of technology, population density and area under maize are important variables for impact assessment. It is necessary to consider how many hectares and how large is the population likely to be affected by planned technology and policy interventions in order to evaluate research priorities and determine allocation of scarce resources.

For these reasons, both the population density and the proportion of area under maize were used to design the farmer survey. Sampling fractions were allocated in proportion to the intensity of maize cultivation i.e. by giving higher weights to zones with a larger proportion of the land under maize. Similarly, more weight was given to regions with higher population density (Table 4). The two weights were combined multiplicatively so that more farmers are sampled from regions with higher concentration of maize production and population density (except for the Very-High population density which was allocated the lowest weight because it represents urban centres with little or no maize).

The last column of table 4 shows the distribution of the 1200 farmers in the sample between the 12 ACZs. At this juncture, the issue of what is the optimal number of survey sites (or clusters of farmers) and farmers per site become critical. The greater the number of survey sites, the more costly the survey becomes. The smaller the number of farmers per survey site, the larger the variance and the less efficient the sample estimates become.

A sample size of 20 farmers per survey site (cluster) was considered to provide adequate representation of farmers and to capture the within site variability. These sites were termed as Sampling Units (SU). For a CP stratum to qualify to be allocated one SU, its weighted share of the total area under maize had to exceed 1.7 per cent (20/1200) which is equivalent to 16,698 Ha of maize).

The following procedure was used to allocate SUs and select CP strata from which farmers will be sampled:

a. The total sample size was allocated between the ACZs in proportion to area under maize adjusted by population density weight (Table 4).

b. ACZs with less than 1.7 percent weighted maize area were dropped. These included the Cold High-Tropical, Dry Low-Tropical and Wet Low-Tropical zones (Table 4).

c. SUs saved by dropping ACZs were added to the zone with the highest percent weighted area among the excluded, which is then retained in the sample. Accordingly, the DLT zone remained in while the WLT and CLDHT exited the sample.

d. Earned SUs were then allocated among population sub-classes within the selected ACZs using the weights established in Table 4. The minimum requirement size of a sampling unit was reduced to 10 farmers because less variation was expected within than between ACZs.

e. It has been decided to select a minimum of two SUs from each of the ACZs entering the sample to capture possible spatial variations not controlled for by the climatic classification.

Table 5 shows the allocation of SUs between the various CP classes. The number of SUs in the (MLT) zone was increased to four sites to represent the three distinct maize production systems at the coastal region (i.e. the cashewnut, coconut, and open zones) and allow for one site outside the coastal strip (Figure 1). As a result, the total sample size increased to 1300 farmers. Notably, all the CP strata in the Very-high population density never qualified as sampling sites for they mainly represented urban centres.

3.3 Selection of Survey Sites

Each survey site or sampling unit comprised of 20 farmers selected at random. A multi-stage stratified random sampling procedure was adopted for selection of sites and farmers within each of the final CP strata that passed the minimum weight test (of 10 farmers). The procedure adopted for selection of survey sites utilized the NSF of NASSEP III.

To locate survey sites within the selected CP strata, a random spatial search was applied on the NASSEP III clusters within the strata. A cluster was randomly selected for every sampling unit. The NASSEP III clusters were coded in a monotonically increasing order across the various districts of Kenya. Clusters falling in each CP stratum were arranged in order of cluster codes. To ensure an even distribution of the selected clusters across districts, systematic random sampling was used to select clusters. A sample of 20 farmers, and a reserve of 5 farmers to provide for non-response, were drawn from each selected cluster using systematic simple random sampling. The distribution of the selected survey sites (clusters) is presented in Figure 5.

4. Summary:

The diverse agro-climatic conditions under which maize is produced in Kenya resulted in a wide range of maize farming practices, productivity levels, and technology needs. Therefore, characterization of the spatial variability in maize production situations, state of technology adoption, and constraints to productivity growth was considered crucial to designing appropriate intervention strategies in the maize sector and for effective planning of maize research.

