The University of Edinburgh

School of GeoSciences

Spatial Pattern Recognition for Crop-Livestock Systems using Multispectral Data

by

Adrián González

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

January 2008

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I am greatly indebted to the Venezuelan Government, specially to the Ministry of Science and Technology, where Marta Rodríguez (FONACIT) and Nuris Orihuela (Venezuelan Spatial Centre) rescued this study from its very early stages, and granted me economical means to complete this research. On the data side, I wish to thank The Venezuelan Geographical Institute (IGVSB), and The National Land Institute (INTI) which gave me much-needed satelite images and farms attributive information respectively. Such support have given me not only a indispensable material assistance to complete the job, but also the opportunities to be open to a completely new application within the agricultural field, to work on a project I am interested in, and to prepare myself to research in a very exiting field of study.

There are three people whom I would like to acknowledge individually. First, I am deeply grateful to Dr. Graham Russell, for his encouragement and careful guidance of my project. Secondly, I owe a great deal to Dr. José Ali Moreno (Universidad Central de Venezuela), my second supervisor, whose supervision and advise have contributed substantially to my study. Thirdly, I am enormously grateful to Dr. James Wright for his faith in me and his ability to provide the right support at difficult times.

Finally, I would like to thank my friends in Edinburgh: Matthew Palmer, Helen Bain, Graeme Allan, Alicia Salazar, Caroline Johnson, Kostas Ververidis, Carlos Acosta, Leyla Rios, Horacio González, Graham Walker, Allister Kidd and the Campanile family who have allways supported me during the long process of completing this project.

Abstract

Within the field of pattern recognition (PR) a very active area is the clustering and classification of multispectral data, which basically aims to allocate the right class of ground category to a reflectance or radiance signal. Generally, the problem complexity is related to the incorporation of spatial characteristics that are complementary to the nonlinearities of land surface process heterogeneity, remote sensing effects and multispectral features. The present research describes the application of learning machine methods to accomplish the above task by inducting a relationship between the spectral response of farms' land cover, and their farming system typology from a representative set of instances. Such methodologies are not traditionally used in crop-livestock studies. Nevertheless, this study shows that its application leads to simple and theoretically robust classification models. The study has covered the following phases: a)geovisualization of crop-livestock systems; b)feature extraction of both multispectral and attributive data and; c)supervised farm classification. The first is a complementary methodology to represent the spatial feature intensity of farming systems in the geographical space. The second belongs to the unsupervised learning field, which mainly involves the appropriate description of input data in a lower dimensional space. The last is a method based on statistical learning theory, which has been successfully applied to supervised classification problems and to generate models described by implicit functions.

In this research the performance of various kernel methods applied to the representation and classification of crop-livestock systems described by multispectral response is studied and compared. The data from those systems include linear and nonlinearly separable groups that were labelled using multidimensional attributive data. Geovisualization findings show the existence of two well-defined farm populations within the whole study area; and three subgroups in relation to the Guarico section. The existence of these groups was confirmed by both hierarchical and kernel clustering methods, and crop-livestock systems instances were segmented and labeled into farm typologies based on: a)milk and meat production; b)reproductive management; c)stocking rate; and d)crop-forage-forest land use. The minimum set of labeled examples to properly train the kernel machine was 20 instances. Models inducted by training data sets using kernel machines were in general terms better than those from hierarchical clustering methodologies. However, the size of the training data set represents one of the main difficulties to be overcome in permitting the more general application of this technique in farming system studies. These results attain important implications for large scale monitoring of crop-livestock system; particularly to the establishment of balanced policy decision, intervention plans formulation, and a proper description of target typologies to enable investment efforts to be more focused at local issues.

Chapter 1

General Introduction

Information about properties of farming systems has been essential in addressing agricultural issues. Such as censuses and surveys have long been the most widely used instruments to gather data on agrarian activities; indeed, historically they have proved to be a useful means of gaining knowledge of such diverse agrarian features as: dominant patterns of farm activities and household livelihood, including field crops, livestock, trees, aquaculture, grazing and forest areas, crop-livestock integration, technology, farm size and land tenure, to mention but a few. Nevertheless, the high requirements in terms of human and monetary resources of censuses and surveys prevent their application with the frequency and extent required to tackle the different levels of agricultural issues.

The rapid development shown by land observation satellites over the last three decades has made a great deal of information about land surfaces available. This has widely been used to study land cover changes; however, there is no evidence of using the spectral data gathered by those remote satellites in recognising patterns associated with agricultural land management due to the inherent complexities surrounding these activities.

It was in this context that the general aim of this research was defined; to pursue and to test the possibility of using learning machines to accomplish the task of pattern recognition for complex mosaics of farm inner land use in crop-livestock systems from multi-spectral data. These methodologies are based on the feature induction from a training set which seeks to estimate unknown dependences using a limited number of instances in order to produce models with a high generalisation capacity.

In basic terms, the general model of the pattern recognition process might be divided into a sequence of three main elements: a) generation of input random vectors with the information to be classified (sensor); b) translation of data into a statistically independent representation code, preserving their most relevant characteristics (feature extraction); and c) a system that, based on extracted features, can develop a function space where an operator might be built and serve as an answer predictor to any input generated by the sensor (classification).

Within the field of pattern recognition, one of the most studied subjects is the idea of approximating relationships from the inner land surface process and its emerging spectral response; using methods that can fit the complexity of these processes. This is vitally important for the study of crop-livestock production systems, given that these are critical to the livelihood of an important portion of the rural population in Venezuela (particularly in the Guárico river catchment) and at a worldwide level (Seré and Steinfeld, 1996; Bouwman et al., 2005). In addition, projections indicate that the demand for livestock food products is increasing globally (Delgado et al., 1999; Wint et al., 2000), and concern about the potential response of these systems is generally justified.

On this issue, a problem that remains open is the spatial monitoring of crop-livestock systems especially for those open range feeding, from which sometimes only time and site specific data can be approximated through field methods; but usually they are not cost effective and suffer from poor spatial connotation. It is also true, in a broader context, that public availability of space-based remote sensing has helped with the monitoring of land surface biophysical properties.

Some approaches have been concerned with the correction of observational data to create valued-added time series (Gleason et al., 2002; Green and Hay, 2002). Others in turn stress the use of optical, thermal and microwave data to model atmospheric and soil moisture (McVicar and Jupp, 2002; Dubayah, 1992); exploit the radiative transfer theory to estimate biophysical properties in vegetation (Wylie et al., 2002; Myneni et al., 1992; Goel, 1987); and macroscale modelling (Kimes et al., 2002; Asrar and Dozier, 1994).

In summary, most methodologies approach monitoring processes mapping land surface by classification or detecting change (Song et al., 2001). Nevertheless, the majority of the methodologies require a process of segmenting the feature space into non-overlapping regions; and hence feature vectors are mapped into one of the classes of interest, treating the problem in an unrealistic way (Brown et al., 1999). This research presents the use of a methodology where an optimal discrimination of pixel mixture might be inferred beyond a training set, by establishing a separating hyperplane between any two classes whose margin is maximum.

Aditionally, this methodology includes the inherent advantage of kernel functions, through which solutions are not built in the input space but into one with a higher dimensionality. In this feature space, it is possible that linear functions are enough to separate classes; given that input data are taken to the feature space by a nonlinear transformation whose diversity adds richness to the process of finding - if it exists - a solution. This flexibility is considered critical within the field of learning machines, making these tools attractive for use in pattern recognition processes for crop-livestock systems.

1.1 Objectives

On the basis of the advantages described above, the two main objectives of the present research are:

- to develop a typology based on census data of crop-livestock systems located in Urdaneta and Monagas-Guaribe counties of Aragua and Guarico states respectively in Venezuela;
- to test the ability of the Kernel-Adatron algorithm to classify the spectral response of farming systems land-cover, using previously identified crop-livestock categories.

As can be observed, this study involves both an unsupervised (1st objective) and supervised classification (2nd objective). The unsupervised classification of this research extends the identification of natural groups of crop-livestock farms in pursuing the following specific objectives:

- to represent and display spatial feature intensity of farming systems in the geographical space using census and survey data;
- to compare the effect of linear and kernel methods of feature extraction on the separation of crop-livestock groups when hierarchical clustering is used.

On the other hand, farm categories provided by the above part of the study are taken within the supervised phase as training instances. The aim of this is to typify the spectral properties of the cluster that training farms represent, using satellite images of their land-cover. To this purpose, the specific objectives for this phase are:

- to compare the kernel Adatron algorithm and linear discriminant analysis on achieving class separation accuracy for the spectral response of identified crop-livestock clusters; and
- to explore how differents kernel functions affect the classification performance of the kernel-Adatron algorithm.

1.2 Study area

The study area research work undertaken was concentrated on the hillside agro-ecological zone in northern Guárico and southern Aragua states in the Venezuelan north central region (lat. 8-10°N, long. 66-68°W) (Fig. 1.1) corresponding to dry tropical forest according to Holdridge's classification (Holdridge, 1967).

This zone is characterised by acute and moderate to gentle slopes with local relief ranging from 150 to 700 meters above sea level. The area is underlain by Ultisols or Inceptisols soils which are mainly originated from Quaternary, Tertiary and Cretaceous deposits leading to the expression of heavy textures and stony areas in those places where parental material outcrops; even though, towards the western part of the study area more clayey textures can be observed (Mogollon and Comerma, 1995). Generally laminar erosion problems are active, and seasonal flooding phenomena are quite commom given low internal drainage of rain water or river flash flooding.

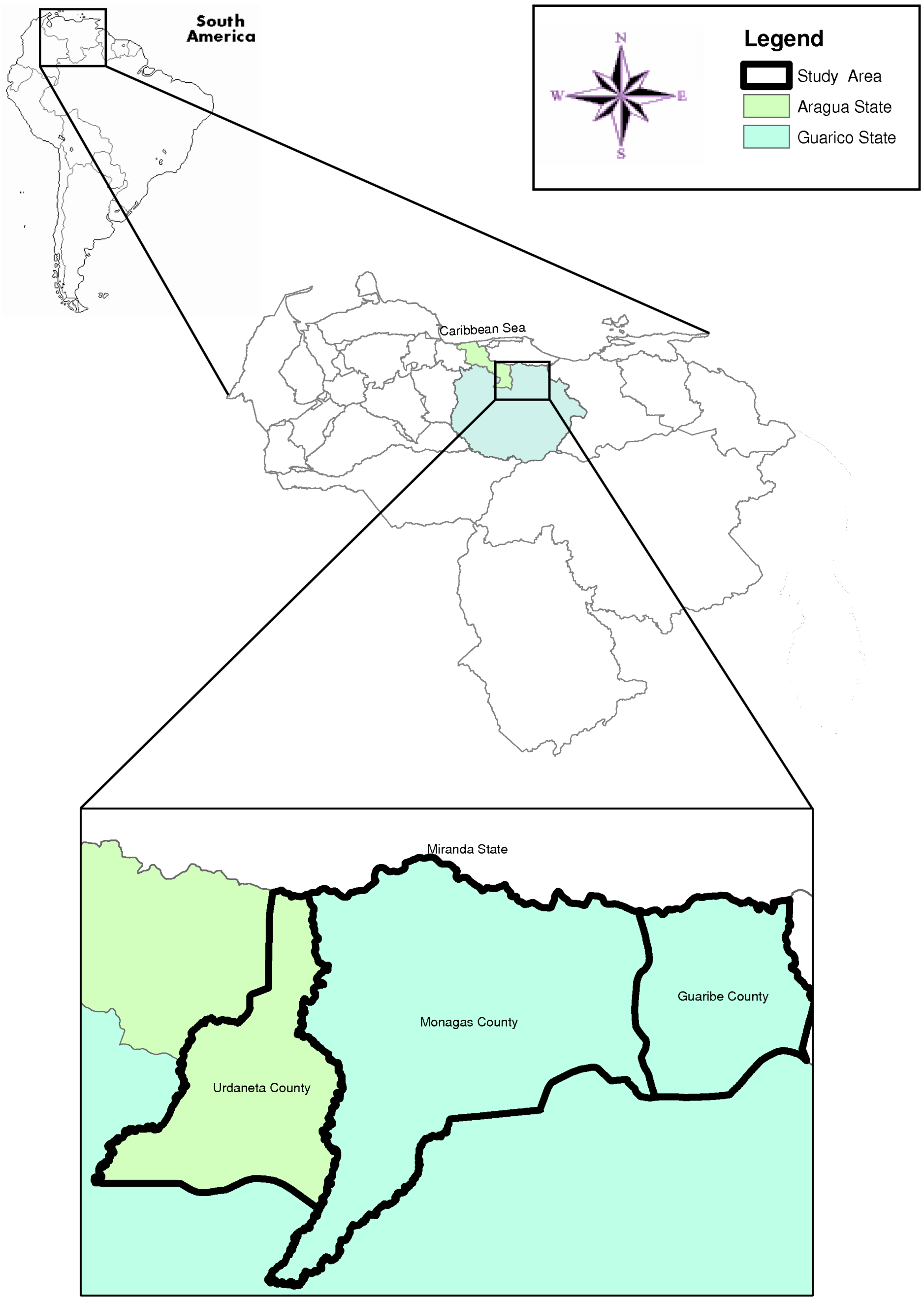


Fig. 1.1. Study area location

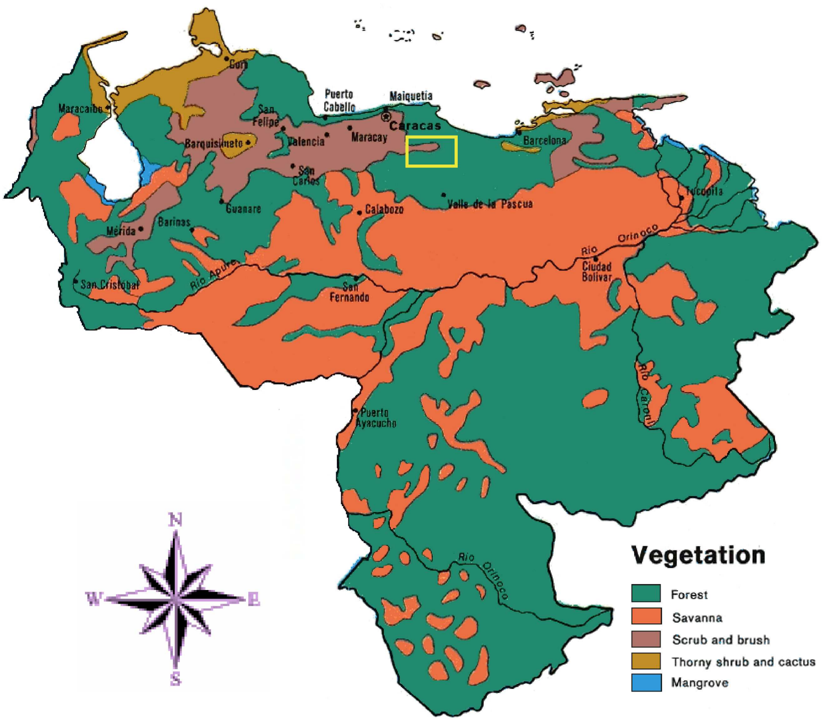


Fig. 1.2. Spatial variations of vegetative covers in Venezuela (Huber and Alarcon, 1988); the yellow box encloses the approximate area where sampling took place.

Climate is strongly seasonal, with a rainfall pattern characterised by 4-6 dry months and an average rainfall of about 1400 mm, evaporation 2000-2650 mm, and evapotranspitation 1500-1987 mm per year respectively. Mean annual air temperature is 28 C°with daily ranges between 10 and 14 C°.

Fig. 1.2 shows the vegetation cover of Venezuela. As can be appreciated, forest is one of the most important forms of land cover in this region occupying more than 50 % of Venezuelan territory, representing about 53 million hectares. It is noteworthy that within the study area (yellow box), the dominating plant physiognomies are represented by forest (deciduous and semideciduous) and scrub-bush covers (Comerma and Chacon, 2002). Both life forms are nowadays subject to considerable intervention by human activities, particularly those associated with selective wood extraction, crop production, and cattle grazing .

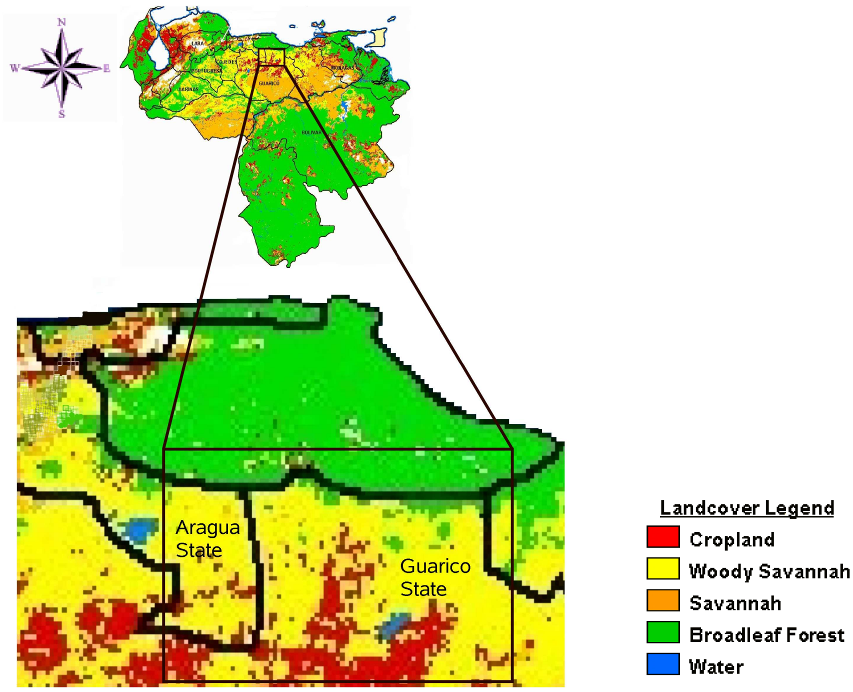


Fig. 1.3. Study area land cover classification from a Landsat ETM+ scene (USGS, 2005).

Fig. 1.3 depicts a more detailed land cover classification of the study area (black square). As can be seen, vegetation physiognomies observed in Fig. 1.2, here appear disaggregated as broadleaf forest, woody savannah and cropland covers. The dominant broadleaf species within woody savannah are: *Acacia macracanta*, *Mimosa arenosa*, *Mimosa tenuiflora*, *Prosopis juliflora*, *Cassia moschata, Entherolobium ciclocarpum, Pithecellobium saman*, and *Guazuma umlifolia* among others (Domínguez, 2006; Quiroz et al., 1997; Berroteran, 1997).

This plant community exhibits the tendency for all its phenophases (plant physiological changes due to seasonal climate variations) to occur according to a bimodal pattern, whose modes concentrate during the transition between rainy and dry season (Wright, 1996). The main consequence of this is that major phenologic changes (leaf-fall, fructification) practically happend during the dry season, leading to a consistent availability of feeding resources for cattle.