GIS techniques were accordingly employed to stratify maize production regions and develop a spatial sampling frame suitable for the country-wide survey of maize farmers in Kenya. Climate surfaces created through multi-variate cluster analysis were combined with population density data, and a coverage of area under maize to perform the spatial classification of maize zones. Sampling fractions were allocated between the various strata in proportion to intensity of maize cultivation weighted by population density factors. The national sample frame developed by the central Bureau of Statistics was adopted for selection of survey sites and farmers within each site using multi-stage stratified random sampling. This approach was appropriate for subsequent integration of survey and other data into the spatial base. Moreover, the data sets compiled for this exercise and techniques developed for designing the geo-referenced survey of maize farmers will be relevant and suitable for many other purposes, especially planning of commodity research and building of spatially integrated information systems.

Table 1. Zone averages: elevation, seasonal precipitation and temperature (March-August) for the maize growing areas of Kenya.

Zone	Elevation in metres	Precipitation in mm	Temperature in °C
Dry Lowland Tropical	290	420	25.5
Moist Lowland Tropical	320	615	25.3
Wet Lowland Tropical	720	1085	26.0
Dry Mid-Altitude Tropical	1033	410	22.7
Moist Mid-Altitude Tropical	1320	620	21.5
Wet Mid-Altitude Tropical	1425	925	21.0
Dry High Altitude Tropical	1925	425	16.9
Moist High Altitude Tropical	1925	620	17.0
Wet High Altitude Tropical	2070	845	15.9
Cool High Altitude Tropical	2525	720	12.2
Cold High Altitude Tropical	2900	690	8.4
Too Dry	630	320	25.6

Table 2. NASSEP III National Sampling Frame

Population Groups	No. of clusters/ district	No. of districts and urban centres	Total No. of clusters
Rural strata District			
0 - 99,999	12	5	60
100,000 - 249,999	16	4	64
250,000 - 499,999	24	16	384
>500,000	36	15	540
Urban Strata(centres)			
All district Hqts and All towns with population of more than 10,000	-	58	324
Total			1372

Source: Compiled from Akach (1990).

Table 3. Area under maize by Agro-climate Zone and Population Density strata (ha)[a]

	Population Strata				
Agro-climate zones ACZ	Low density <100 persons/Km²	Medium 101-400 persons/Km²	High 401-800 persons/Km²	Very High >800 persons/Km²	Total[b]
NO MAIZE ZONE	40,809	n.a.	n.a.	n.a.	40,809 (4.08)
Dry Low Tropical (DLT)	9,925	1,838	114	39	11,916 (1.19)
Moist Low Tropical (MLT)	15,756	4,447	802	155	21,160 (2.12)
Wet Low Tropical (WLT)	4,696	591	1	n.a.	5,288 (.53)
Dry Mid-Altitude Tropical (DMAT)	57,703	42,133	1,944	342	102,122 (10.21)
Moist Mid-Altitude Tropical (MMAT)	55,597	165,573	54,955	13,658	289,783 (28.99)
Wet Mid-Altitude Tropical (WMAT)	11,423	89,221	25,465	5,888	131,997 (13.20)
Dry High-Altitude Tropical (DHAT)	22,197	64,760	8,937	3,010	98,905 (9.89)
Moist High-Altitude Tropical (MHAT)	33,344	89,816	35,369	6,104	164,633 (16.47)
Wet High-Altitude Tropical (WHAT)	28,200	46,329	4,371	938	79,838 (7.98)
Cool High-Altitude Tropical (COLHAT)	16,976	17,124	113	11	34,224 (3.42)
Cold High-Altitude Tropical (CLDHAT)	389	78	n.a.	n.a.	5,288 (0.05)
Too dry (VDRY)	12,897	4,492	848	n.a.	18,737 (1.88)
Total[b]	309,912 (30.99)	526,902 (52.70)	132,919 (13.29)	30,145 (3.02)	999,878 (100)

(a) Computed from the DRSRS maize coverage (1992).
(b) Figures in brackets denote % of the grand total.