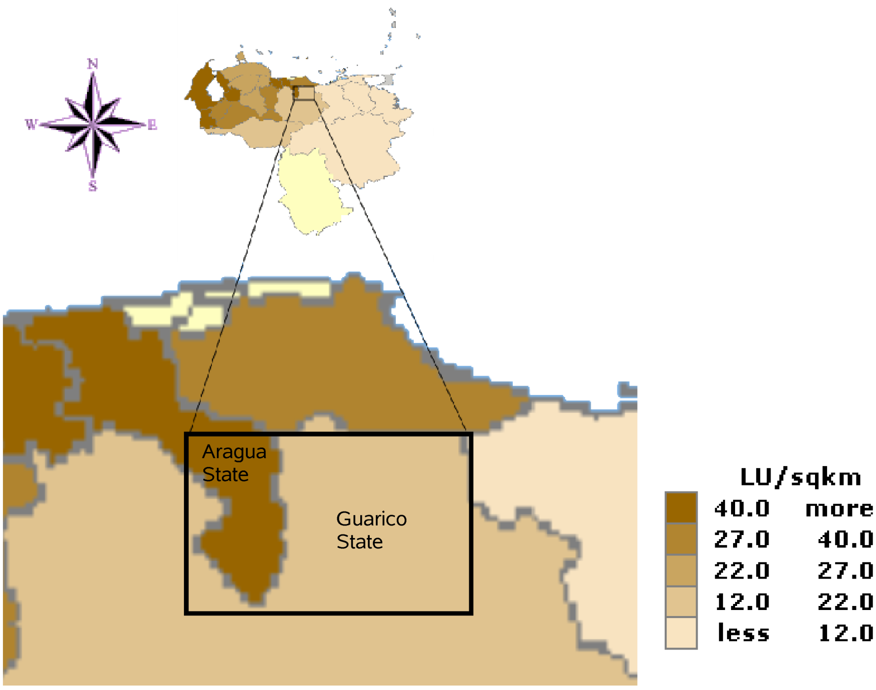


Fig. 1.4. Demography variation of Livestock units (LU) sqkm^{-1} in Venezuela (FAO, 2004). The black square corresponds to the approximate extension of the study area.

The dry season in this area is characterised by important shortages of herbage biomass in quantity and sufficient quality before high grazing pressure. In this sense, forest forage provision alongside crop (maiz and sorgum) harvest residues, enhance large herbivores' ability to optimise their diet choice, digestive balance and spatial utilisation of pasture (Murray and Illius, 2007; Baumont et al., 2005).

This alternative source of feed resources is particularly important in those regions that show high density of cattle population, as the portion of Aragua state (Urdaneta county) in the selected study area, which duplicates Guarico's counties in terms of the livestock units (LU) density as shown in Fig. 1.4. These changes in livestock density in Latin American countries, typically respond to the presure that urbanisation and crop production exert on grazing and mixed-farming systems (Delgado et al., 1999).

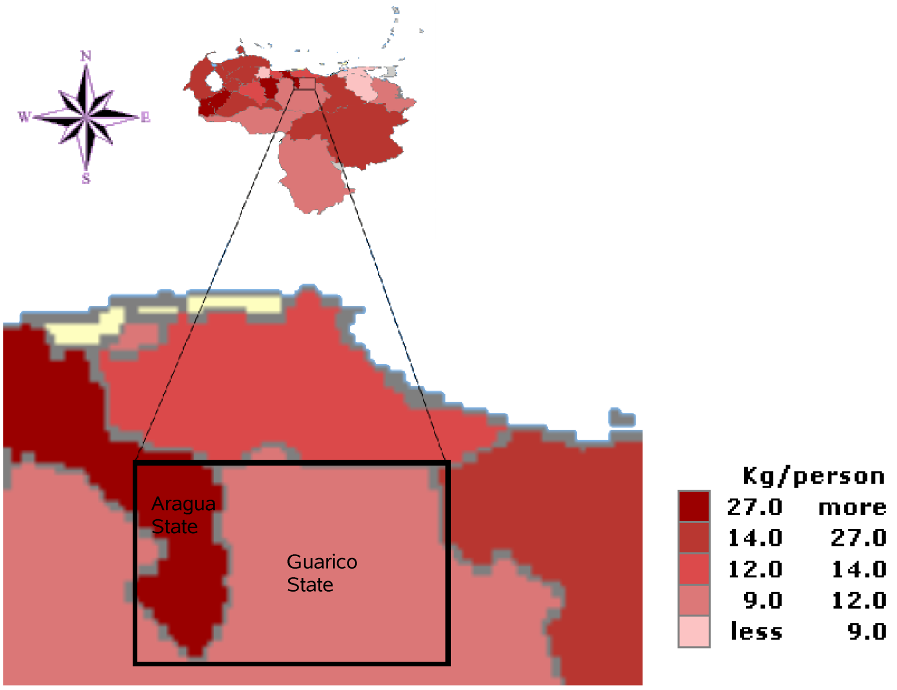


Fig. 1.5. Variation map of meat production kg person^{-1} in Venezuela (FAO, 2004). The black square corresponds to the approximate extension of the study area.

Fig. 1.5 depicts a map of meat production intensity related to human population. It is noteworthy that Aragua State, the most densely populated (211 people per square km) within the study area (black square), shows higher meat production per inhabitant when compared with Guarico state (9.8 people per square km). It has been found that urbanisation promotes the adoption of new feeding preferences in humans that generally lead to the consumption of more animal-source protein. Such variations in eating habits have been recognised as one of the global drivers that probably shape recent adaptations of the livestock sector toward more intensive production systems (Steinfeld et al., 2005).

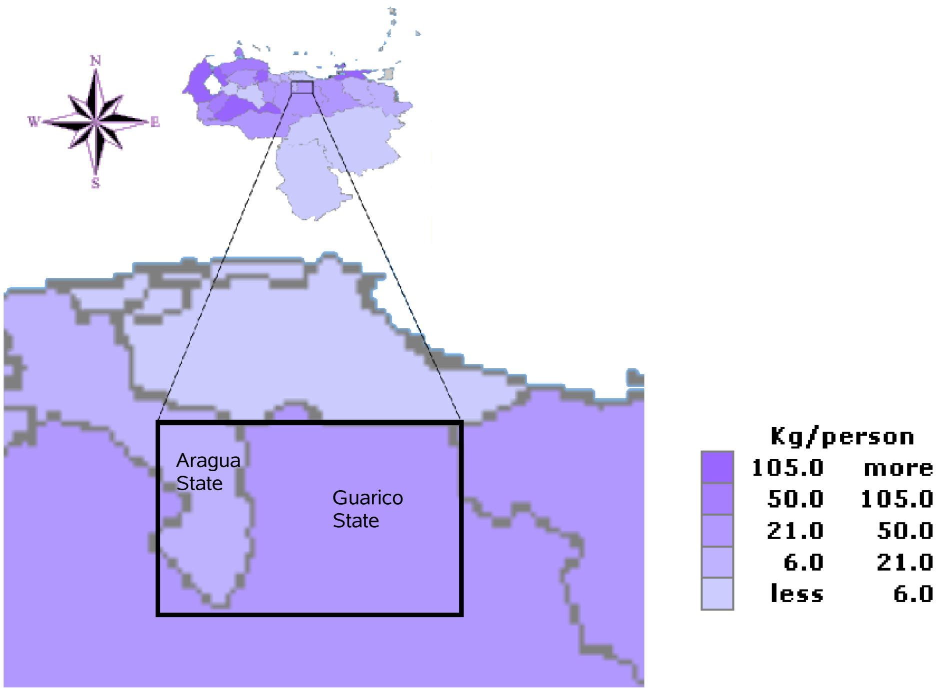


Fig. 1.6. Variation map of milk production kg person^{-1} in Venezuela (FAO, 2004). The black square corresponds to the approximate extension of the study area.

In contrast to meat production, which tends to be concentrated in more populated parts of the study area, milk production per person (Fig. 1.6) is higher in the less inhabited sampling region (Guarico state). This trend is consistent with recent findings about crop-livestock integration around the world, in which milk transportation from rural areas to cities is easier and cheaper than moving crop residues and feedstuff to support livestock feeding. As a result, the viability of urban dairies seems to be no longer possible given the above-mentioned transport cost constraints of consumables, and several environmental issues (Entz et al., 2005).

Cattle production systems settled in this area exhibit a wide scope of dynamic behaviour. Forage resources are highly diverse and strongly variable in quality and quantity, given the prevailing seasonal climate (wet and dry seasons); and their surface represent about 0.28 % of the total native pastures, and 2.14 % of cultivated pastures at national level (Table 1.1).

Table 1.1. Grassland-forest areas and cattle population at different ambits in Venezuela

Ambit	*Livestock*	*Native Pasture*	*Cultivated Pasture*	*Forest*
	$(Cattle - heads)$	$(hectares)$		
Venezuela	13,168,692	11,000,000	6,000,000	8,987,468
Urdaneta[a]	52,409	13,966	18,434	73,500
Monagas[b]	74,238	16,540	68,885	81,887
Guaribe[b]	35,221	951	41,614	29,005

a: County located in Aragua State
b: County located in Guarico State

Cattle population within the study area is about 161,868 head; mainly a mix of *bos indicus* -Zebu- (Brahman, Guzerat and Nellore) and crossbred *bos taurus* x *bos indicus* (Fig. 1.7). Moreover, different variants of crossbred might be found, such as Angus, Indu Brazil, Red Brahman, and also Italian breeds (Chianina, Romagnola and Marchigiana); however their proportion makes them less important (Payne and Hodges, 1997).

The performance of cattle within mixed crop-livestock systems (rainfeed and irrigated) under different livestock production technology can be seen in Table 1.2. These parameters are typical within the scope of mixed farming systems in humid-subhumid regions located in low tropical land according to the classification of Seré and Steinfeld (1996).

Currently, the dynamics of these systems are being estimated via field work studies, surveys and censuses (Domínguez, 2006; Espinoza et al., 2005). However, measured variables are not enough, and results are site and time specific and also difficult to extend to other regions beyond the sampling area. For this reason, traditional techniques for controlling and monitoring these systems are not very effective.

The two major tasks to be completed by a monitoring system in this area should include data gathering and processing in order to have an up to date knowledge of the system typologies. At this stage the main question that the system should answer is the actual spatial distribution on the geographical space of different farm typologies (Duvernoy, 2000).

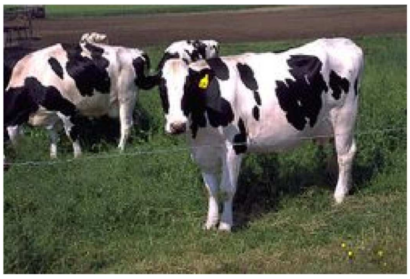
Even though the dynamic of farm typologies is highly dominated by climate, one of the major sources of perturbations are the socio-economical reforms (elimination of

(a) Guzerat breed

(b) Nellore breed

(c) Gyr breed

(d) Brahman breed



(e) Holstein breed

(f) Browswiss breed

(g) Crossbreed Browswiss x Zebu

(h) Crossbreed Holstein x Zebu

Fig. 1.7. Most used cattle breeds within the study area: *bos indicus* (a, b, c, d); *bos taurus* (e, f) and their crossbreeding (g, h).

Table 1.2. Performance parameters for cattle of mixed crop-livestock systems (rainfed and irrigated) in the Venezuelan lowlands.

Index	CR^a	CM^b	CCI^c	CI^d	*Stocking rate*^{e}	$Milk^f$	$Meat^g$
	(%)		$(days)$		$(UA\ ha^{-1}\ y^{-1})$	$(kg\ ha^{-1}\ y^{-1})$	
National Average	58	10	140	450	0.5	900	32
Improved System	77	≤ 5	70	400	1.0	5000	110
High Tech. System	89	$\simeq 2$	48.7	370	3.0	≥ 7000	427

CR: calving rate; CM: calf mortality; CCI: calving-conception interval; CI: calving interval
a: Vera (1998); b,c: Vaccaro (1998); d: Chicco et al. (1977); e,f: Combellas (1998)
g: (Chacon et al., 2004)

subsides, credit costs, price of agricultural consumables and demographic pressure) that impact the agriculture sector directly. The majority of these perturbations are very well known, since they come out as a result of variations at macro and micro economical level, even during those seasonal periods of high or low demand. Normally in these situations, statistics about previous experience might be used to characterise the relevance of the perturbation, predict the potential effects and also decide actions to be taken.

A very different case is perturbation resulting from climate change, meteorological systems and accelerated human demographic growing. Control in these scenarios is quite difficult given that previous experiences are scarce and responses to these changes are not easily predictable.

The complex pattern of land uses within similar crop-production systems, makes much more difficult the establishment of balanced policy decision and the formulation of intervention plans including investment efforts to be more focused on local issues.

To offer a basis for comparative analysis, researchers have proposed that existing models (crop and livestock) may be adapted in order to address the scope of integration potential. However, problems relating to rising dataset dimensions (especially from remote sensed data), the increasing complexity of analysis, problems associated with their extension to cover large territories and poor knowledge about specific land surface variables, are preventing the extensive use of this tool to monitor the depicted systems (Yang et al., 2007); and are placing a severe demand on computational power, preventing processors from completing simulations in a reasonable run-time (Armstrong, 2000).

In consequence, methods of performing complex pattern recognition aiming to construct farm typologies to inform specific development questions on local, regional or national scales are required. Currently, the massive availability of remote sensed data from this area, good quality agricultural censuses and validation field survey information, and boundaries of individual geo-referenced farms, have provided a unique opportunity, and motivated researchers to proceed with research of such nature based on the statistical learning theory. Two of the main advantages of the solution presented under this scheme are their broad spatial scope and the reduced spatial temporal parameterisation of biophysical variables.

1.3 Learning from instances in agriculture

1.3.1 General problem

The process of estimating an unknown input-output dependence and generalising it beyond a limited training set of observations is acknowledged as learning from instances, which had its origin in the pioneering work of Rosenblatt (1958). During the 1960's this paradigm was seriously hampered as a result of the work of Minsky and Papert (1969).

By this time it was thought that complex applications in the real world would require representation hypotheses much more expressive than linear functions. Given that the target concept could not normally be represented as a simple linear combination of data attributes, as a result some fields of study such as learning machine and pattern recognition were affected, preventing their use on applied research including farming systems. Later on it was demonstrated that the theories of Minsky and Papert (1969) were wrong.

Typification of farming systems has been one of the major approaches within the field of agricultural systems in which research has been conducted. This paradigm mainly refers to those methods characterised by induct non-supervised clustering of farms within a taxonomy; where farm likeness is represented according to a finite set of m-dimensional variables (Köbrich et al., 2003; Berdegue and Escobar, 1990; Kostrowicki, 1977).

During the 70's most of the learning techniques used in the agricultural system field were influenced by the wave of learning linear decision surfaces (Hart, 1990; Capillon, 1985; Kostrowicki, 1977). That kind of representation was preferred given that its theoretical properties were well understood. Beyond the 80's, researchers trying to move away from the limitation of linear models started using non-linear models in the application of decision trees and artificial neural networks. It should be said that these techniques were rapidly assumed within the agriculture domain with several applications. However, the main problems of these approaches were their theoretical weakness and that their solution space was full of local minima.

The consolidation and application of the statistical learning theory during the mid-90's allowed the development of efficient algorithms to learn non-linear functions. These ideas completely recast the pioneering work of Rosenblatt (1958); and were theoretically supported in the statistical learning theory (Vapnik, 1998, 1995; Vapnik and Chervonenkis, 1974).

Vapnik and Chervonenkis (1974) formalised the learning problem as a function estimation; where given an empirical data set generated by a regular stochastic distribution, the algorithm pursues the extraction of regularities in the data. The general model of learning from instances might be summarised in a sequence of components: a) an input vector generator; b) a system that produces an output value and c) a linear machine.

Contrasting with the statistical learning theory, which appeared on the scene quite recently, one current solution implementation is based on kernel functions (Aronszajn, 1950; Mercer, 1909), whose study was recognised about a century ago, and which has been playing an important role in increasing the representation capacity of the solutions especially in agricultural fields involving remote sensing. Its application within the learning task relates closely to data preprocessing; and along with the learning machine, constitutes a compact body.

1.3.2 Particular cases

Supervised and unsupervised learning are among the most investigated applications in agriculture. The former approach pursues building relations between input vectors and target outputs. The outputs may be expressed in different scales: categorically or numerically; corresponding to classification and regression problems respectively. The unsupervised approach, rather than approximating input data to a target label, seeks to approximate data by similarity expressions, generally, distance functionals, from which groups of data that resemble each other can be built. This paradigm is usually referred to as clustering (Bishop, 2006).

The remote sensing works of Hermes et al. (1999) and Huang et al. (2002) are precursors of the classification approach in an agriculture related field, where, given a spatially dispersed set of pixels, different forms of land cover (closed forest, open forest and woodland) were classified according to their spectral response. Other research of this kind is: the work of Keuchel et al. (2003) which progressively compares land cover classification using three methods (support vector machines, maximum likelihood and iterated conditional models); and the work of Su et al. (2007) which uses the multi-angle approach and its corresponding spectro-radiometer image to accurately map grassland types by support vector machines. A good application of learning machines on the regression problem is the work of Yang et al. (2007) within the forestry field. In this research the target vector used was eddy covariance-based gross primary production (GPP) and three remotely sensed variables (land surface temperature, enhanced vegetation index and land cover) in order to predict flux-based GPP at continental scale.

Regarding the clustering problem in the unsupervised ground, Diez et al. (2006) combine a kernel based similarity function and a support vector machine to permit the identification of public beef product preferences as they were stratified by market segments. In addition, within the unsupervised family can be found density estimators, which mainly pursues projecting data from a high onto a lower dimensional space seeking to determinate its distribution in the input space in order to add visual richness to the solutions represented (Bishop, 2006).

In summary, these methodologies are based on feature induction from a representative set of instances, where it may be possible to produce a model able to generalise beyond the training instances. In this way a description of relationships present in the original data is possible, and their representation is simplified at the same time that their main features are preserved. Today there is still a wide usage of linear paradigms in farming systems studies (Milá et al., 2006; Köbrich et al., 2003; Dobremez and Bousset, 1995). Nevertheless, extensive applications of linear machine techniques in agriculture are still scarce. The forerunners have shown that models generated are flexible, theoretically robust and provide expressive solutions. Some of the preliminary results of the present research may be found in González et al. (2007); and for those seeking a deep understanding the following publications are suggested: Bishop (2006); Shawe-Taylor and Cristianini (2006); Cristianini and Shawe-Taylor (2000) and Vapnik (1998, 1995).

1.4 Thesis outline

The whole document is structured as follows: initially the unsupervised part of this research is approached by both linear and nonlinear methods. Chapter Two is dedicated to the linear approach, alongside introducing the complementary methodology of representing spatial feature intensity of farming systems in the geographical space; while Chapter Three depicts the nonlinear apsects of the unsupervised classification process, studying and comparing it with the linear approach. Then based on these results, Chapter Four introduces the supervised classification stage of the study, describes experiments, and compares results with other standard supervised methods. Following this, a general discussion is presented in Chapter Five; and finally conclusions are summarised in Chapter Six.

Chapter 2

Spatial Geovisualisation and Unsupervised Classification of Crop-Livestock Systems

2.1 Abstract

The most common challenge in regional studies of crop-livestock systems has been the establishment of the operational basis for comparative analysis, given the multidimensionality of interactions within real-world earth surface and agriculture processes. Many scientists have explored crop-livestock systems using multivariate approaches that lead to the development of farming typologies (clusters) (Capillon, 1985; Berdegue and Escobar, 1990; Köbrich et al., 2003). These classifications are rarely spatially explicit, and the whole population is generally segmented based on previous experiences. In the present research, a complementary technique is proposed; this methodology makes use of a (3D) GIS-based spatial density model, that represents the spatial feature intensity of crop-livestock systems before they are segmented by hierarchical clustering. These methodologies were implemented using census data from geo-referenced farms of the central rural region of Venezuela (South Aragua and North-eastern Guarico states). The main findings were that surface feature density, in combination with cluster analysis, permits qualitative interpretation of continuous gradients in farming activities; enables its visualisation in geographical space without disruptions posed by political boundaries; provides guidance about the number of classes to use; and improves cluster analysis by

making explicit the spatial relationships between features that form the multidimensional density pattern.

2.2 Introduction

A farming type or modality is a representation of a population of farms that share the same n dimensional traits. Typically, farming system studies seek to define separate groups of farms by looking for a natural structure among the observations. The objective is to maximize homogeneity within clusters and heterogeneity between them (Hair et al., 1998; Dixon et al., 2001).

An issue that remains open is deciding the number of clusters into which a population should be segmented (Kostov and McErlean, 2006). It is not clear whether subjectively choosing the number of groups, while quantifying the structural features of the sets inevitably leads to meaningful and spatially explicit classes. The main difficulty is the reduction of geo-spatial information from farms and their constituent land uses into a group profile. As a result, geo-spatial data are treated as a multidimensional whole from a large number of attributes which are generalised into patterns that are sometimes difficult to observe.

Interesting solutions to this problem have been tested recently within the field of geo-visualization, where the analysis of geo-referenced activity data using GIS-based surface modeling is a very active area of study (Kwan, 2000; MacEachren and Kraak, 2001). Basically, it attempts to find a function that distributes the magnitude of point or line observations over an area to create a continuous surface. The main issue here is the assigning of meaningful features values to a Z component in a given bi-dimensional geographical framework (X, Y). The outcomes are visual settings of the data that not only facilitate a graphic exploration of multi-dimensional datasets, but also add spatial elements within data structures (Hernandez, 2007).

It can be hypothesized that for agricultural systems which represent complex and multiple related dimensions of a set of farm attributes the surface pattern might help to explain the role of spatial organisation in cluster assemblage. It would thus be of interest to learn how the patterns related to different activities might be compared between and within populations. If these results were confirmed, would they provide strong evidence for cluster number decisions?

To this end, this chapter has the following objectives: firstly, to visualize land use and productive crop-livestock data that permit exploration, display and comparison of the modalities by examining the activity density distribution across the spatial domain; secondly to classify farms into subsets according to common attributes from census data, and thirdly, to identify which variables are primarily responsible for the pattern observed among the classes.

This chapter first reviews the general aspects and concepts related to how processes that generate heterogeneous spatial responses in crop-livestock systems might be summarized by a set of indicators. This is followed by a brief justification of the use of PCA within this unsupervised classification phase, even when the dimensionality of the problem posed was low enough. The next session includes a description of the main exploratory procedures and data used in this research. Then, results and discussion are presented; and finally conclusions are summarized.

2.3 Visualisation of indicators and sources of spatial complexities in crop-livestock systems

In a broad sense, the process of geo-visualizing spatial information from biophysical data of crop-livestock systems, can be divided into the following sequence: firstly making spatially explicit the objects under study in terms of their extent and location (geo-referencing); secondly, the decomposition of reality into attributes that serve as indicators of the farming system to be studied (representation); and thirdly, to interpolate the value of these attributes and decompose them into geometric structures of easy manipulation for rendering purposes across a simulated geographical domain.

According to these terms the selection in crop-livestock systems of a number of the variables that permit their characterization belongs to the representation phase. Depending on the representation choice a semantic and meaningful farm classification might be achived. However; in most farming system characterization studies, the selection of atributes to represent farms is based on a certain degree of experience, or expert knowledge to guarantee a proper cluster segmentation. The present review was focused on those methodologies of attribute selection from which the topology of the data distribution may be spatially inferred by visualization rather than human expertise seeking to simplify the clustering processes and enhancing the applicability scope of the spatial approach. Aspects related to the spatial point pattern analysis were also reviewed with emphasis on the interpolation processes.

2.3.1 Representation of spatial pattern from crop-livestock systems features

Traditionally, hierarchical clustering seeks to group patterns based on similarity criteria (generally distance) looking for structures in the data whose domain might be recursively divided according to the user making the judgement (Ward, 1963; Johnson, 1967; Jain et al., 1990; Xu and Wunsch, 2005). It is certain that this approach has yielded some advances in unsupervised farm classification, and the multidimensional data has been incorporated to generate more sophisticated and realistic clustering (Köbrich et al., 2003; Domínguez, 2006). However, given that sets of similar patterns often result in quite complex taxonomic separations and that clustering performance is highly dependent on pattern representation, it would seem natural to accompany these procedures with much more expressive approaches to gain insight into the system under study, as in the studies of White et al. (2001) and Kruska et al. (2003), where a worldwide atlas of attribute variation for crop-livestock systems was created over the geographical space as a complementary tool for the global characterization of animal production systems.

Spatial phenomena involve those elements and processes of systems that are spatially manifested in a geographical position (Huber and Scheneider, 1999). Some common central instances of these elements are soil properties, vegetation and climate. The spatial evolution of these elements seems to be both cause and effect of significant processes, such as: land cover and land use; hydrologic response of catchments (Singh and Woolhiser, 2002); hydrologic effect of grazing (Fiedler et al., 2002), or probabilities of continuation of crop (Hansen and Jones, 1996; Berroteran and Zinck, 2000) and livestock systems (Thornton and Herrero, 2001; Powell et al., 2004). On the other hand, the evolution in space and time of these processes is dominated by complex interactions between discrete events such as the begining or end of the growing season, the heterogeneity of soil pedology, and the starting/finishing of any physiological or phenological stage (Ho, 1989; Rodríguez-Iturbe et al., 1999).

Recently, the role of attributes such as proportional presence in the landscape of forest and unplanted grasses, was used to make the above concepts spatially explicit through computational techniques. For example, Wijk and Rodríguez-Iturbe (2002) combined the representation of death and reproductive chances of trees and grasses to permit a dynamic description of water stress using a simple cellular automaton. The basic principle behind this technique is the use of the cell-space conceptualization to represent spatial heterogeneity (Li and Yeh, 2004). This approach has also been very useful in addressing other land surface problems (Tobler, 1979; White and Engelen, 1993).

The process of identifying farm attributes that truly represent spatial outcomes has been a problem of active research within the modeling field. The most common crop and livestock attributes used are particularly rich in time scale outcomes but require the redefinition of their indicators in terms of their ability to handle spatial scales (Huber and Scheneider, 1999; Thorne, 1998). On the other hand, the spatial heterogeneity of resources for crop and livestock interactions, has long been recognized by Fisher (1935), Preston and Leng (1987), Otte and Chilonda (2002) and Powell et al. (1993). These publications have provided a useful insight into the attributes that best describe varied tropical livestock production systems that match the wide range of existing resources. Emphasis has to be placed on the overall optimization of crop and livestock productivity from available resources, through the spatial integration of processes and technology, and by using multipurpose crops, animals, residues and by-products.

Employing much of this equivalent philosophy, Powell et al. (1993) also looked at indicators adapted to capture the spatial dimensions of nutrient cycling as attributes of the probable success and continuation of mixed farming systems. In other words, sustainability was considered as a function of nutrient cycling efficiency in crop-livestock systems. The data needed to monitor crop-livestock systems may be found in the report of Thorne (1998). He suggested that the impact of livestock mixed farming systems can be described by a minimum of six datasets, namely: feed resources, individual or aggregate animals, livestock holding, animal outputs/organic inputs to soils, and draught power. These aspects are recognized by the author as potentially spatially driven variables, when the simulation is scaled to a regional level. An application of this approach has been made in DSSAT (Beinroth et al., 1998).

At this stage, most research has focused on the development of more accurate spatial predictions for crop-livestock interactions linked to the definition and conceptualization of the spatial units where the process to be studied is mapped. The key issues in the construction of this mapping are, on one hand, the degree to which spatial units affect state variables and on the other, how the state of one variable might affect the future state of another (Hunsaker et al., 1993).

Given that distinct attributes are supposed to have different impacts on the spatial manifestation of the system under study, the focus is centered on some common configuration of the range of variables selected to represent appropriately the features of crop-livestock systems. For instance, the insertion of livestock milk production at the cow level of resolution permits farmers' responses to changes they may face in the economic and environmental settings to be summarized. Closely related to this attribute is the variable milk production per unit area (hectare), where grazing cattle productivity seems to be inversely correlated to milk production per head. Hence, when milk production per unit area tends to augment as a result of increasing the number of animals in a given grazing area, milk production per cow approaches its minimum. However, as this is a non-linear dependence after a certain level of cow yield decline, milk production per area also falls. This variable is also expected to be driven by fluctuation in growing period and stocking rate (Wilson and Macleod, 1991; Coppock, 1994).

Post-weaning management in those farms that retain weaned calves as stockers (Wegenholt, 2004) is an attribute that provides information about the proportion of male and female animals between calfhood and reproductive age in the whole herd. This indicator accounts for the near-future beef and heifer production in the farms, and it is highly driven by sanitation, and the number of young females that are intended to be used as replacements.

Herd replacement capacity can be incorporated to provide information about the proportion of mature cows (reproductive age) to the total females on the farm. The inclusion of this variable might indirectly account for the differences in response between farmers dedicated to cow-calf systems, and those inclined to cow stock-rearing beyond calfhood. It is expected that in cow-calf farming, range land and forest grazing is used to reduce feeding costs while in cow stock-rearing systems, characterized by higher revenues, annual crop residues play a more important role in the feeding system (Fitzhugh, 1978).

Land cover management might be inferred as the result of household decisions about the particular feeding system (Köbrich et al., 2003; Andersen et al., 2004); and in consequence, as a function of the proportion of a farm's area dedicated to annual crops, forest and range land. This depends mainly on the patterns of indirect management i.e. grazing or browsing described by Thorne (1998) where the organic resource-livestock interface might be represented by a non-structured distribution of feeding choices integrated by sets of crop residues, planted pasture, range land, and forest.

Grazing has been identified as one of the primary management aspects of vegetation dynamics within crop-livestock systems (Gillen et al., 2000). Variable stocking rate is a dynamic version of grazing management that dictates the intensity of plant defoliation by cattle during a finite period of time (Scarnecchia, 1985). The initial architecture of a complex above-ground vegetative community, from the organic resource-livestock interface, is gradually harvested at different levels of efficiency by cattle through grazing management (Gillen and Sims, 2002), and it is expected to set off heterogeneous spatial responses as part of complex non-linear effects on vegetation dynamics (Peters and Havstad, 2006; Rodríguez-Iturbe et al., 1999) and impacts of grazing on soil hydrology (Fiedler et al., 2002; Warren et al., 1986).

In summary, although using the above representation strategy does not guarantee meaningful clustering it is possible to increase its reliability by permitting a degree of flexibility in its formulation. This flexibility might be achieved by including additional variables according to a hierarchy of agriculture subsystems introduced by Hart (1982) and Van de Ven et al. (2003).

2.3.2 Spatial point pattern interpolation for crop-livestock attributes

Without being exhaustive, and with the idea of introducing some of the concepts alluded to the beginning of section 2.3, the process of predicting attribute values in unsampled areas is known as interpolation. In this section are included some aspects of the spatial analysis used to create continuous surfaces from geo-referenced point data, which generally involve such processes.

Fig. 2.1 displays a taxonomy of the most common types of spatial analysis. As it reveals, two main groups of analysis can be identified based on the Longley et al. (2003) classification. Query and analysis methods are of particular interest within the present research, given that the user can activate procedures using keyboard or pointing devices through which it is possible to ask simple questions of a database. These procedures are known as queries and reasoning. Other options open to the user within this set are measurements, which involve the description and summary of datasets, using numerical values determined by several algorithms; and finally transformations; where interpolations are contained here the users can alter the database contents by comparing and combining them using geometric operations (Burrough and McDonnell, 2005).

Within the spatial analysis domain two interpolation methods can be distinguished: global and local, which mainly differ in the scope for sampling interpolators. The former use all data point to produce predictions, while local methods work with small areas around the points involved in a predetermined neighborhood. Given the purposes of this research, local methods are of particular interest. Within this category, Thiessen polygons (Voronoi diagrams and their corresponding triangulation Delanuay)

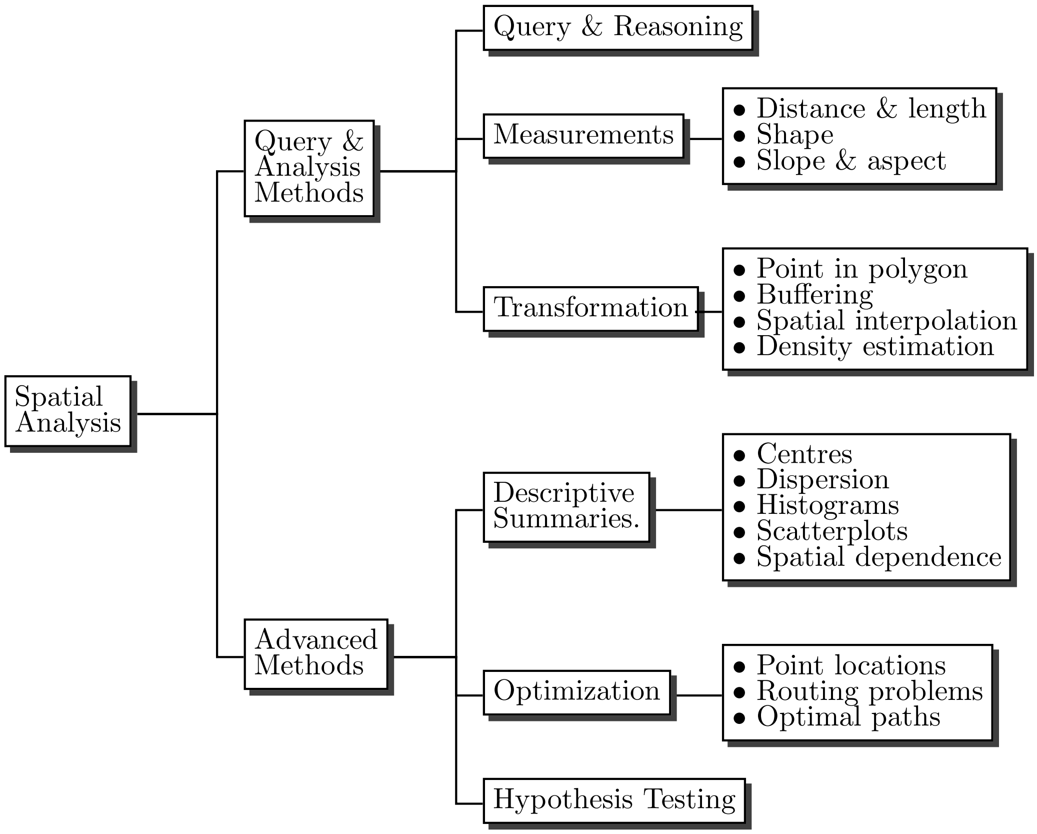


Fig. 2.1. Spatial analysis classification scheme (Longley et al., 2003)

(Gold, 1991), represent one of the most common forms of interpolation from data points; being nowadays a standard technique included in most geographical information packages (Edelsbrunner and Mücke, 1994; Burrough and McDonnell, 2005).

However, Thiessen polygons seek to interpolate point sets within homogeneous geometric expressions that coincide with the sampled point. This approach often yields abrupt approximations because continuity of attributes is ignored and changes of quantities measured effectively only occur at polygons borders. By way of contrast, the requirement of smooth variation between transitions can be achieved by inexact methods which can predict values at point locations that differ from their sampled value (Burrough and McDonnell, 2005).

Currently available inexact interpolation functions as density estimation (Fig. 2.1), might be used to fulfill the smoothness requirement. In particular kernel density functions seek to approximate a surface represented by the variation of point density events across an area (Silverman, 1986). The surface surge as a result of the intensity estimation shows the predicted distribution of an event across an approximated bi-dimensional grid that represents the landscape (Bailey and Gatrell, 1995). The main advantage of kernel density functions is that they fit the observations without any *a priori* assumptions about the generating data distribution and perform well with relatively small samples (Bowman and Azzalini, 1997; Levine, 2002).

Kernel density estimation produces a probability density surface approximating the sampling domain by a grid, or something similar, using a kernel function that depends on the distance to those points in the surface that lie within the radius of the kernel (Gatrell et al., 1996), where estimates of intensity are thus represented at each equally sized cells (Amatulli et al., 2007). This use of simple geometric structures to approximate the geographical domain is analogous to that used to represent the hydrological and similar systems where diverse representations for visualizing surface have been tested. They are more accurately defined as spatial subunits (Maidment, 1993).

Expressing kernel density estimations as spatial unit similarities through a grid seems to be particularly relevant for crop livestock systems. This is because of the complex set of attributes needing to be analyzed, the heterogeneous resource allocation and use by plants and animals and the varied distribution of water, nutrient and relief (Wade et al., 1998; McIntire et al., 1992). In summary, an interesting way to tackle spatial point analysis for the heterogeneity of crop-livestock interactions for visualisation purposes might be through the paradigm of continuous fields using kernel density functions; which constitute a discrete way to represent the variation of attributes over the space, using spatial units such as regular grids or administrative boundaries.

2.4 Feature extraction and dimensionality reduction

The main idea within a feature extraction context is to isolate those statistical characteristics of the data that portray essential elements of them, and to provide a better understanding about the underlying processes that generate the data (Guyon and Elisseeff, 2003). The traditional way to extract this information is by measuring properties in the object that integrate the pattern to be recognised; and to characterise them in terms of values that are similar for objects within the same catergory and different with respect to those located in another category (Duda et al., 2001).

One of the standard tools to perform this kind of task has been Principal Component Analysis (PCA) (Hotelling, 1933a). Using this technique, features appear as a result of a linear transformation of original attributes, leading to new variables, principal components (PC), whose covariance is set to zero (decorrelated). Other approaches (feature selection) adjust a subspace from the general pull of variables and discard those attributes that might not be of interest (Landgrebe, 2007). However, within PCA rather than discard certain attributes, the new PCs integrate the available variables into a weighted average of all the data.

This variance recounting on the PCs has been capitalised on unsupervised classification of farming systems, where feature extraction processes are not necessarily undertaken to reduce dimensionality, as is the case of Köbrich et al. (2003), who characterise farming system in Chile on the basis of just few attributes (11 variables). Domínguez (2006) uses 15 variables to represent Venezuelan populations of crop-livestock systems, and in both cases the use of PCA is permitted in order to provide measure distances for eventual unsupervised classification.

Graphical representation of the data has been one additional reason to use this technique (Drury, 2001) in an unsupervised classification context, and also the creation of domain knowledge (Guyon and Elisseeff, 2003), in the sense that eigenvectors for a set of variables can be seen as a concept described by linear combinations of feature values, which enables semantic meaning to be elicited.

Even when the few original attibutes used in this research appear not to require further reduction, the use of PCA is justified given the general improvement of the vector space representation. Such amelioration ensures that famrs containing equivalent attributive values, are mapped to similar features. This and previously mentioned properties demonstrate that even within low dimensional problems, feature extraction techniques might lead to a more flexible, expressive and compact representation of the data. A good summary of PCA application and its role within unsupervised classification can be found in Jolliffe (2002).

2.5 Data preprocessing and methods

2.5.1 Census data

The data used in this study were assembled from the sixth Venezuelan agricultural census. The census was organised by the former Ministry of Agriculture and Livestock of Venezuela, during the period between July 1997 and January 1998. The coverage included 500,959 holdings occupying an area of 30,071,192 hectares; which represented all holdings of different agricultural activities in the country; and was framed on county and parish maps from the XIII Population and Housing census. The quality parameters were delineated on the basis of a 10% sample (34 sectors); and to certificate census quality, the country was divided in 10 segments, integrated by sectors (330 sectors in total), with 15 agricultural holdings each.