Table 4. Weighted Percent Maize area by Climate-population Strata

	Population Density Classes				Total	
	Low	Medium	High	Very high		
Weight[a]	0.50	0.75	1.00	0.25	2.5	
Factor[b]	0.20	0.30	0.40	0.10	1.00	
	Weighted Percentage Maize Area[c]					No. of farmers[d]
DLT	0.74	0.21	0.02	0.00	0.96	**12**
MLT	1.18	0.50	0.12	0.01	1.80	**22**
WLT	0.35	0.07	0.00	n.a.	0.42	**5**
DMAT	4.31	4.72	0.29	0.01	9.32	**112**
MMAT	4.15	18.53	8.20	0.51	31.39	**377**
WMAT	0.85	9.98	3.80	0.22	14.86	**178**
DHAT	1.66	7.55	1.33	0.11	10.35	**124**
MHAT	2.49	10.09	5.28	0.23	18.04	**216**
WHAT	2.10	5.18	0.65	0.03	7.98	**96**
COLHT	1.27	1.92	0.02	0.00	3.20	**38**
CLDHT	0.03	0.01	n.a.	n.a.	0.04	**0**
VDRY	0.96	0.56	0.13	n.a	1.65	**20**
TOTAL[d]	1.74	65.85	24.89	7.52	100.00	1200

[a] Density of Population Weights.

[b] Weight divided by sum of weights (2.5)

[c] Calculated as: (Maize Area) * (Factor)/Adjusted Total Maize Area.

[d] (Weighted Percent Maize Area in the ACZ) * (Total Sample Size, 1200).

[e] Area under the No-Maize-Zone is excluded.

Table 5. Allocation of sampling units among the selected climate-population strata by weighted share of maize area

ACZ	Sampling Units by Population Density Classes				Sampling Units per ACZ	Number of farmers per ACZ
	Low	Medium	High	Very High		
DLT	1	1	0	0	2	40
MLT	3	1	0	0	4	80
DMAT	2	3	0	0	5	100
MMAT	3	11	5	0	19	380
WMAT	1	6	2	0	9	180
DHAT	1	4	1	0	6	120
MHAT	2	6	3	0	11	220
WHAT	1	3	1	0	5	100
COLHT	1	1	0	0	2	40
VDRY	2	0	0	0	2	40
TOTAL	17	36	12	0	65	1300

REFERENCES

Akach, S.O. 1990. The National Sample Surveys Program (NASSEP). In proceedings of the Symposium of producers and Users of Statistics. UNDP/UNDTCD and Central Bureau of statistics, Ministry of Planning and National Development, pp. 54-68, Nairobi, Kenya.

Corbet, J.D. 1993. Classification of Maize-specific agro-climates in Kenya: A cluster Analysis approach. KARI/CIMMYT Maize Data Base project Document No. 3, Kenya Agric. Research Institute, Nairobi Kenya.

Hassan, R.M., J. Corbett, D. Karanja, and P. Okoth. 1992. The Spatial Sampling frame and design of farmers' survey. KARI/CIMMYT Maize Data Base Project Document No. 4, Kenya Agricultural Research Institute (KARI), Nairobi, Kenya.

Jaetzold, R., and H. Schmidt. 1983. Farm Management handbook of Kenya. Ministry of Agriculture, Nairobi, Kenya.

Jones, P.G. 1988. Climate version 3.5. World Tropical Climate Database. Agro-ecological studies unit, CIAT, CALI, Columbia.

Kendall, M. 1980. Multivariate analysis. Second edition. McMillan Publishing Co. Inc. New, York.

Ottichilo, W.K., and R.K. Sinange. 1991. Long rains maize and wheat production in Kenya in 1990. Technical Report No. 140, department of Resource Surveys and Remote Sensing, Ministry of Planning and National Development, Nairobi, Kenya.

Pollak, L.M., and H.N. Pham. 1989. Classification of maize testing Locations in Sub-Saharan Africa by using Agro-climate data. Maydica (34):41-43.

Pollak, L.M., and J.D. Corbett. 1993. Using GIS datasets to classify maize-growing regions in Mexico and central America. Agronomy Journal (Forthcoming).