The quality acceptance level of this census was 4%; in other words, this was the maximun percentage of defective farms that was tolerated within a particular sector, in order to acept the sample. A merit index was used to asses the response quality, and when such merit was superior to 6 the holding was rejected.

2.5.2 Sample size and sampling

In order to establish the sample size, for both proportion and continuous variables, the following equations were used (Casley and Kumar, 1990):

$$n^1 = \frac{NZ^2\,SD^2}{Ni^2 + Z^2\,SD^2} \tag{2.1}$$

Where (2.1) are the sample size for proportion and continuous variables respectively. N represent the population size; Z is the correspond $z - scores$ within a Gauss distribution when probability α equeal 0.05 or 0.01. SD is the espected standard deviation; and i the spected error.

Once the sample size was set (18%), a list of the crop-livestock farm population within the study area was created using the agricultural census database. After this, each farm in the list was identified with a number, and then using a random number generator (microsoft excel) the 168 corresponding farms were selected after verification that they still exist and remain involved in the same activity as reported in the agricultural census.

Table 2.1. List of attributes examined in the unsupervised classification of farms.

Symbol	*Units*	*Description*	*Source*
STR_a	AU	Stocking rate	Scarnecchia (1985)
PWM_a	%	Post weaning management	Wegenholt (2004)
HMP_a	$\mathrm{l\,ha^{-1}}$	Milk production hectare^{-1}	Wilson and Macleod (1991)
CMP_a	l	Milk production cow^{-1}	Coppock (1994); Vaccaro (1998)
HRC_a	%	Herd replacement capacity	Vaccaro (1998); Vera (1998)
SOR_b	%	Land to growing sorghum	Fitzhugh (1978); Renard (1997)
MAI_b	%	Land to growing maize	Renard (1997); Thorne (1998)
FOR_b	%	Land occupied by forest	Duvernoy (2000); Domínguez (2006)
PAF_b	%	Land dedicated to pastures	Gillen et al. (2000)
$MASL_b$	m	Meters above sea level	Comerma and Chacon (2002)

a: Variables associated to livestock productive-reproductive management
b: Variables associated to land-cover management
AU: animal unit (450kg)

Based on the feedback from random farm selection, the census database was queried on a sample of 168 households from a population of 1321 farms located in the hillside landscape of Southern Aragua (Urdaneta county) and Guarico (Monagas and

Guaribe counties) states in Venezuela. Each record of the agricultural census included as primary key an 11 digit code associated with its corresponding location at state, county, parish, sector level and farm number.

The spatial location for the 168 farms was derived by a real-time GPS receiver (Trimble 4700) and processed within a Geographical Information System in order to link geo-referenced data and digital cartography. On the basis of the literature review described above, attributive data was delineated from the census database at each sample point, and comprised 10 variables (Table 2.1).

These attributes were selected for two main reasons. Firstly they provided information on the two main farm-associated constructs: livestock productive-reproductive management, and land cover management as potential indicators of farm types. Secondly because of the posibilities that census database offered for their calculation.

Sampled data were also partitioned into two comparison groups: The Aragua-Guarico data set including the whole 168 household information corresponding to both states; and Guarico data comprising a sub-set of 103 farms from the former group 1 (Fig. 2.2). Data separation permits farm comparisons between two very different administrative units (Aragua and Guarico states).

It also enables a more detailed exploratory analysis of lower hierarchies within Guarico state, by both spatial density activity estimation and hierarchical cluster analysis. The recognition of these hierachies is important, in the sense that classes in the whole study area (group 1), are made up of class mixtures in a lower order region (group 2).

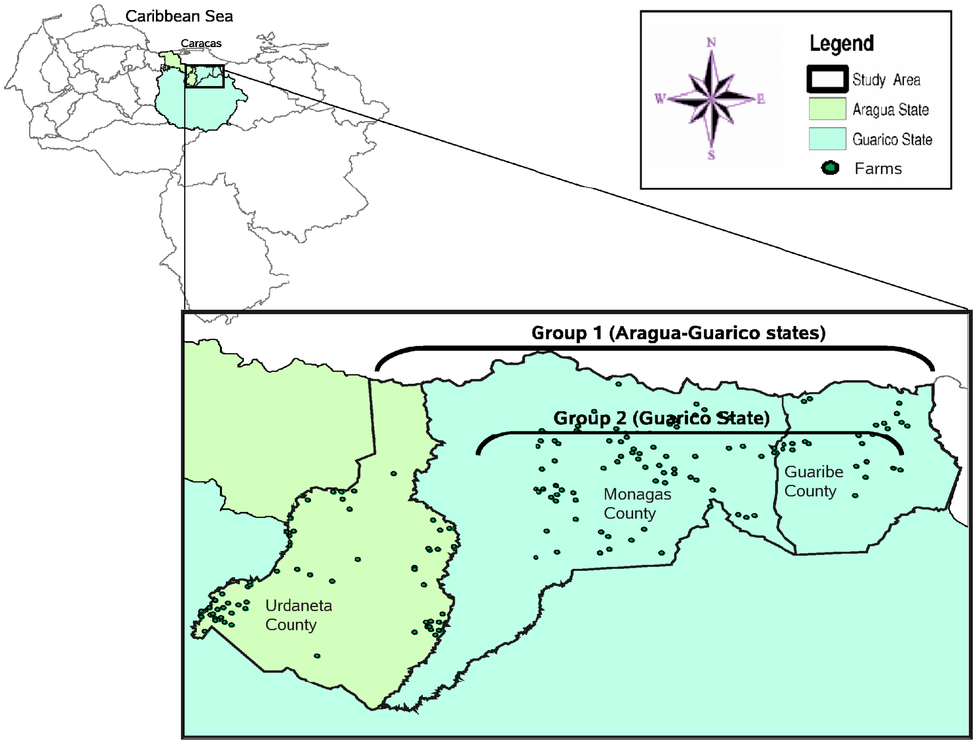


Fig. 2.2. Geographical distribution of experimental group 1 (168 farms located in Aragua and Guarico states) and 2 (a subset of group 1 integrated by 103 farms located only in Guarico state).

2.5.3 Spatial activity density estimation

In order to explore how farm attributes are concentrated in geographical space, an estimated density surface of each variable observed in the farms data set was built. The algorithm to generate such density surfaces from point activity distribution according to (Bailey and Gatrell, 1995) is formalized as:

$$\lambda_h(x) = \frac{1}{\delta_h(x)} \sum_{i=1}^{n} \frac{w}{h^2} k\left(\frac{x - x_1}{h}\right), \qquad x \in \mathbb{R} \tag{2.2}$$

where $\mathbb{R}$ represents the area under study, x a location on $\mathbb{R}$, $h > 0$ is a band accounting for the smoothing; w is weight vector, which represents the value of the attribute to be visualized; $\delta(x)$ is a correction factor; $x_1, x_2, ...x_n$ feature activities locations; and k a quadratic kernel function defined by Silverman (1986):

$$k(x) = \begin{cases} 3\pi^{-1}(1 - x^T x)^2 & \text{if } x^T x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

To compute the density surfaces the Spatial Analyst extension of the software ArcMap (ESRI, 2004) was used; and the results were rendered into a 3D format using the software ArcScene (ESRI, 2004). Cardinality of the band accounting for the smoothing was established according to the procedures described by Levine (2004):

$$\text{Mean}(h) = \sqrt{\frac{N(p) * A}{N * \pi}} \tag{2.4}$$

where N is the events sample size located in a region of area A; and $N(p)$ represents the number of obsevations to be considered within the kernel radius of search. To compute this number of observations, a $k - nearest\,neighbour$ analysis was used to calculate the number of centroids around which data is centered; following this, each centroid is checked for the number of farms aroud it; then the smaller cluster of farms around one of the centroids within the area of interest, is set as the maximum number of obsevations to be included in the kernel radious (Moreno, 1991; Guerra, 2004).

2.5.4 Statistical analysis

Histograms and their parameters (mean and standard deviation) were generated to accompany density surfaces for variables grouped by comparison sets (Aragua-Guarico and Guarico). Data sets were preprocessed by normalizing their variances; then in order to extract the systematic variation and to reduce dimensionality of the data matrix, a Principal Component Analysis (PCA) was carried out.

The first six components generated within the PCA provided the input obsevations required to perform a hierarchical clustering using the methodology described by Ward (1963); to confirm the farm sets visually observed in the surface density analysis. Both PCA and hierarchical clustering were implemented with the software CSTAT (CIRAD, 1989). Then farm types were determined following the methodology and descriptive statistics of Berdegue and Escobar (1990) using the SPSS software (SPSS-Inc, 1999). Finally a linear discriminant analysis was applied in the interest of determining the original variables that were helpful on separating groups. The analysis was carried out through the 7M routine of the sofware BMDP (Dixon et al., 1981).

2.6 Results and discussion

The methodologies were applied to the two comparison sets (Guarico-Aragua and Guarico) which varied in complexity. The objective was on the one hand to characterize the illustrative capacity of feature intensity representation in terms of the number of typologies to be defined from the data and on the other, to assess quantitatively the quality of the clusters based on *a priori* information. These correspond to a numeric measurement from PCA and cluster analysis but are more qualitative for those experimental observations where the criterion is based on the visual interpretation. Firstly the output corresponding to Aragua-Guarico set (168 farms) is shown, then results for the Guarico set are included (103 farms).

2.6.1 Aragua-Guarico set

2.6.1.1 Multivariate density visualisation

Initially, to gain insight into the spatial distribution of the whole data set farms of both states were studied and treated as if they came from a single entity. This experimental level explores how the spatial intensity of productive-reproductive management and land cover occupy the geographical space using real data. The most relevant aspect to study is the use of density estimates to define cluster structures that might help in dividing the farm population into a number of classes.

The study area for this comparison group was covered by a grid estructure of 512 columns and 252 rows. Each output cell showed a size of 265.44 m on X and Y axes and the search radius was 30 km. This value results from the selection of 59 farms as the number of observations to be considered within the kernel radius; a decision that was made on the grounds of $k-nearestneigbour$ analysis; which based on distances between observations, indicated that farms for Aragua-Guarico group were clustered around three centroids and the smaller cluster integrated 59 farms. Hence, the kernel function sequentially scanned the grid area summarizing the number of events within the search threshold until all the surface was completely covered and produced a map of smoothed intensity estimation. Fig. 2.3 and 2.4 show kernel density estimates for productive and land cover management respectively.

It can be appreciated from Fig. 2.3 (a) that the highest densities for stocking rate activity lie in the western end of the study area which corresponds to Aragua farms. A visual comparison of this surface density depicts that there is a reduction of intensity

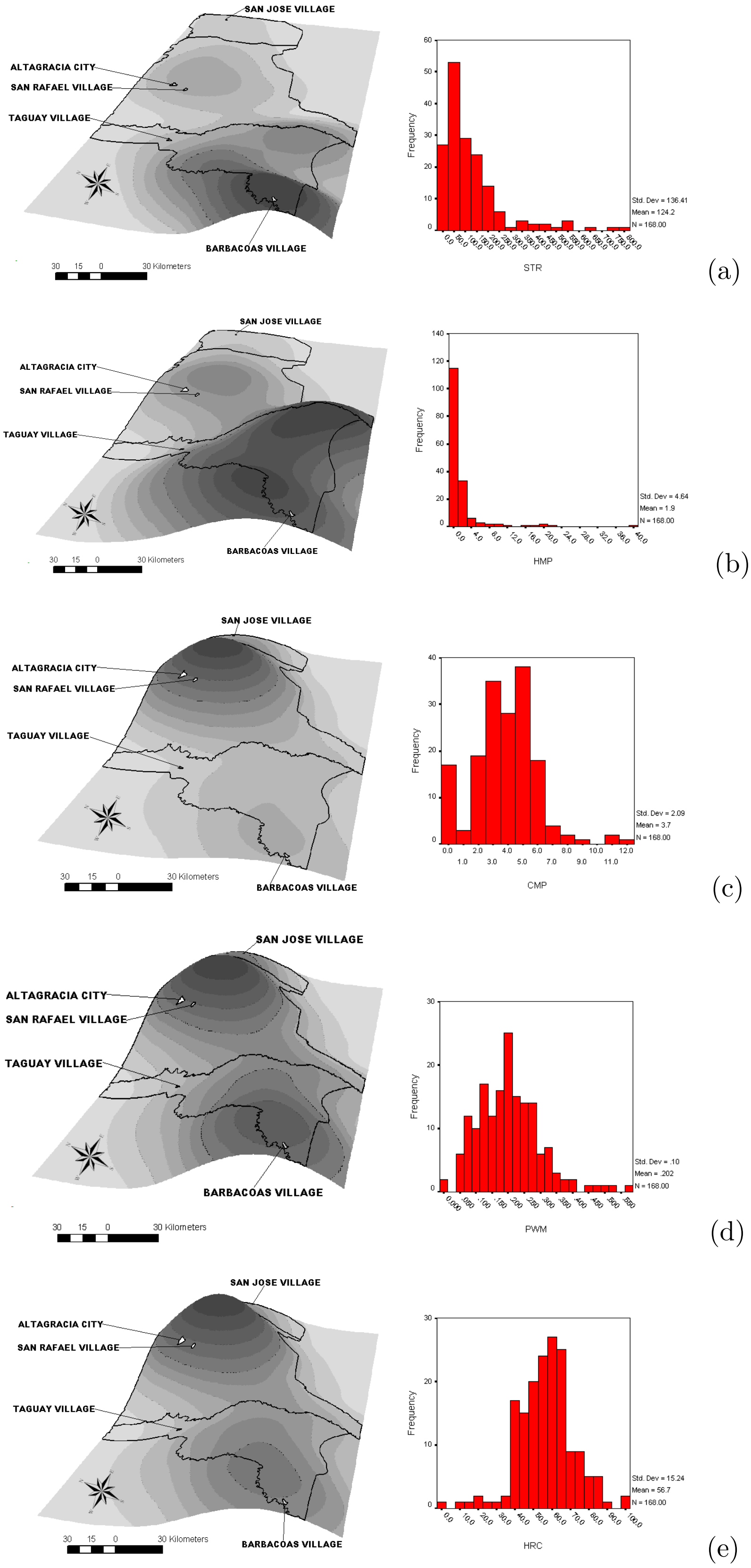


Fig. 2.3. Kernel density estimates of intensity for productive-reproductive attributes of 168 crop-livestock farms, search radius 30 km.

and continuity towards the eastern end of the area, turning it into small density focus organized around villages that act as municipality centroids. If milk production per hectare is considered (Fig. 2.3 b) a similar pattern can be seen, while by way of contrast milk production per cow had the opposite density estimates (Fig. 2.3 c) showing rings of high density activity for this variable within Guarico farms (Urdaneta et al., 1999; Betancourt et al., 2005).

This activity density analysis indicated that the intensity of farms attributes relating to milk production, seems to be occupying a different spatial position within the sampling domain. These results are consistent with earlier findings of Coppock (1994), who reported a configuration of milk production per hectare with an underlying production per cow that tended to be inversely associated in tropical semi-arid rainfed conditions. Similarly, Páez and Jiménez (2000); Páez et al. (2003) and Urdaneta et al. (2004) encountered the same responses under humid-subhumid rainfed and irrigated conditions in grazing livestock in Venezuela. Additionally, it is interesting to note that stocking rate was highest in those farms with more milk production per area, which is consistent with the views of Gillen and Sims (2002). Animal response changes in terms of its productivity, and there are some attributes, such as production per area, that might be positively influenced by stocking rate as a consequence of a non linear relationships (Wilson and Macleod, 1991; Van de Ven et al., 2003).

Fig.2.3 d and e correspond to post-weaning management (PWM) and herd replacement capacity (HRC) respectively. It can be observed, that the breaks between neighboring counties are diluted by smooth progressive density rings; the modes of the distribution are spatially polarized among the Aragua and Guárico area as the pattern

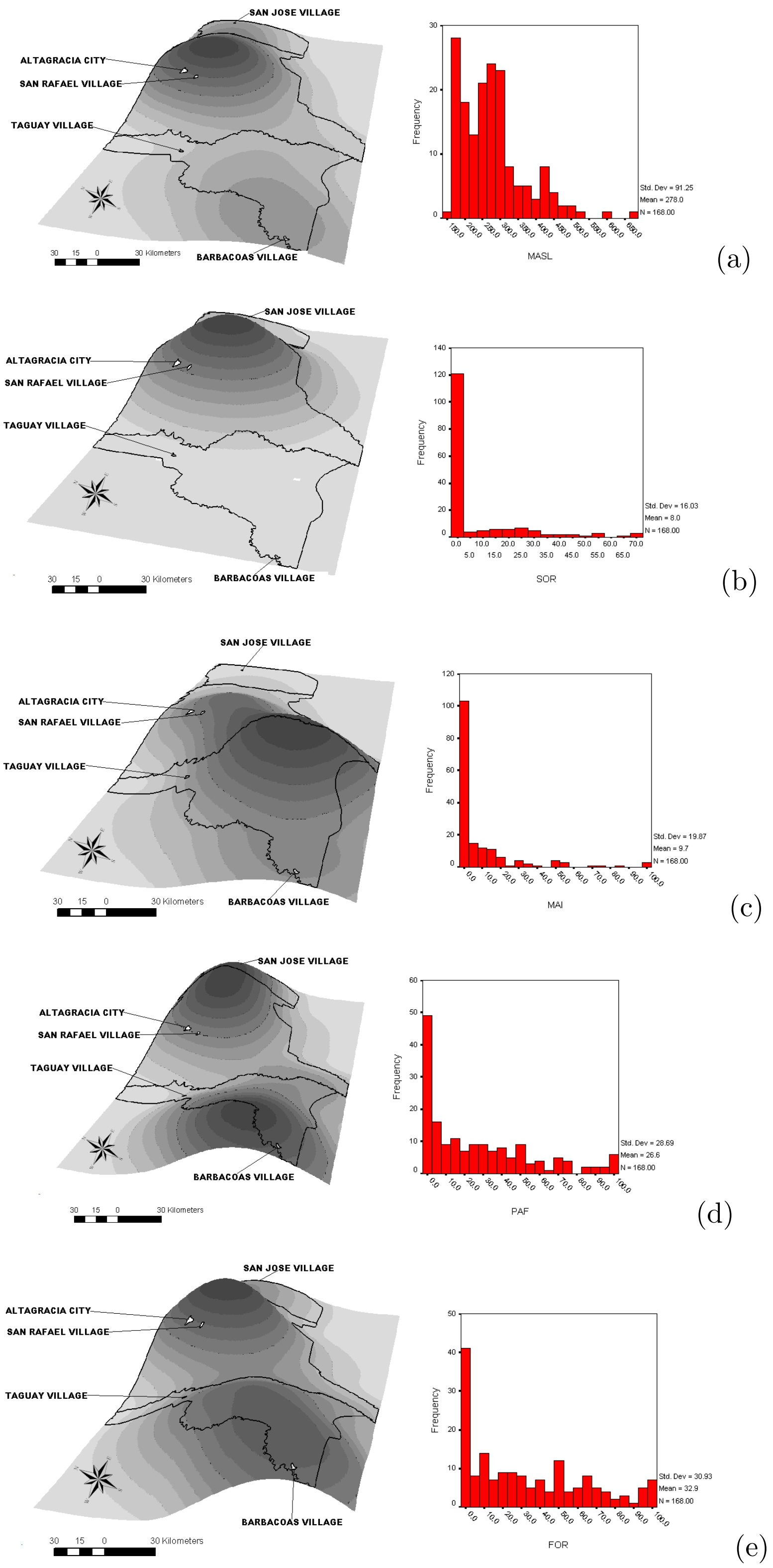


Fig. 2.4. Kernel density estimates of intensity for land cover attributes of 168 crop-livestock farms, search radius 30 km.

gets more complex. It is worth noting that farms showing the highest density for PWM activity were concentrated in the west end of the study area which corresponds to Urdaneta county of Aragua state. Production systems in this area have been associated with a high propotion of growing and feeding systems where young animals are raised or purchased after calfhood to be fattened and then slaughtered in the near future (Fusagri, 2001; Mosquera, 2005). On the other hand, those farms having low PWM activity were clustered towards Guarico's counties which have been characterized by the high frequency of cow-calf systems (Espinoza et al., 2005; Domínguez, 2006).