Sombroek, W.G., H.M.H. Braun and B.J.A. van der Pouw. 1982. Exploratory Soil Map and Agro-climatic zone map of Kenya. Ministry of Agriculture-National Agricultural Laboratories-Kenya Soil Survey, Nairobi.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. (58): 236-244.

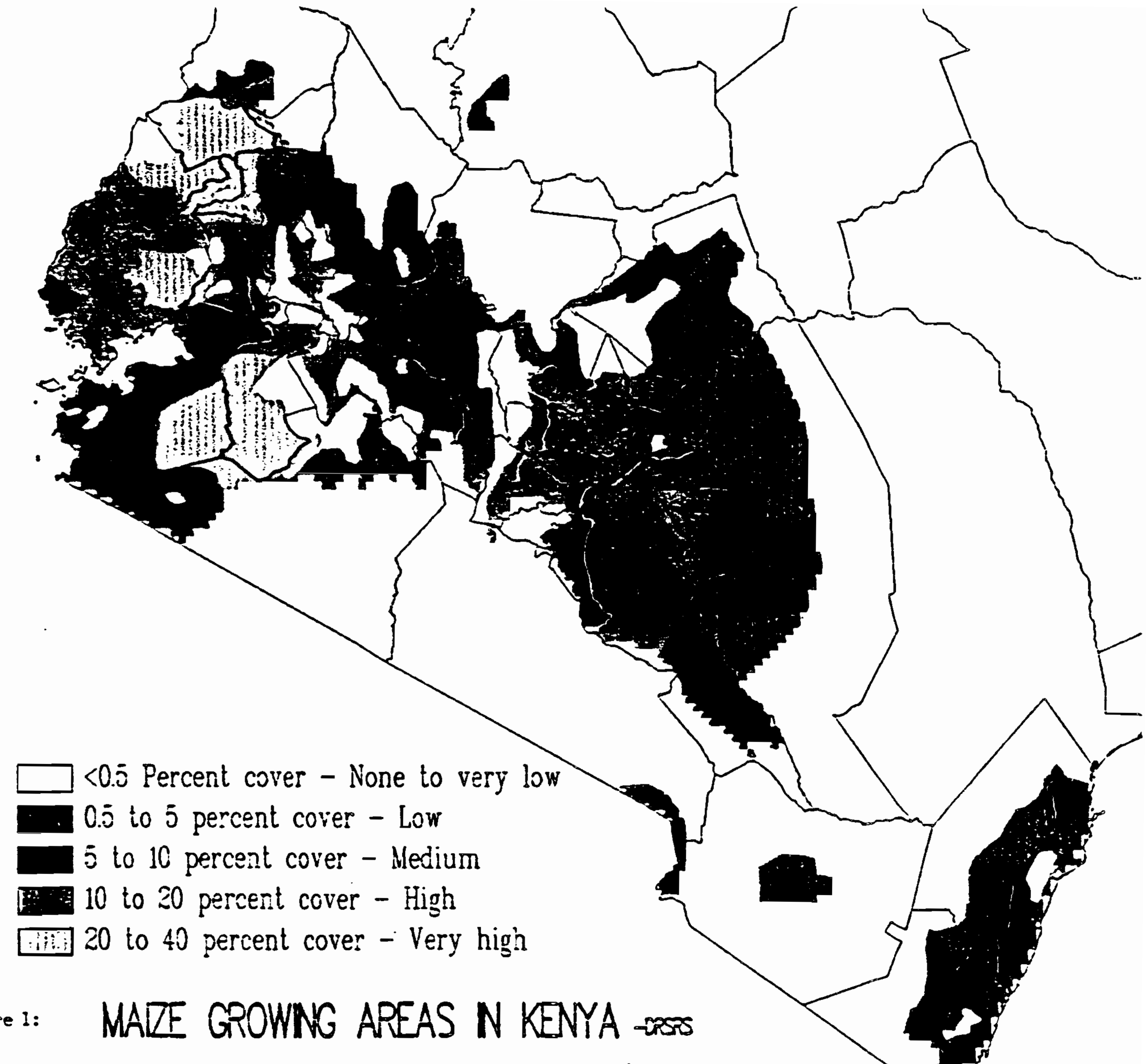


Figure 1: MAIZE GROWING AREAS IN KENYA -DRSRS

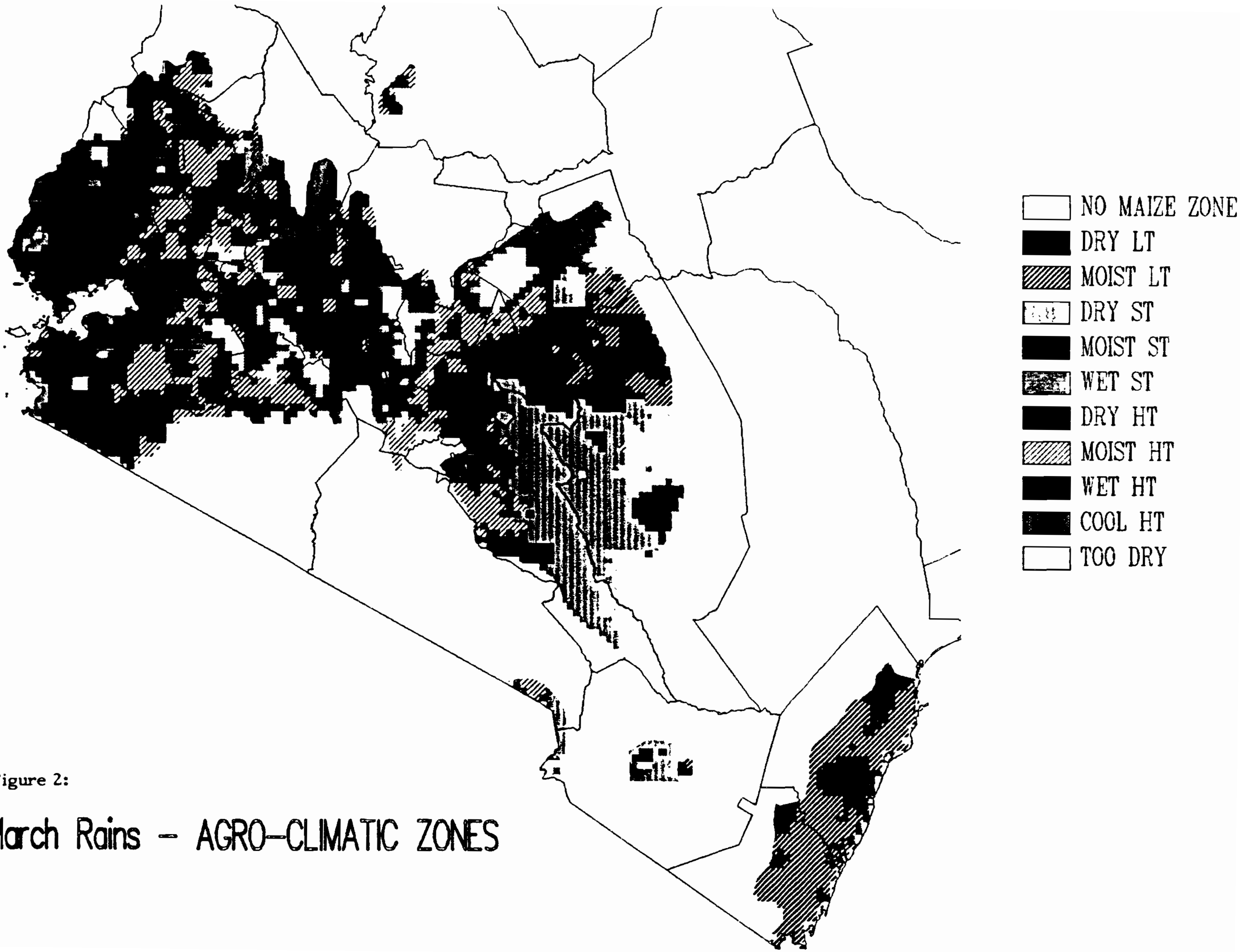
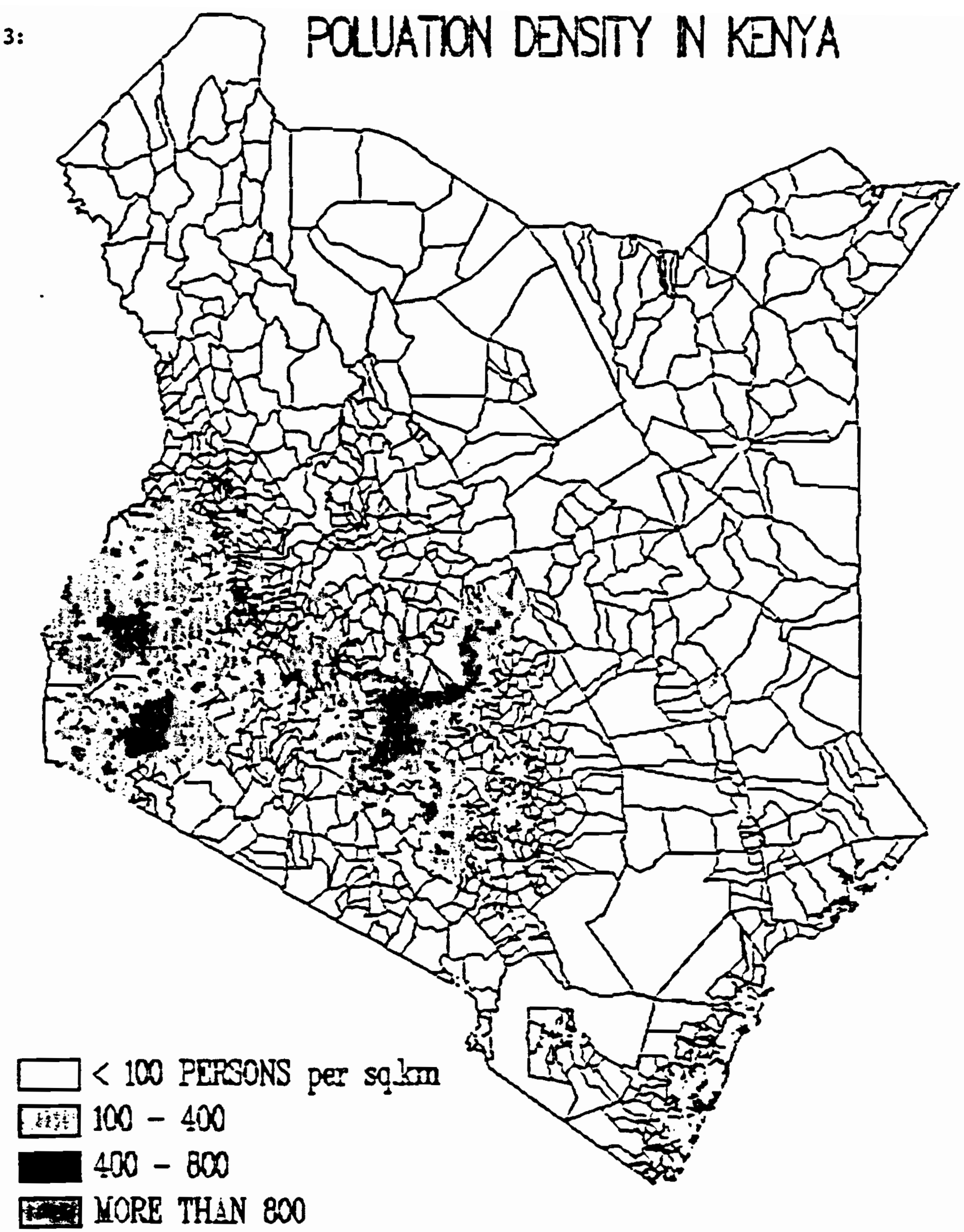