Density surfaces of farms' land cover attributes are shown in Fig. 2.4. The elevation density of farms within the sample space is depicted in Fig. 2.4 a. In the topographic characterization, farm values range from 153 to 680 meters above sea level showing a density distribution with more than one local maximum. As can be gathered from visual comparisons, the lowest altitude occurs to the western end of the study area in Aragua state, while Guarico farms appear to be mainly concentrated at higher altitudes.

Generally, there is a bimodal distribution across the spatial domain for the majority of variables, with sorghum and maize cover (Fig. 2.4 b and c) being the attributes that show a unimodal distribution of density, and that sorghum is only cultivated in Guarico's counties. The next three surfaces correspond to maize, forage and forest covers (Fig. 2.4 c, d and e); which represent the basal livestock feeding resources for most farmers across the study site. As can be seen, density of annual crop activities does not occupy the same geographic domain. Most farms with high density for MAI activity appear concentrated in the center of the study area, between Aragua and Guarico states while farms that showed high values for SOR are grouped towards the east. This might be indicative of a crop substitution strategy (Herrero et al., 2007a), where sorghum replaces maize to cope with shorter growing periods (270-299 vs 210-239 days) in the eastern end of the study area (Fig. 2.5) (Rodríguez and González, 2001).

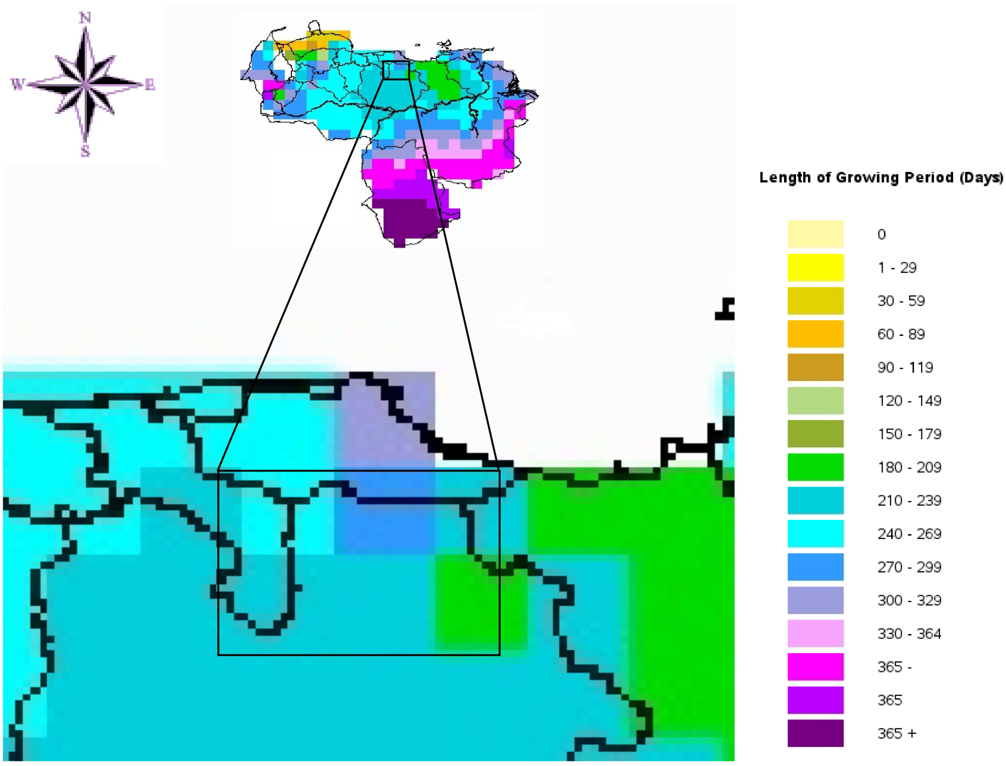


Fig. 2.5. Length of growing period (Fischer et al., 2000).

Pasture and forage density (Fig. 2.4 d) appear polarized to the west and east extremes of the sampling area, coinciding with the lowest values (210-239 days) for lenght of growing period within the study area (Ewell and Madriz, 1978; Rodríguez and González, 2001). A similar pattern was observed on the variable FOR, although farms with a high proportion of semi-deciduous seasonal forest tend to be concentrated towards the areas with growing periods of 240-299 days. These results resemble those of Domínguez (2006), who additionally found that due to their tendency to rely more on grassland and secondary forest than annual crops, cow-calf farms include a higher proportion of forest and also utilize more land in total than growing and feeding systems. These aspects of farm orientation (cow-calf/feeding) require specific attention, given the increasing interest of the public in the enviromental consequences of agricultural expansion (Steinfeld et al., 1997), particularly those associated with deforestation and extensive grazing systems (Nicholson et al., 1999; Fisher and Thomas, 2004), the commanding role of market-labor relationships (Herrero et al., 2007b) and the probabilities of sustainability of different productive modalities for meeting future human needs for animal products (Delgado et al., 1999; Bruinsma, 2003; Ortega et al., 2004; Bouwman et al., 2005).

The overall pattern indicates that most attributes showed two modes, suggesting the presence of two farm subgroups occupying different geographical spaces (Silverman, 1986). In this case, cow-calf and growing-feeding systems seem to be the two main derivations after examining previous published studies (Domínguez, 2006), and attribute density surfaces. The use of this data visualisation method resulted in a very effective description of activity distribution across Aragua-Guarico.

2.6.1.2 Exploratory cluster analysis

As part of the previous step before proceeding with the cluster analysis of a population; a dimensionality reduction is performed in order to eliminate much of the noise present in the data. One of the standard techniques to undertake such reduction of dimensionality is principal component analysis (PCA) which basically attempts to reorganize original data into a hierarchy of linear combinations based on their variance. The usual procedure involves the analysis of the variance-covariance matrix to explore similarities of attributes' pattern differences derived from a pre-selected central tendency measure (normally the mean).

---Correlation Matrix---

	2 STR	3 HRC	5 PWM	6 CMP	7 HMP	8 MAI	9 SOR	10 PAF	11 FOR
2	1.000								
3	-0.052	1.000							
5	0.108	-0.393	1.000						
6	0.077	0.196	-0.054	1.000					
7	0.002	0.028	0.090	-0.025	1.000				
8	-0.044	-0.102	-0.038	-0.111	0.361	1.000			
9	0.368	0.063	0.073	0.190	-0.117	-0.071	1.000		
10	-0.017	-0.016	-0.054	-0.054	0.091	-0.251	-0.123	1.000	
11	-0.140	-0.039	0.101	-0.143	-0.254	-0.155	-0.190	-0.495	1.000
12	0.165	0.163	-0.129	0.438	-0.218	-0.228	0.200	-0.090	-0.030

Table 2.2. Correlation matrix for the ten variables examined from the Aragua-Guarico dataset.

A standarized version of the covariance matrix for the 10 variables used in this study is shown in Table 2.2. This represents the degree of association between pairs of dimensions in the form of a correlation matrix. As can be seen, each variable shows a similar or related pattern of differences for at least one of the other variables within the group, meaning that variables are not perfectly independent from one another. This is a basic action to be carried out before looking for clusters of variables that measure similar underlying constructs (Tabachnick, 2001). These results were confirmed by the Bartlett test of sphericity showing that most off-diagonal values in the matrix were significantly different from zero ($p < .0001$).

Additionally, for testing multicolineality the determinant of this matrix was calculated (=0.178), indicating that variables involved in this study were not perfectly correlated with each other (Pedhazur and Schmelkin, 1991). The magnitudes of the linear components estimated from this dataset are provided in Table 2.3. Conceptually, they represent an indicator of the weight of each component within the model and are hierarchically ordered according to the amount of variance they explain, as can be observed on the third column of this table.

---Eigen Values---

	Eigen Value	%	% Acumul	Histogram
1	1.983	19.83	19.83	**
2	1.654	16.54	36.36	***************************************
3	1.488	14.88	51.25	***********************************
4	1.315	13.15	64.40	*******************************
5	0.941	9.41	73.81	**********************
6	0.752	7.52	81.33	******************
7	0.669	6.69	88.02	****************
8	0.490	4.90	92.92	************
9	0.432	4.32	97.24	***********
10	0.276	2.76	100.00	*******
TOTAL	10.000			

Table 2.3. Histogram of components eigenvalues.

According to the criteria of Kaiser (1960), a substantial amount of variability may be explained after extracting only the four components whose eigenvalues are greater or equal to 1 (Table 2.3). However, given the sample size of this experimental group (n=168) and the resulting average communality (0.6) the criterion of Jolliffe (1972, 2002) was used. Inclusion of six components made possible the explanation of slightly more than 80% of the accumulated variance with this model.

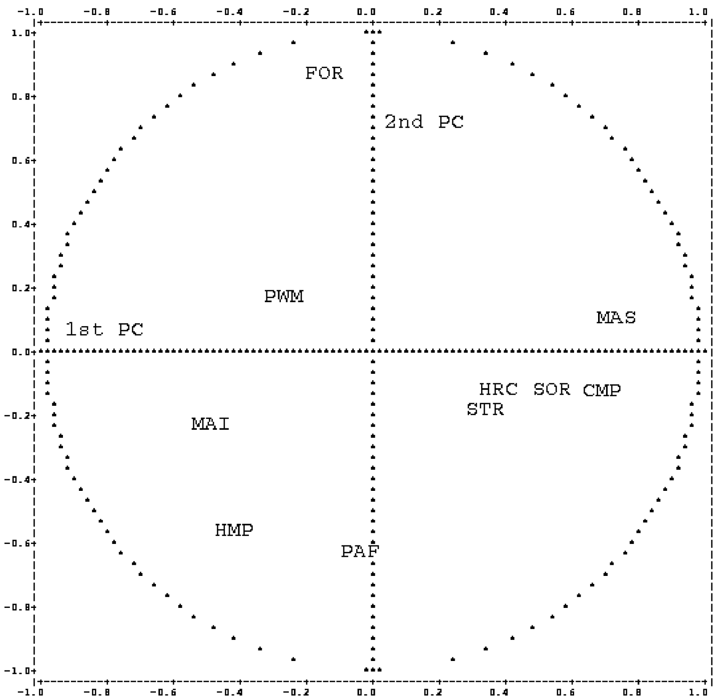


Fig. 2.6. PCA loadings plot of the first two principal components of the productive and land cover attributes data set.

A geometric perspective of the relationship between components and the original variables is presented in Fig. 2.6. The position of the original variables on the plane formed by both axes represents its relationship with each component. In consequence, variables that form clusters in this plane are hypothesized to measure the same underlying construct (Bray and Maxwell, 1985).

Given that in these cases an exact mathematical expression for the loadings of each variable onto components is available (cosine of the angle between former variable positions and new components) the contribution of original dimensions can be inferred. In Fig. 2.7 a map of loadings for the 10 original variables is plotted on the three principal components. As revealed, an emerging pattern of the variable set can be appreciated resembling the common variance reflected in these three components in the form of two pontential underlying constructs.

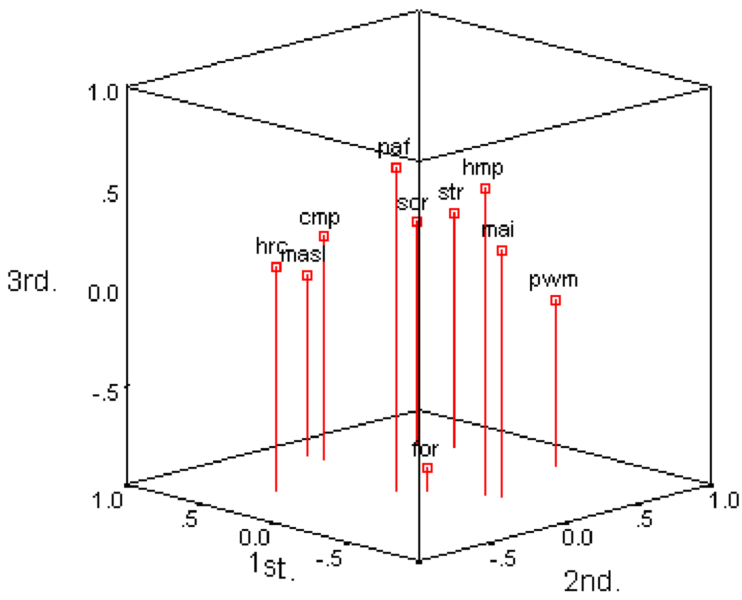


Fig. 2.7. Scatterplot of variables loadings for the first three principal components.

If a variable set form a cluster on the plane depicted among these three axes (Fig. 2.7), they are probably contributing different information to a common dimension behind them (Morrison, 2005). Hence, the model would appear to suggest that apart from the variable forest (FOR), which is inversely correlated with pasture and forage (PAF) in a significative way (Table 2.2) the remaining variables are distributed within the volume enclosed by these components, showing a pattern of differences that resembles the constructs used to guide the information gathering on farms. This may indicate that assumptions made on attribute selection from census data were right, given that variables selected actually measured the underlying expected dimension (Stevens, 2002).

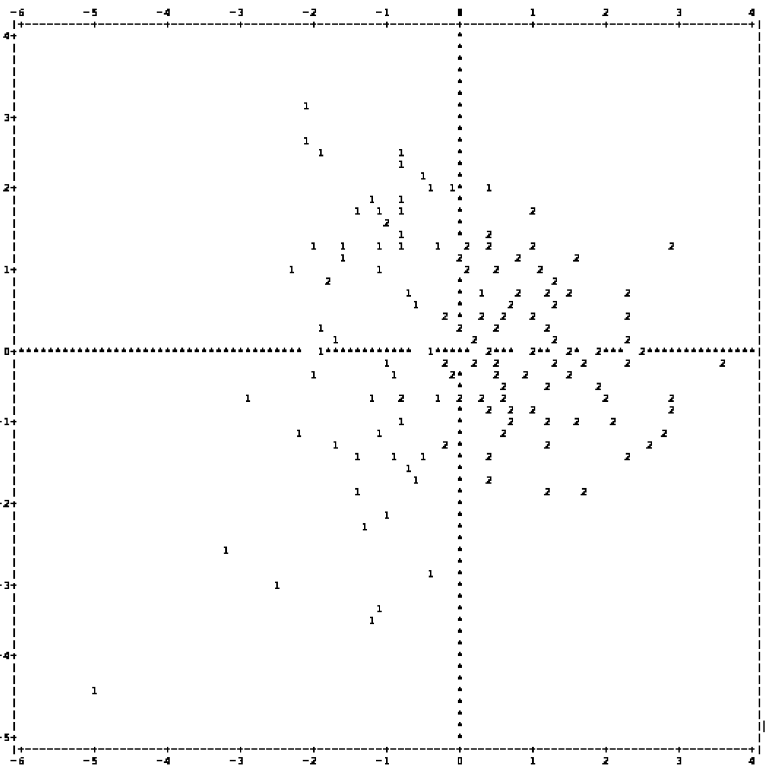


Fig. 2.8. PCA scores plot of farm's classes for the first two principal components of Aragua-Guarico group dataset.

As was pointed out in section 2.5.4; after having original data represented in terms of the new principal directions, the next step was to proceed with the classification process. It is worth noting that the decision about the number of clusters into which the data were segmented, was made on the basis of the overal spatial pattern observed on activity density surfaces depicted in the previous section. Some results of this clustering process are provided in the form of farms' composite scores by class projected on the first and second components in Fig. 2.8. As can be seen, the Aragua-Guarico experimental set, group 1, described in section 2.5 was segmented into two farm classes (farms type 1 and 2) using a hierarchical clustering method. If individual coordinate scores along the components are considered, it can be observed that most type 1 farms are clustered to the top and bottom left-hand of the plot, largely representing Aragua farms, and type 2 farms on the opposed side of the plot including most Guarico farms.

2.6.1.3 Summary of farm's classes (Aragua-Guarico set)

By way of analogy; when variable loadings (Fig. 2.6) are compared with farm scores projections (Fig. 2.8), even when both farm classes share a comon boundary, it can be appreciated that original variables such as stocking rate (STR), cow milk production (CMP), sorghum (SOR), herd replacement capacity (HRC) and altitude (MASL) clearly appear to have influenced Guarico farms to cluster together. Unlike the Guarico set, Aragua farms' segmentation seems to be governed by showing high values on the original variables forest (FOR), post weaning management (PWM), maize (MAI), pasture and forage (PAF) and milk production per hectare (HMP).

Farm class 1 (growing-feeding systems)

These results seem to confirm what was observed from density surfaces in Fig. 2.3 and 2.4, where two farm types could be inferred (cow-calf and growing-feeding livestock systems). In this analysis, with few exceptions two group of farms appear linearly separated. These reflect not only the spatial distribution of farms but also the effect of attributes that are related to these farm types. According to the attribute structure observed, class 1 farms correspond to mixed rainfed systems (Seré and Steinfeld, 1996) that, apart from dairy production, raise or purchase weaned calves which are fattened until they reach a suitable weight to be slaughtered. The livestock feeding systems in these farms rely basically on annual crops, mainly maize although sown grass is often encountered on these holdings.

The use of forest ecosystems as grazing areas is restricted in this system to non-milking pregnant cows, replacement heifers and growing weaned females. Milking cattle are normaly crossbred (*Bos taurus* x *Bos indicus*) types although pure Zebu (Guzerat, Nelore or Brahman breds) are often used within the milking herd.

Farms class 2 (cow-calf systems)

As in growing-feeding, cattle in class 2 farms (cow-calf system) may be crossed *Bos taurus* x *Bos indicus* breeds for milking cows, but like class 1, these farm types also could present whole herds composed only of pure Zebu (*Bos indicus*) cows. In this typology, the feeding systems rely mainly on non-planted grassland and forest ecosystems. Annual crops might be included within the land cover, occupying significant proportions of the farm surface where sorghum plays an important role, as a high density biomass forage. Maize can also be cultivated as a commercial crop, but it is far less important than sorghum. Semi-deciduous forest can be very important in this typology as a grazing area, particularly during the dry season when forest litter is nutritionally rich and grass supply is generally not abundant.

Fig. 2.9 shows boxplots about the distribution of variables scores on each class; it can be appreciated that apart from variables FOR and PAF, outliers are common between the remaining dimensions. Also from these data differences in spread and centering seem

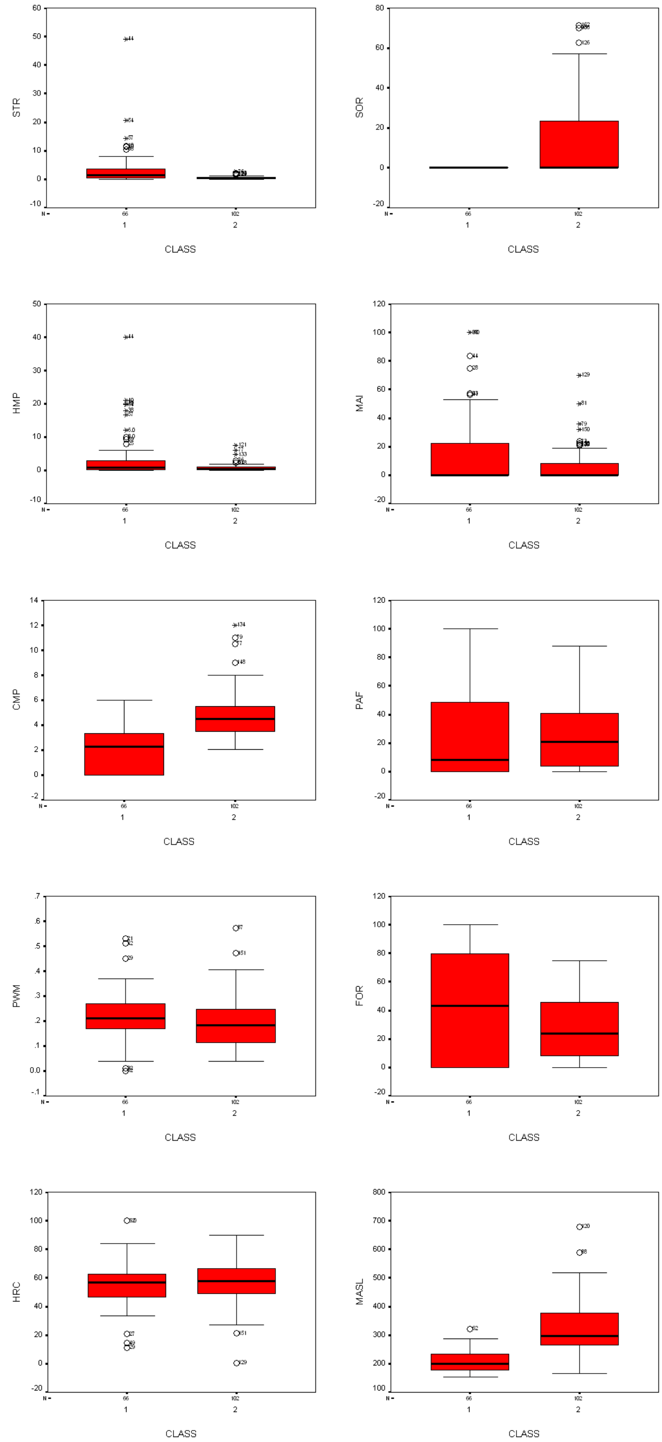


Fig. 2.9. Boxplots of productive and land cover attributes variations.

to be quite obvious between classes, reinforcing the ideas that were drawn from density surfaces.

The sequence of boxplots reveals that for most original variables (6 from 10) there is a variation trend between classes. Major variations seem to be more common on those dimensions associated to land cover-feeding system (SOR, MAI, PAF, FOR); and some variables connected to productive management such as CMP, HMP and STR.

Farm metrics estimated from census data are shown in Table 2.4. This summary is organized by farm class and provides an overview of several central tendency and disperssion measurements for each variable involved. As can be seen, the information about skewness and kurtosis sugests that no variables showed a normal distribution. However, given that generalisation of results beyond the sampled data was not one of the objectives of this study, limitations because of lack of normality could be ignored. Nevertheless, additional central tendency measures such as mode, median and trimmed mean, have been included which are much more informative when data do not show a Gaussian distribution.

Table 2.4. Summary statistics of 168 crop-livestock systems attributes by farm's class.

			STR		HRC		PWM		CMP		HMP		MAI		SOR		PAF		FOR		MASL	
CLASS			Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E
1	Mean		3.4362	.84579	55.4848	1.92011	.2194	.01228	2.2809	.21556	3.5142	.86132	16.0845	3.39488			27.8136	4.35141	42.2630	4.91589	208.461	4.81350
	95% Confidence Interval for Mean	Lower Bound	1.7471		51.6501		.1949		1.8504		1.7941		9.3045				19.1233		32.4453		198.848	
		Upper Bound	5.1254		59.3196		.2439		2.7114		5.2344		22.8646				36.5040		52.0807		218.074	
	5% Trimmed Mean		2.3421		55.5387		.2157		2.2177		2.4412		12.4004				25.3485		41.4034		206.262	
	Median		1.3200		56.5000		.2100		2.2500		.7800		.0000				8.2150		43.1650		198.800	
	Variance		47.213		243.331		.010		3.067		48.964		760.664				1249.70		1594.96		1529.21	
	Std. Deviation		6.87120		15.5991		.09974		1.75121		6.99742		27.5801				35.3511		39.9369		39.1051	
	Minimum		.00		11.00		.00		.00		.00		.00				.00		.00		153.61	
	Maximum		49.09		100.0		.53		6.00		40.0		100.0				100.0		100.0		321.46	
	Range		49.09		89.00		.53		6.00		40.0		100.0				100.0		100.0		167.85	
	Interquartile Range		3.3025		16.5000		.1050		3.3550		2.9225		23.6100				49.0025		80.6150		57.4425	
	Skewness		4.996	.295	-.019	.295	.618	.295	.230	.295	3.139	.295	1.852	.295			1.031	.295	.164	.295	.852	.295
	Kurtosis		30.605	.582	2.085	.582	1.701	.582	-.799	.582	11.611	.582	2.569	.582			-.368	.582	-1.660	.582	-.154	.582
2	Mean		.5192	.04604	57.5588	1.48999	.1904	.00959	4.6462	.17145	.8343	.10827	5.5905	1.08408	13.12	1.870	25.8268	2.33339	26.7882	2.12543	322.946	8.62172
	95% Confidence Interval for Mean	Lower Bound	.4279		54.6031		.1714		4.3061		.6195		3.4400		9.41		21.1979		22.5720		305.843	
		Upper Bound	.6106		60.5146		.2094		4.9863		1.0491		7.7410		16.83		30.4556		31.0045		340.049	
	5% Trimmed Mean		.4610		57.9630		.1840		4.4737		.6639		3.8864		10.91		24.1980		25.8355		317.376	
	Median		.3600		58.0000		.1850		4.5000		.4800		.0000		.00		20.7150		23.5400		296.250	
	Variance		.216		226.447		.009		2.998		1.196		119.874		356.501		555.361		460.778		7582.08	
	Std. Deviation		.46501		15.0482		.09688		1.73160		1.09348		10.9487		18.881		23.5661		21.4657		87.0751	
	Minimum		.04		.00		.04		2.07		.06		.00		0		.00		.00		166.14	
	Maximum		2.52		90.00		.57		12.0		7.50		70.00		*		88.00		75.00		679.19	
	Range		2.48		90.00		.53		9.93		7.44		70.00		*		88.00		75.00		513.05	
	Interquartile Range		.4775		18.0000		.1400		2.0000		.7125		8.0825		23.67		37.5025		37.7550		112.702	
	Skewness		2.047	.239	-.481	.239	.990	.239	1.707	.239	3.938	.239	3.241	.239	1.495	.239	.778	.239	.518	.239	1.279	.239
	Kurtosis		4.842	.474	1.255	.474	1.620	.474	4.740	.474	18.907	.474	13.681	.474	1.480	.474	-.261	.474	-.854	.474	2.320	.474

a. SOR is constant when CLASS = 1. It has been omitted.

2.6.2 Guarico set

2.6.2.1 Multivariate density visualisation

This level of experimentation explores the complexity of a more compact farm sample generated from real data. The most important features to study are the variables that are responsible for the farm's grouping and how these variables are correlated given a finite number of sets.

As stated in section 2.5, the data used consisted of a 103 farm subset from the 168 used in the previous experiment. This group of farms represents a mixed livestock oriented rainfed agricultural system, located in tropical lowland with a humid/sub-humid growing period (240-299 days; Fig. 2.5) (Seré and Steinfeld, 1996; Fischer et al., 2000; Kruska et al., 2003). Typically these systems exhibit a mix of European, Criollo and Zebu breeds converging on a configuration of dual purpose for milk and meat production.

Establishing the productive ambit of these systems is particularly challenging for these Venezuelan stakeholders given the difficulties of segmenting farms that are generally nested into a much more complex arrangement of holdings, making it difficult to draw spatial boundaries for each group as could be seen in the Aragua-Guarico analysis. The major complication for getting practical application from understanding these systems, is the complex nonlinear dynamics that govern agricultural processes; in particular, the proportionality between input and output.

In oder to provide a context to make comparisons within this farm subset a mask grid of 422 colums and 252 rows was created to cover the sampling area. The resulting cell size was 169 m for axes X and Y respectively, and the search radius used was 14 km.

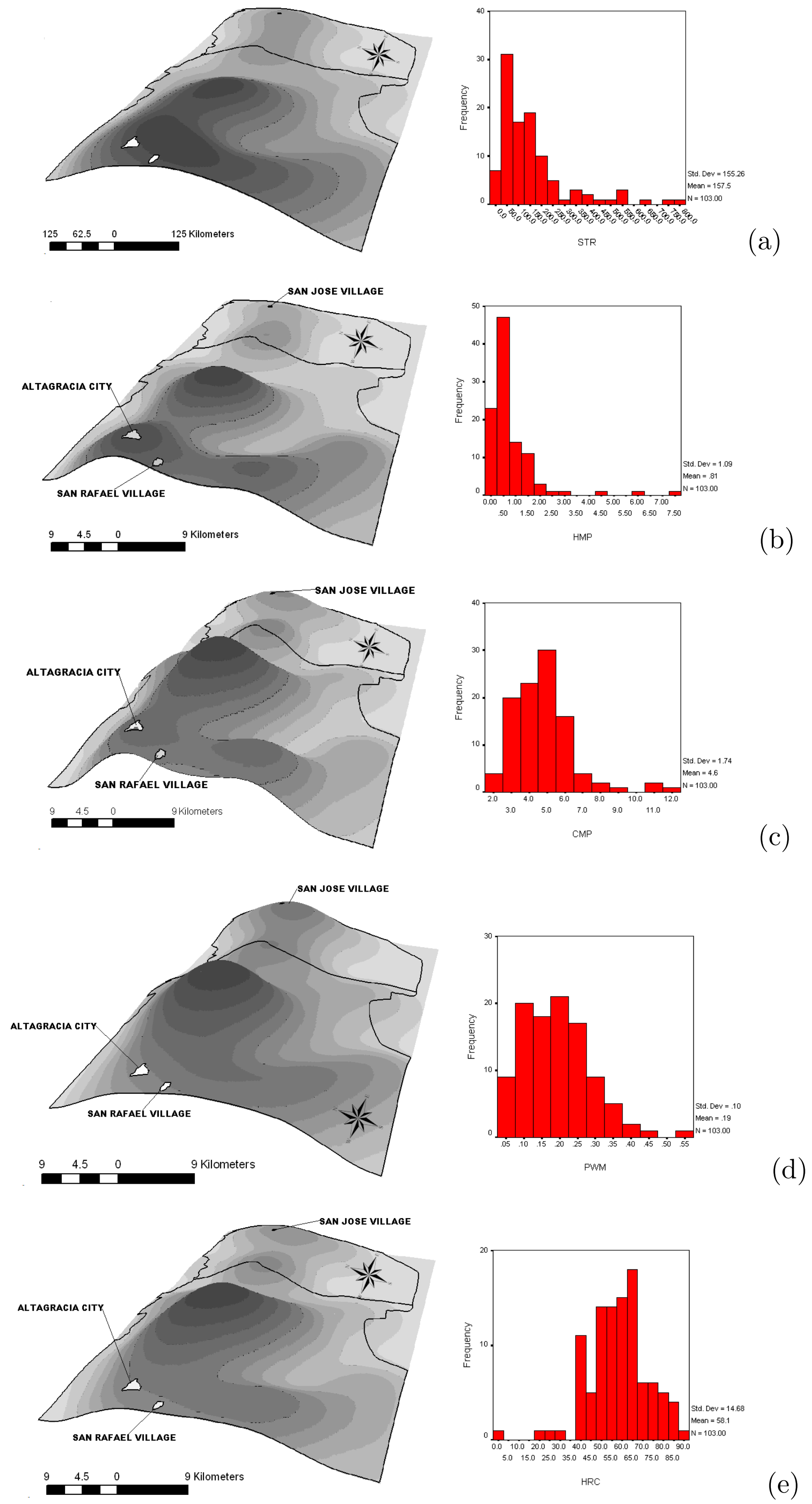


Fig. 2.10. Kernel density estimates of intensity for productive-reproductive attributes of 103 crop-livestock farms, search radius 14 km.

As in theAragua-Guarico group, this value results from the selection of 107 farms as the maximum number of obsevations to be considered within the kernel radius; this decision that was made on the grounds of $k - nearestneigbour$ analysis; which based on distances between observations, indicated that farms for the Guarico group were clustered around one centroid and the smaller cluster can be integrated by 107 farms within the study area.

Activity density surfaces for productive and reproductive management are gathered in Fig. 2.10. It can be noticed that the gradient of density appears to be concentrated in the western end of the analysis area for all productive variables, while the eastern part of the area shows lower density for all these activities (Fig. 2.10). Overall, apart from the HMP and CMP variables, all productive dimensions are characterised by three modes which were not evident from the histogram. The configuration of this density pattern shows that the taxonomy of Guarico farms is much more richer than what was observed in the former level of experimentation with group 1 (section 2.6.1). Previously described cow-calf systems shows that at this scale of resolution, a new subset may emerge, revealing for instance, the existence of an important growing-feeding activity given the high density observed for PWM. These results confirm earlier findings of Domínguez (2006) who described groups for this area characterized by keeping young animals for fattening.

Fig. 2.11 shows the result of density activities for the five land cover management variables. In the figure the density of SOR was higher in central and western parts of Monagas county clustered around two modes, while a third mode was clearly separated towards the eastern end of the study area in San Jose county (Fig. 2.11a). In the

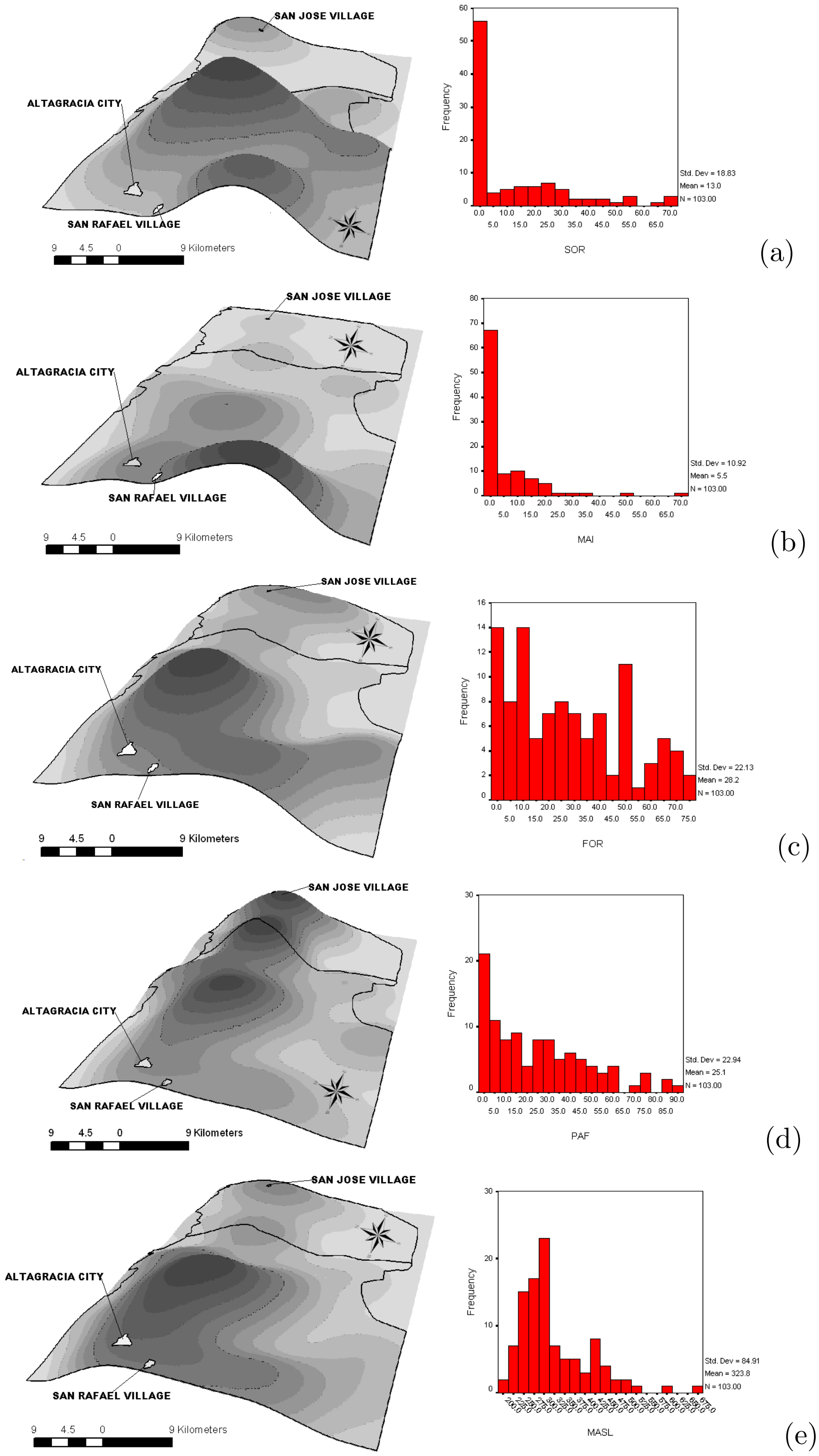


Fig. 2.11. Kernel density estimates of intensity for land cover attributes of 103 crop-livestock farms, search radius 14 km.

spatial pattern of the variable MAI (Fig. 2.11b), farms appear clustered within a high density focus in the west that gradually grades to lower density modes towards the east without any evidence of polarization between county centroids. It is noteworthy that SOR activity density modes are not spatially concentrated in any particular subregion of the field, but spread out along both counties; where growing periods range from 210 to 299 days in length (Fischer et al., 2000). This suggests that sorghum growth could form part of some agricultural management strategy in response to mid-season drought and lack of market access to alternative commercial crops like maize (Nhemachena and Hassan, 2007).

It can be appreciated from Fig. 2.11 c that the forest proportion (FOR) within farms describes a similar pattern to cash crops focusing intensity in Monagas county; and farms were also clustered around three modes. Conversely, pasture and forages (PAF) participation within farm land cover shows an opposite pattern, with systems of density rings concentrated in the eastern end of the sample area. On the other hand, elevation density (MASL) showed a similar pattern to cash crops and forest, with a main mode spatially located in the western of the area, and a couple of minor modes focused to the east region. Two points are worth mentioning: first, in the study location altitude density coincides with the longest growing period (270-299 days) of the region which also overlaps density distribution for annual crops and forest. This attribute configuration leads to an agro forestry-farming system interface that is probably permitting the segmentation of those farms that have managed to mitigate unfavourable climate and enhance farms' sustainability through a particular combination of the land cover choice (FAO, 2007). The other issue is the close relationship between changes of growing period, an attribute extensively used for agro-climatic classification (White et al., 2001); and the variable PAF which resembles the patterns of forest degradation, cropping and grazing observed in other tropical areas (Loker, 1994; Fujiska et al., 1998; McCracken et al., 1998), and which is probably acting in this classification as a complemetary indicator of farm adaptation response to its geographical surroundings.