Figure 2:

March Rains - AGRO-CLIMATIC ZONES

Figure 3:


POLUATION DENSITY IN KENYA
< 100 PERSONS per sq.km
100 - 400
400 - 800
MORE THAN 800

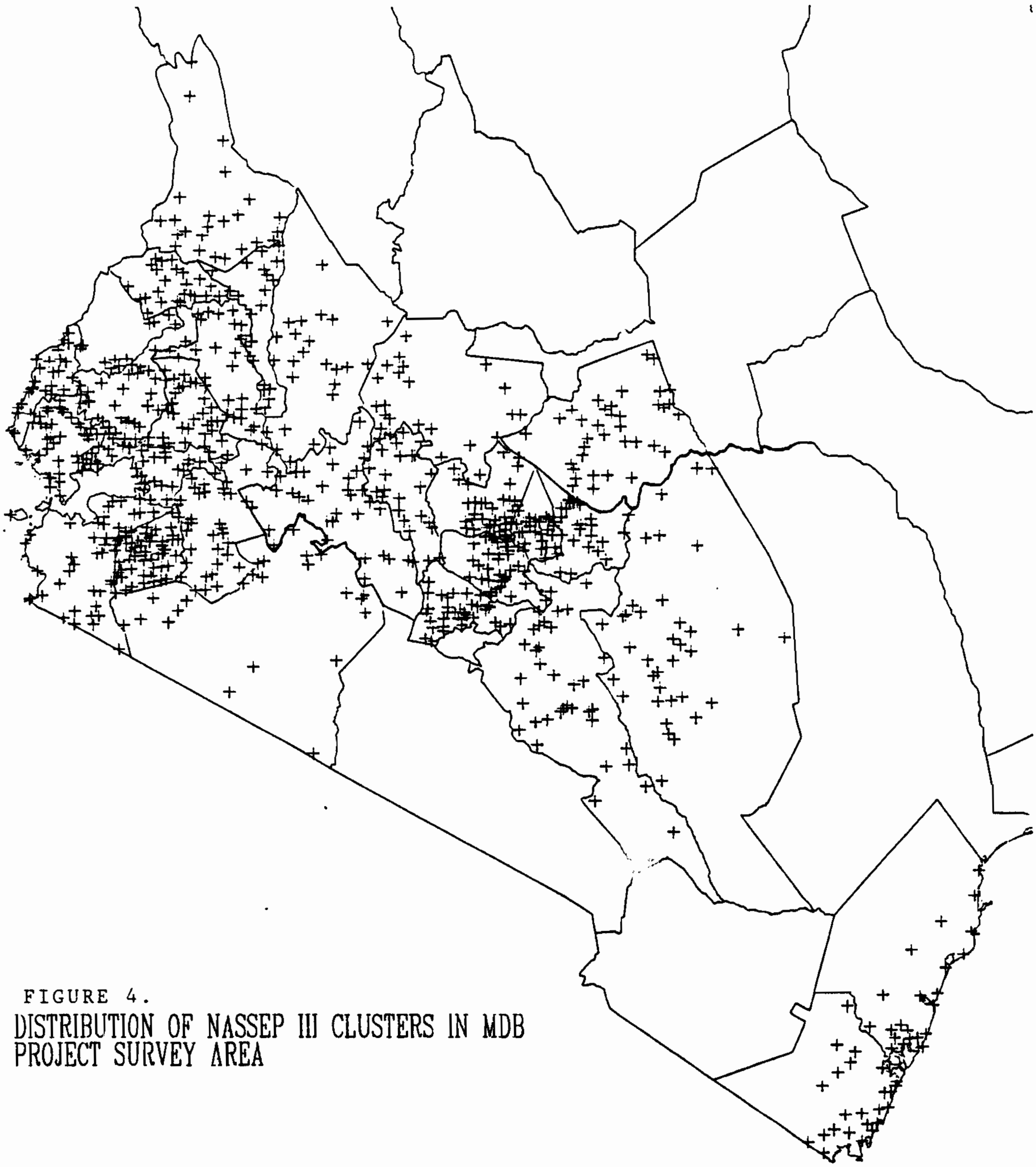