The use of density surface poses several advantages for describing data accross the sample space when the modality of the distribution is displayed. Basically, this approach attempts to gain insight about the underlying data structure based on the location of modes, their number, width and height. The results observed in Fig. 2.10 and 2.11 provide an effective contextual visualisation of the spatial relation between attributes; and based on their modalities it might be inferred that density activity is suggesting three farm clusters.

2.6.2.2 Exploratory cluster analysis

The correlation matrix of the 10 variables involved in this study is presented in Table 2.5. A first glance shows that groups of strongly related variables can be identified (Barlett's test of sphericity $p < .0001$). For instance, on one hand STR, HRC, CMP, HMP and SOR showed high correlation each other and, on the other hand, variables MAI, PAF and FOR also described high levels of correlation between them, signaling that there were potential underlying dimensions that may capture most of the variability in this data set.

Potential multicolinearity problems were discarded calculating the determinant of this matrix (=0.234), confirming what was shown visually by the correlation matrix.

--- Correlation Matrix ---

	2 STR	3 HRC	5 PWM	6 CMP	7 HMP	8 MAI	9 SOR	10 PAF	11 FOR
2	1.000								
3	-0.120	1.000							
5	0.182	-0.473	1.000						
6	-0.216	0.083	0.057	1.000					
7	-0.186	-0.012	0.089	0.444	1.000				
8	-0.037	-0.270	0.019	0.074	0.135	1.000			
9	0.301	0.024	0.176	-0.039	-0.005	0.089	1.000		
10	0.052	-0.050	0.015	0.017	0.155	-0.306	-0.170	1.000	
11	-0.134	0.032	-0.041	-0.031	-0.273	0.054	-0.222	-0.293	1.000
12	-0.021	0.121	-0.063	0.265	0.131	-0.038	-0.076	-0.059	0.074

Table 2.5. Correlation matrix for the ten variables examined from Guarico dataset.

In Table 2.6 all orthogonal components calculated appear organized hierarchically by eigenvalue. Based on this information and following the Jolliffe (1972, 2002) criterion, six components were retained to perform clustering. As can be seen, 80 % of the common variability present in the data can be explained using only these components.

The distribution of componet loadings showing how variables are related to the first and second principal direction is plotted in Fig. 2.12. Only variables STR and SOR appear to highly influence the first component, and FOR the second, while remaining variables had large co-ordinate values on different axes, which is indicative of a shared contribution to both orthogonal directions of the model. They seem to be measuring the same aspects of a common underlying construct.

The positioning of variables on diagonally opposed quadrants indicate that they are inversely correlated, meaning that when the annual crop proportion (MAI, SOR) in farms increases the proportion of forest (FOR) decreases or *vice versa*. However, this

	Eigen Value	%	% Acumul	Histogram
1	1.813	18.13	18.13	**
2	1.696	16.96	35.09	***
3	1.506	15.06	50.15	**************************************
4	1.237	12.37	62.52	********************************
5	1.039	10.39	72.91	***************************
6	0.791	7.91	80.82	*********************
7	0.598	5.98	86.81	****************
8	0.483	4.83	91.63	*************
9	0.471	4.71	96.35	*************
10	0.365	3.65	100.00	**********
TOTAL	10.000			

Table 2.6. Histogram of components eigenvalues.

might be an uncertain conclusion given the closeness of variable MAI to the center of the co-ordinate system.

Overall, the loads plot also reveals that the first principal direction seems to correspond mostly to the stocking rate variable which best characterizes a density management construct. On the other hand, the second principal direction appears to be influenced mainly by those variables that related to land cover management encoding those aspects that support grazing processes in the system, given that this axis is strongly driven by the variable FOR at its negative end and showing a moderate impact of variables related to annual crops and pastures on its positive side.

Fig 2.13 shows that there is not a visual metaphor on how the variable configuration used in this study contributes underlying constructs alluded to from a three principal components perspective. It can be observed that there are two sets of variables apparently grouped together, but it cannot be assumed that they contribute with different information to common underlying dimensions. In this case, the predicted land cover and productive management dimensions do not appear entirely probable because

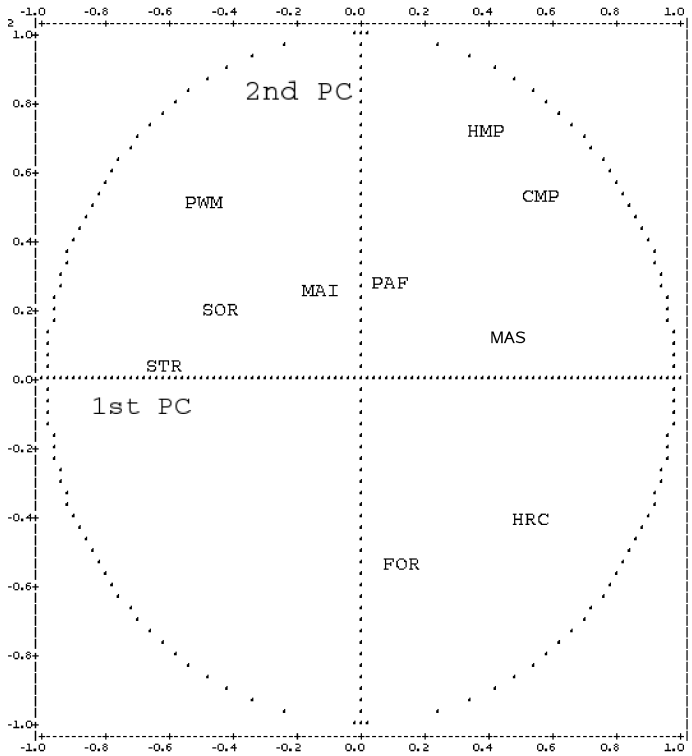


Fig. 2.12. PCA loadings plot of the first two principal components of the productive and land cover attributes data set.

certain attributes of the census database were not representative of the whole range for the referred dimension space. Moreover, some variables were grouped into unexpected sets, and it is clear that extracted salient features showed no structure in the data, confirming what was observed in the correlation matrix. Most variables did not result in clusters confirming to predictions from the census database.

Individual farm score projections are displayed in Fig. 2.14. Given that there is a direct Euclidian link between farm scores and variable loadings plots it can be observed that the position of farms labelled with the number 1 appear weakly governed by variables such as HMP and CMP when a comparison with Fig. 2.12 is made. On the positive side of PC 2 the variable PAF can also be found having a moderate impact

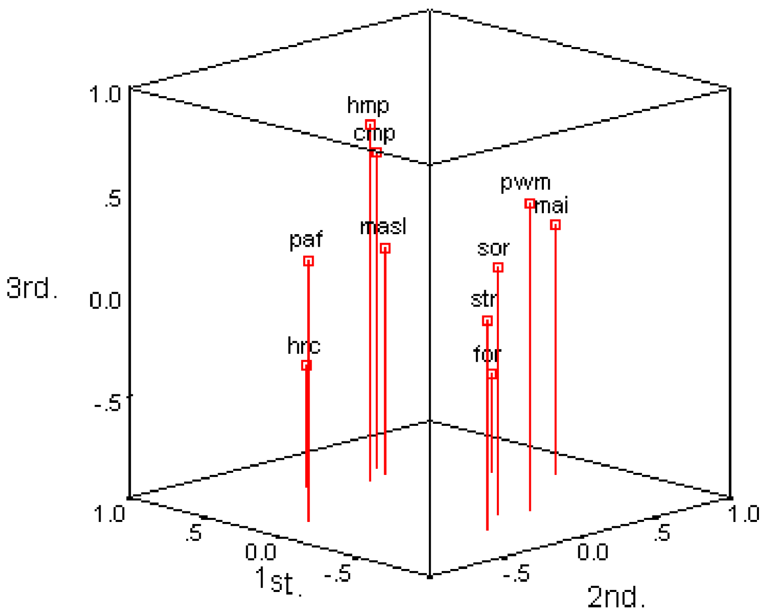


Fig. 2.13. Scatterplot of variables loadings for the first three principal components.

on this group and inversely correlated with FOR. Thus, if variables SOR and MAI are considered, this suggests that farms vary according to a complex set of local constraints apparently driven by the land cover devoted to support grazing.

This comparison also might indicate that farms type 2 seem to be pulled apart because of the influence of variables FOR and HRC; however there is not a clear separation between these farms and other goups, just a weak trend to cluster into separated sets can be seen.

Farms labeled as 3 were generally characterized by having high values on variables SOR, MAI, STR and PWM but low values on FOR and HRC. However, again there was not a clear decision boundary between this group and other classes, especially to the center of the co-ordinate system where individual farms show average properties with respect to their descriptor variables.

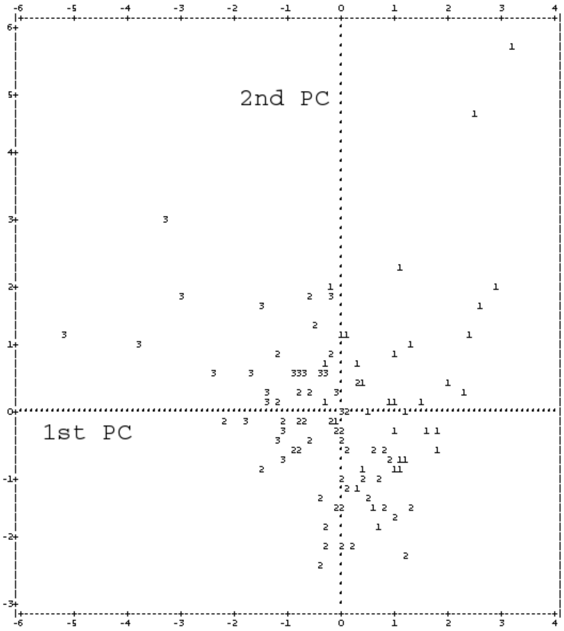


Fig. 2.14. PCA scores plot of farms' classes for first (x axis) and second (y axis) principal directions for Guarico group.

2.6.2.3 Summary of farm's classes (Guarico set)

Information about typical values by farm class for each variable and their dispersion can be found in Table 2.7, where a statistical summary is provided. These metrics include additional central tendency and dispersion metrics in order to avoid any biases because of outliers. Class 1 were characterized by showing high values for variable HMP whose trimmed means were 75% and 42% higher than classes 2 and 3 respectively while for milk production per cow (CMP) trimmed means were 28% and 23% higher when compared with classes 2 and 3 respectively. Nevertheless, this higher productivity of milk production does not seems to be supported by cash crops; in this sense, these result do not replicate the findings in most tropical countries (Renard, 1997; Milne, 2005; Landers, 2007).

The percentage of forest (FOR) within the farm appears as one of the variables responsible for clustering farm class 2. The trimmed means for this variable were 60% and 59% higher than for classes 1 and 3 respectively. Furthermore, these farms (class 2) also showed lower milk production per cow and per hectare with a moderate bovine density (0.44 AU/ha). It is informative to note that the variables SOR and MAI also showed the lowest values for this group while the trimmed mean for variable PAF was 11% and 6% higher in this set than farms labeled as 1 and 3. Indicating that grazing in this kind of farm mainly relies on forest and pastures with not much dependence on cash crops.

Class 3 farms were mostly influenced by variables SOR and MAI which registered the highest values in this group against farms classes 1 and 2 that registered lower values of 81% and 91% for SOR and for MAI, 12% and 61% respectively. The fact that stocking rate resulted in the lowest trimmed mean for this group might be interpreted as indicating that farms with dominance of annual crops have a lower livestock density. However, this conclusion is unreliable given the presence of nonlinear complexities within this variable; and also contradict experimental evidence (Renard, 1997; Ruthenberg, 1980; Entz et al., 2005; Domínguez, 2006).

Post-weaning management (PWM) was another variable that played an important role in farm class 3 differentiation. The highest values were reported for this group; followed by class 2 showing moderate values (21% less than class 3), and class 1 with the lowest population of young bovines per farm (41% less than class 3). These values of PWM suggest that class 3 farms keep a high proportion of young cattle in relation to their total bovine population, indicating that these farms as well as milk or dairy products also sell young animals. It is interesting to appreciate that HRC, which is a variable that accounts for the proportion of mature females within the herd, is negatively correlated with PWM; which makes sense, because as the stocking density of young animals approaches its minimum, mature animals (including females) must be proportionally increased in these farms.

Fig. 2.15 reveals additional insights into the layout of the data for the three farm clusters. As can be seen, variables are rendered as box plots by class and organized in such a way that for each class all variables were lined up to facilitate visual interpretation of the performance of classes across variables. As a result, plots were organized in two colums, where the group of variables that encode productive and reproductive information are laid to the left; and to the right those descriptors that encode for land use management.

The main distinction between box plots presented, when observed from top to bottom, is that each class shows a characteristic profile across the additive effect of the salient features that confirms their behaviour contributing to differences between groups. It is noteworthy to see that apart from few exceptions most variables were plagued with outliers and not normally distributed.

Finally, the central result of this part of the study is that there were no reasons to think that data structure in the Guarico set conform to the existence of separated classes in this group of farms. This was true for the three clusters suggested by the visualization analysis, despite the tendency observed on farm classes, to occupy different areas of the input space (described by the three principal components). As was shown above, these tendencies seem to be important in some attributes, but there are no clear parameters

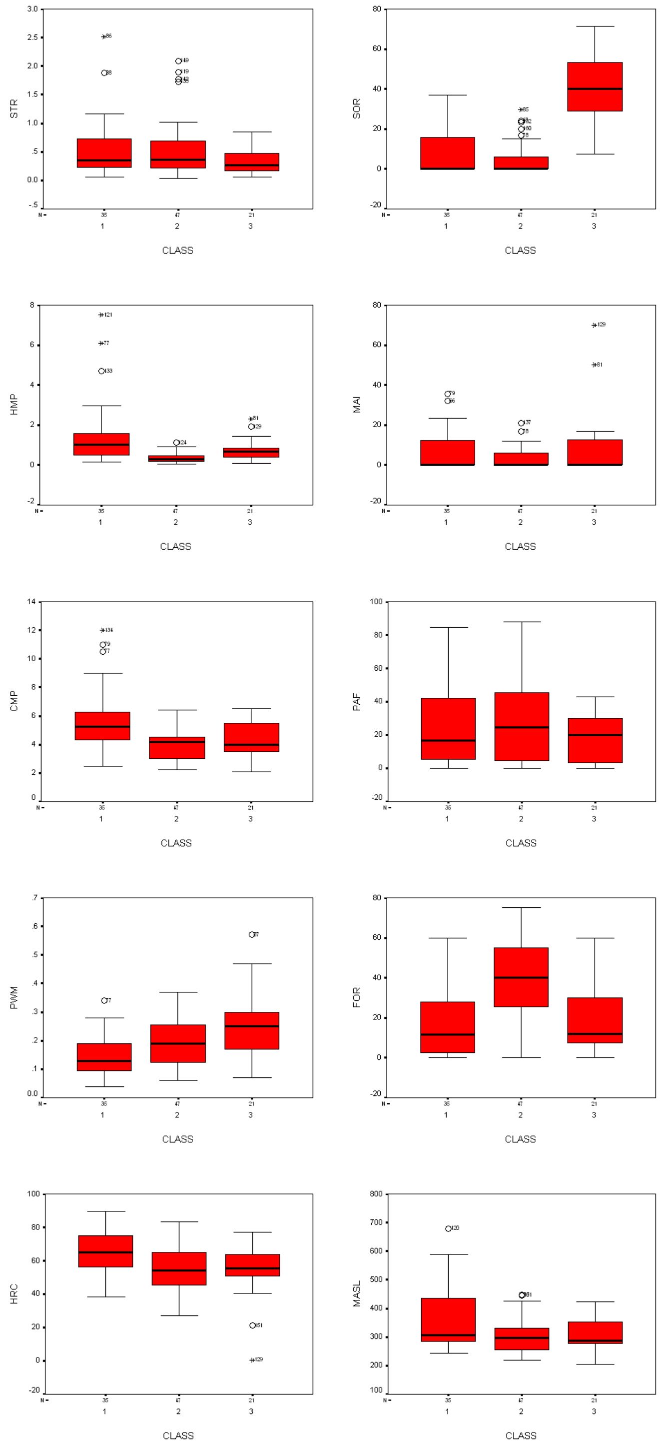


Fig. 2.15. Boxplots of productive and land cover attributes variations.

that permit us to draw any conclusion about potential clusters, because classes appear very overlapped.

Table 2.7. Summary statistics of 103 crop-livestock systems attributes by farm class.