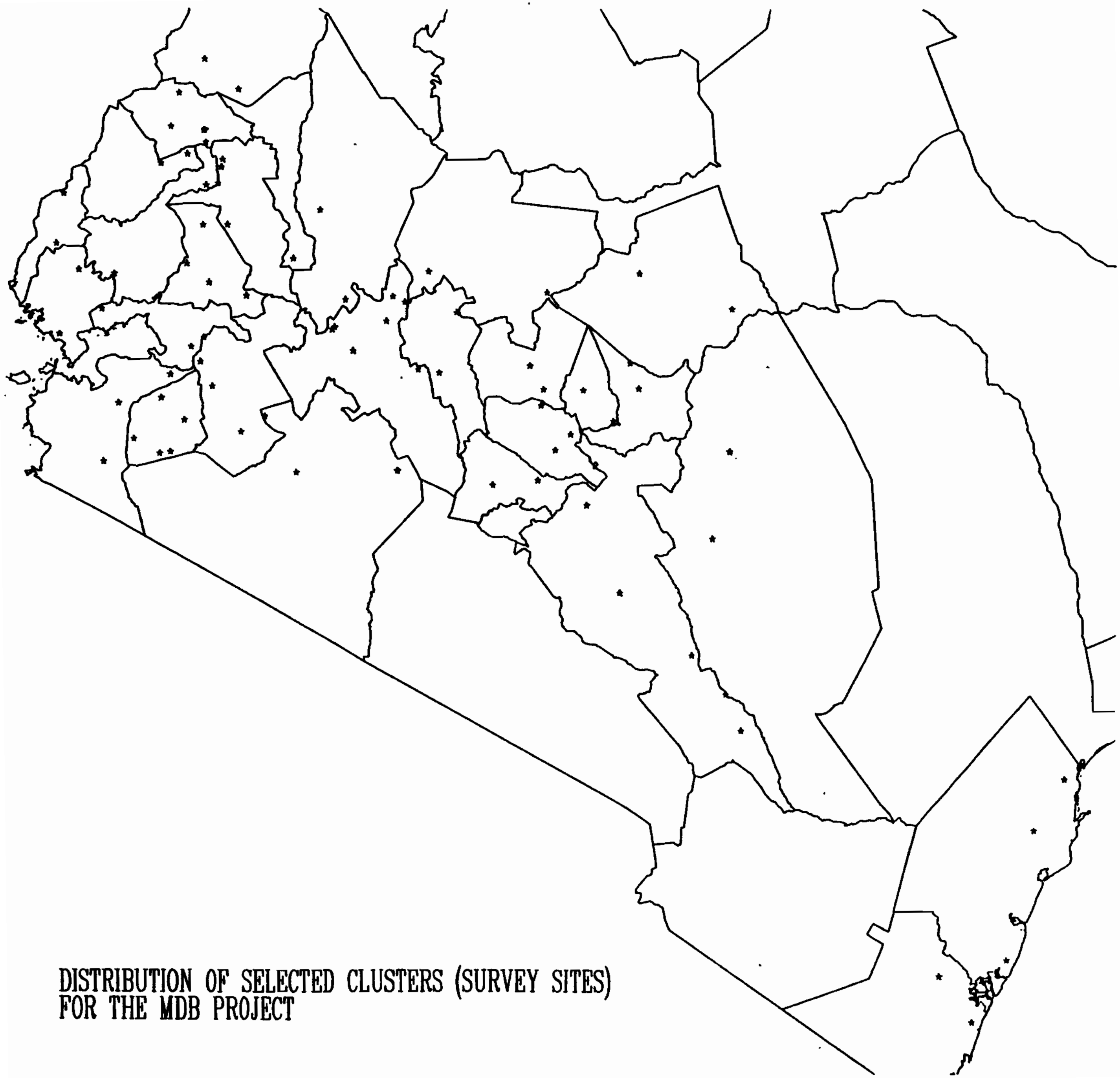


FIGURE 4.
DISTRIBUTION OF NASSEP III CLUSTERS IN MDB PROJECT SURVEY AREA



DISTRIBUTION OF SELECTED CLUSTERS (SURVEY SITES)
FOR THE MDB PROJECT