			STR		HRC		PWM		CMP		HMP		MAI		SOR		PAF		FOR		MASL	
CLASS			Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E	Statistic	S. E
1	Mean		.5501	.0871	64.8640	2.210	.1491	.0121	5.6857	.3715	1.4503	.2718	6.6114	1.776	7.1440	1.956	25.5634	3.890	17.4497	2.996	360.952	18.45
	95% Confidence Interval for Mean	Lower Bound	.3730		60.3721		.1246		4.9307		.8979		3.0019		3.1680		17.6575		11.3616		323.447	
		Upper Bound	.7272		69.3559		.1737		6.4407		2.0026		10.2210		11.1200		33.4693		23.5378		398.457	
	5% Trimmed Mean		.4822		64.9350		.1458		5.5286		1.2190		5.4505		6.0136		23.9925		16.1233		351.862	
	Median		.3563		64.9400		.1300		5.2500		1.0100		.0000		.0000		16.6700		11.6700		307.580	
	Variance		.266		170.989		.005		4.831		2.586		110.412		133.967		529.687		314.107		11920.3	
	Std. Deviation		.51558		13.0763		.07135		2.19788		1.60797		10.5077		11.5744		23.0149		17.7231		109.180	
	Minimum		.06		38.46		.04		2.50		.15		.00		.00		.00		.00		242.77	
	Maximum		2.52		89.66		.34		12.00		7.50		35.71		37.04		84.58		60.00		679.19	
	Range		2.46		51.20		.30		9.50		7.35		35.71		37.04		84.58		60.00		436.42	
	Interquartile Range		.5546		21.4900		.1000		2.3000		1.1200		14.2900		16.6700		39.4800		27.0700		159.720	
	Skewness		2.231	.398	.065	.398	.580	.398	1.330	.398	2.569	.398	1.429	.398	1.286	.398	.779	.398	1.029	.398	1.074	.398
	Kurtosis		6.037	.778	-.751	.778	.034	.778	1.803	.778	6.859	.778	.886	.778	.196	.778	-.219	.778	.056	.778	.645	.778
2	Mean		.5228	.0712	55.1991	1.893	.1953	.0123	3.9689	.1491	.3566	.0334	3.0700	.7312	4.4119	1.127	27.9074	3.678	40.6319	3.107	302.398	8.761
	95% Confidence Interval for Mean	Lower Bound	.3795		51.3880		.1705		3.6688		.2894		1.5982		2.1442		20.5049		34.3776		284.762	
		Upper Bound	.6662		59.0103		.2201		4.2691		.4238		4.5418		6.6796		35.3100		46.8862		320.034	
	5% Trimmed Mean		.4686		55.1290		.1937		3.9332		.3387		2.4209		3.4396		26.3247		41.0380		298.889	
	Median		.3600		54.0500		.1900		4.2000		.2900		.0000		.0000		24.5100		40.0000		296.190	
	Variance		.238		168.489		.007		1.045		.052		25.126		59.652		635.650		453.745		3607.89	
	Std. Deviation		.48832		12.9803		.08436		1.02225		.22890		5.01262		7.72346		25.2121		21.3013		60.0657	
	Minimum		.04		27.18		.06		2.25		.06		.00		.00		.00		.00		220.21	
	Maximum		2.09		83.33		.37		6.44		1.12		21.00		29.53		88.00		75.00		446.67	
	Range		2.05		56.15		.31		4.19		1.06		21.00		29.53		88.00		75.00		226.46	
	Interquartile Range		.4988		22.0100		.1400		1.5000		.2900		6.2500		6.6700		43.6200		35.0000		82.0900	
	Skewness		1.849	.347	.049	.347	.261	.347	.302	.347	1.213	.347	1.777	.347	1.805	.347	.736	.347	-.145	.347	.860	.347
	Kurtosis		3.210	.681	-.410	.681	-.971	.681	-.520	.681	1.603	.681	2.974	.681	2.406	.681	-.386	.681	-.903	.681	.114	.681
3	Mean		.3584	.0523	53.4786	3.775	.2543	.0272	4.2495	.2903	.7633	.1271	9.0776	3.985	41.9290	3.975	18.1205	3.466	18.0667	3.427	309.780	14.65
	95% Confidence Interval for Mean	Lower Bound	.2492		45.6045		.1975		3.6441		.4983		.7656		33.6371		10.8915		10.9187		279.211	
		Upper Bound	.4676		61.3526		.3111		4.8550		1.0284		17.3897		50.2210		25.3494		25.2146		340.350	
	5% Trimmed Mean		.3473		55.0758		.2472		4.2464		.7162		6.2503		42.1735		17.7588		16.7937		309.335	
	Median		.2714		55.2500		.2500		4.0000		.6800		.0000		40.0000		20.0000		12.0000		288.000	
	Variance		.058		299.228		.016		1.769		.339		333.444		331.833		252.207		246.587		4510.13	
	Std. Deviation		.23981		17.2982		.12484		1.33013		.58234		18.2604		18.2163		15.8810		15.7031		67.1575	
	Minimum		.06		.44		.07		2.07		.08		.00		7.50		.00		.00		203.69	
	Maximum		.86		76.92		.57		6.50		2.31		70.00		71.43		42.86		60.00		424.15	
	Range		.80		76.48		.50		4.43		2.23		70.00		63.93		42.86		60.00		220.46	
	Interquartile Range		.3354		13.9350		.1400		2.0500		.6700		12.9150		27.5750		31.1500		22.7500		98.7800	
	Skewness		.980	.501	-1.678	.501	.933	.501	-.070	.501	1.326	.501	2.588	.501	.089	.501	.288	.501	1.132	.501	.467	.501
	Kurtosis		-.167	.972	3.761	.972	.818	.972	-1.101	.972	1.457	.972	6.624	.972	-.839	.972	-1.531	.972	.951	.972	-.675	.972

2.6.3 Discriminant analysis

Stepwise discriminant analysis was performed to analyze the differences between groups and to have an idea about how the variables taken from the census database discriminate the different farm classes. It is Hoped that if a discriminative direction is found in the experimental groups, it can be held to have occurred as a result of a good choice of the number of clusters into which the sample was segmented; and as the outcome of an effective feature extraction process.

Table 2.8 shows a summary of statistics of the multivariate discriminant analysis for both Aragua-Guarico and Guarico experimental groups. It appears that, regardless of the percentage of farms correctly classified, the proportion of variance explained by the model variates (r^2) seems not to be very high, particularly for the Guarico group, where the low accuracy makes difficult the use of this information as labels in an eventual supervised classification. Nevertheless, dissected classes appear to be meaningful in the Aragua-Guarico experimental set, since the wilks' lambda test, which represents the ratio of within-group variance to total sample variance, was statistically significant ($p < 0.01$), as the remaining statistics displayed in the summary shown.

Table 2.8. Multivariate stepwise discriminant analysis summary statistics by experimental group.

$Group$	$\%C$	$W\lambda$	PT	T^2	RM	r^2
$Aragua - Guarico$	91.7	0.37	0.62	1.65	1.65	0.62
$Guarico$	78.3	0.15	1.21	3.36	2.30	0.60

%C: percentage classified correct; $W\lambda$: Wilks' lambda; PT: Pillai's trace; T^2: Hotelling's test; RM: Roy's minimum root; r^2: squared average canonical correlation

The variables that best separated classes in the experimental grouping Aragua-Guarico included STR, CMP, HMP, SOR, FOR, and MASL. On the other hand, with low accuracy, the best discriminators for Guarico group membership were STR, HRC, CMP, MAI, SOR, FOR and MASL. As can be appreciated, cash crops and forest tend to dominate the land use variables that were definitely important for classification in the Aragua-Guarico group, and relatively important for the Guarico group. On the production side, the role played by milk production at cow and hectare level of resolution is remarkable. These findings partially corroborate Domínguez (2006) and Espinoza et al. (2005) initial categorization of livestock farming systems in this area. In these studies, land use management and milk production associated variables, served to discriminate between modalities of production systems.

Another important aspect of farming class predictability seems to be the altitude (MASL), which is not surprising given the high spatial correlation between topography and the length of growing period in this area (Fischer et al., 2000; Rodríguez and González, 2001). The presence of forest as a significant predictor of farm modality agrees with the literature in which forest patterns of degradation may be symptoms that permit differentiate particular ways in which evolution of farming systems is taking place (Nicholson et al., 1999; Fisher and Thomas, 2004).

A notable finding that might require additional research involves the role played by replacement females with respect to the total female herd (HRC). It can be speculated that the inclusion of this variable is probably helping with differentiate management response between farmers inclined to a particular farm typology; however there is not strong evidence in the literature to support this idea.

In summary, lack of accuracy within the Guarico set, reveals that a lineal model that discriminates all farm classes with minimum overlap, could not be found (Hastie et al., 2001). However, based on the feedback from the PCA, there were several compelling tendencies suggesting the existence of potential farm subsets that remain hidden. So in the interest of generating maximally useful data, a nonlinear mapping of the original attributes (Schölkopf et al., 1998) might project them onto crucial directions where it is possible to discharge meaningless aspects of the data; and then classes can be effectively separated.

2.7 Conclusion

In this chapter density surface visualisation of land cover and productive attributes in crop-livestock data have been addressed. It has been demonstrated that between farms variation can be captured succesfully in a spatially concrete context, without any *a priori* assumption about data distribution. This permits the local examination of attribute interactions in a given geographical domain and enables the use of activity density as ancillary methodology for additional quantitative approaches. Thereby, with the exception of the Guarico set, there was a common configuration between the spatial distribution of the attributes examined by geovisualisation techniques, and the underlying dimensions that emerged as a result of the principal component analysis. One of the major advantages of this ancillary methodology is that it provides a simple and spatially explicit means to make decisions about the number of clusters into which a farm population may be segmented. In this way, findings of visualization served as a basis for the unsupervised classification process part of this research. It was shown that farm subsets identified through 3D visualisation of attribute density, were confirmed by clustering analysis in the Aragua-Guarico set. The intrinsic farm classes in the sampled data resembled the spatial attribute gradients that were observed on density surfaces; and the taxonomy of farms encountered corresponded to meaningful features at ground level. It is noteworthy that farm classes yielded were most influenced by variables associated with the land cover domain and within the productive variables, those related to milk production exhibited more weight on the segmentation.

Although semantically relevant classes were identified in the Aragua-Guarico group; that was not the case for the Guarico set, given that informational classes in this group of farms were overlapped. One possibility to overcome this poor performance, is by the implementation of a nonlinear feature extration approach before proceeding with the unsupervised classification. Such an approach is delineated in the next chapter, where it is explored whether nonlinear mapping might help on the class separation for the Guarico set, and confirming the findings in the Aragua-Guarico set.

Chapter 3

Kernel Based Unsupervised Classification of Crop-Livestock Systems

3.1 Abstract

Feature extraction works by finding a suitable transformation of attributive data into a low dimensional feature space. When the standard algorithm used is PCA, such a transformation might be orthogonal and could result from the solution of an eigenvalue problem leading to a new coordinate system that contains the original data-set projections. This new coordinate system consists of linear decomposition attributes which are not always appropriate for crop-livestock pattern recognition, given that such data is normally full of nonlinearities. The main objectives of this research are to classify the given crop-livestock obsevations within the study area into groups, by using a nonlinear feature extration method; to compare nonlinear and linear approaches on their ability to empower better quality information for eventual clustering; and, in the interest of generating useful features, to assess within the chosen nonlinear method, the impact of different kernel functions. Discriminant analysis was used to assess the discriminative power of features and a t-test was performed to compare resulting Mahalanobis distances between linear and kernel PCA. Results reveal an improved discriminative power of kernel methods in comparison to linear PCA ($p < 0.001$).

3.2 Introduction

Feature extraction constitutes an important task within multidimensional crop-livestock pattern classification. The idea behind it is, among others, to isolate those statistical characteristics of the data that portray essential elements of them, and to provide a better understanding about the underlying process that generates the data (Guyon and Elisseeff, 2003). Avoiding redundancy of even low dimensional input data, this characterises crop-livestock systems (crop production, land use, livestock production, management, etc) by finding meaningful projections into a feature space.

Principal components analysis (PCA) is one of the standard techniques to obtain features from input data (Jolliffe, 2002). This is achieved by maximising the projected variance onto mutually orthogonal eigenvectors along the directions of higher eigenvalues through iterative algorithms that minimise information losses. PCA basically performs a linear decomposition of input vectors, into a space whose coordinate system is hierarchically organised by data variability (Bishop, 2006).

PCA has demonstrated good performance in previous studies related to the farming system field, especially for dimensionality reduction and for interpreting multiple crop-livestock signals (Köbrich et al., 2003). However, crop-livestock systems variables interact in a non-linear dynamic, which in turn usually produces complex outcomes of landscape heterogeneity, livestock activity, and vegetation interactions. In consequence, most of these crop-livestock systems traits are subject to limited description within the second order correlation approach of linear PCA.

One solution to this problem is the generalisation of linear PCA setting to an application of kernel principal component analysis (KPCA) (Schölkopf et al., 1998). This algorithm combines linear PCA simplicity with the capability of integral operators, known as kernel functions; to express data from input space as dot products in the feature space. This method enables the construction of nonlinear versions of the original variables in a high dimensional context (Shawe-Taylor and Cristianini, 2006).

On the basis of this knowledge, the primary aims of this study were to classify the given crop-livestock obsevations within the study area into groups, by using a nonlinear feature extration method; to compare nonlinear and linear approaches on their ability to empower better quality information for eventual clustering; and, in the interest of generating useful features, to assess within the chosen nonlinear method, the impact of different kernel functions.

The plan of this chapter is as follows. Section 3.3 presents some theoretical aspects of the principal component analysis and its nonlinear generalisation with kernel; this section also includes some general formalisms about how these tools might be constructed. This leads to section 3.4, which is concerned with data and all methodological aspects that were followed during this research. Then the main results and discussion are included in section 3.5; and finally, conclusions are summarised in section 3.6.

3.3 Principal components analysis

Feature extraction through principal component analysis (also referred to as *Karhunen-Loève* transform) can be traced back to the pioneering work of Pearson (1901) and Hotelling (1933a,b). Today PCA is one of the feature extraction methods most used in farming systems (Köbrich et al., 2003; Berdegue and Escobar, 1990), and there has been considerable research surrounding the application of this technique in different topics of pattern recognition (Duda et al., 2001; Jolliffe, 2002; Bishop, 2006). Basically, the method pursues the finding of a lower dimensionality space by the orthogonal transformation of the coordinate system where a given data set is described, with the aim of identifying directions of maximum variability. Let us consider a set of observations such that:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & & & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \tag{3.1}$$

where X is the original data set $m \times n$ matrix, n is the number of samples, which conform m-dimensional vectors ($\alpha = x_1 \dots x_m \in \mathbb{R}^N$) of random variables in an arbitrary space. These vectors are linearly decomposed into another coordinate system whose first axis is a projection of each observation and respond to the linear function $\alpha_1^T x$. This new $m = 1$-dimensional subspace is oriented to the direction where the elements of X show their highest variability.

The subsequent axes are orthogonally aligned in X to the following highest direction through recursive linear decompositions until m vectors have been aligned $\alpha_m^T x$. The axes of this new coordinates system are organized hierarchically according to data variability, and are normally referred to as principal components. It might occur that those components in directions of very low variability are practically near-constant for all vectors (Jolliffe, 2002), and can be eliminated since they do not contribute new information. Therefore, a substantial dimensionality reduction ($<< m$) of the problem is usually achieved, given that typically a few axes are enough to retain most of the data structure, if this exists.

Generally the feature extraction and dimensionality reduction proceeds as described above. However, it is worth pointing out the following precisions: to obtain the new coordinate system data must be projected to the direction aligned with the maximum variance; this best fit axis passes through the mean of the data cloud which is given by:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x \tag{3.2}$$

In order to establish this direction, data is projected onto the $d = 1$-dimensional vector whose scalar value projection is defined by $\alpha_1^T x$ with a projected data variability such that:

$$\frac{1}{n}\sum_{i=1}^{n} \{\alpha_1^T x - \alpha_1^T \bar{x}\}^2 = \alpha_1^T S \alpha_1 \tag{3.3}$$

Variability maximisation is pursued in such a way that the sum of squared of element on α_1 equals 1 ($\alpha_1^T S\alpha_1 = 1$), where S is defined by:

$$S = \frac{1}{n}\sum_{i=1}^{n}(x_n - \bar{x})(x_n - \bar{x})^T \tag{3.4}$$

At this stage, the main task is the minimisation of redundancy present in the covariance and maximisation of useful information provided by the variance. Diagonal elements of the covariance matrix summarise the data dynamic of interest as long as they are high, otherwise, they are associated with noise. Maximisation of $\alpha_1^T S\alpha_1$ is performed incorporating a lagrange multiplier λ:

$$\alpha_1^T S\alpha_1 + \lambda_1(1 - \alpha_1^T\alpha_1) \tag{3.5}$$

whose derivative with respect to α_1 yield:

$$S\alpha_1 = \lambda_1\alpha_1 \tag{3.6}$$

Considering that the eigenvalues are ordered in a decreasing way ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$) being $\lambda' = \lambda_{max}$ and proceeding by mathematical induction, it is assumed that principal components from 1 to $m-1$ can be found along the first $m-1$ directions of eigenvectors. The principal component m_{th} is constrained to be orthogonal to such directions, then in the variance expression in this direction, it must be stated that $\alpha_1 \cdots \alpha_{m-1} = 0$. So maximising S subject to this condition and being α an unitary vector $|\alpha| = 1$, or $S\alpha = 1$

$$\alpha_1^T S \alpha_1 = \lambda_1 \tag{3.7}$$

Hence, the principal component m^{th} can be found along with the eigenvalue m^{th} and it can be established that the variance equals the eigenvalue m^{th} when α_1 is aligned to the direction of the m^{th} principal component (Jolliffe, 2002; Bishop, 2006).

In the literature it can be encountered that correlation and covariance matrix are alternatively alluded. To be completely accurate, covariance matrix is the mean scalar product of patterns minus the mean, while correlation matrix is a standarized version of the covariance matrix, given that correlation is originated from the mean scalar products of the patterns divided by the product of multiplying the standard deviation of patterns (Field, 2005). Nevertheless, this kind of analysis is performed from centred data ($\sum_{i=1}^{m} x_i = 0$) hence both matrices are equivalent.

Principal component analysis has been shown to be a very powerful technique in finding orthogonal derived variables that in succession maximise the variance of a given data set (Mardia et al., 1979; Jolliffe, 2002). However, sources of nonlinearities and complexities in real-world problems might require to be hypothetised in sub-spaces much more rich than linear combination of features (Cristianini and Shawe-Taylor, 2000). Therefore, nonlinear generalisations of principal components analysis are playing an important role in pattern analysis through the inclusion of kernel functions notions.

3.3.1 Kernel principal components analysis

The kernel "trick" permits the generalising of any algorithm that uniquely depends on inner products (Aizerman et al., 1964). This approach has proven to be particularly helpful for those statistical problems that involve feature extraction (Schölkopf et al., 1998); classification (Boser et al., 1992); regression (Williams, 1998) and clustering (Graepel and Obermayer, 1998; Crammer and Singer, 2002). Generally it can be said that kernel methods serve to induct non-linear functions in feature spaces usually of high dimensionality, and also may be incorporated into the dual form of most algorithms in such a way that it is not necessary to calculate explicitly the transformation to the feature space (Shawe-Taylor and Cristianini, 2006).

A result of the inclusion of the kernel idea within the dual representation, is that the computation task is not affected by the feature space dimensionality (Cristianini and Shawe-Taylor, 2000), and given that the gram-matrix is the unique information used in the feature space, the amount of work required to calculate the inner product is not necessarily proportional to the feature number; that way the use of kernels can be seen as a means to establish an implicit correspondence between the original data and the feature space, without the limitations associated with the computation of such correspondence.

Within a broad context, the study of statistical aspects of pattern analysis has been approached from two main paradigms: the Bayesian approach (Duda et al., 2001) and empirical processes (Vapnik, 1995). The work of Boser et al. (1992) pioneered the merging of kernel methods and the statistical learning theory (empirical processes approach) through large margin classifiers. However, most of the theoretical development on kernel methods has its origin in Mercer (1909) and Aronszajn (1950) research, where fundamental issues of Mercer's theory and Hilbert's spaces were treated respectively.

After the crisis of the main linear approaches of common use in the learning machines field (Fisher, 1936; Rosenblatt, 1958) as a result of the publication of Minsky and Papert (1969) about the limited computational power of linear methods, one of the alternatives proposed was the threshold multilayer structures, which led to the development of neural networks (generalised perceptron) with associated algorithm as back propagation (Hertz et al., 1991) .

The other approach was data preprocessing: in other words, the projection of data into a higher dimensional space to increase the computational power by including redundancies in their representation and assuring an effective feature extraction process from very complex data. An interesting alternative method to accomplish the above task, was the use of kernel methods, whose functions and corresponding feature spaces theory derive from integral operators study (Aronszajn, 1950; Berg et al., 1984; Sahitoh, 1988). The inclusion of these constructs into a nonlinear generalisation of principal components analysis was led by Schölkopf et al. (1998). One of the main achievements of the study was to express the feature extraction based on eigen-decomposition, as a process that pursues the finding of orthonormalized directions in a kernel-defined feature space by dual representation, along which data variability is maximised (Shawe-Taylor and Cristianini, 2006).

Nonlinear PCA might be expressed as an eigenvalue problem. Consider a feature space $\mathcal{H}$ associated to the input space $\mathbb{R}^m$ by a non-linear transformation:

$$\Phi : X \Rightarrow \mathcal{H}, \qquad x \Rightarrow \Phi(x) \tag{3.8}$$

The feature space $\mathcal{H}$ can show a dimensionality arbitrarily big $(m \times m)$, and potentially infinite. Assuming that in this space data are centered according to $\sum_{i=1}^{m} x_i = 0$, covariance matrix can be written in $\mathcal{H}$ as following:

$$Cov = \frac{1}{p} \sum_{i=1}^{p} \Phi(x^j)\Phi(x^j)^T \tag{3.9}$$

Having a feature space that possesses infinite dimensions, $\Phi(x^j)\Phi(x^j)^T$ can be considered the linear operator in $\mathcal{H}$ that performs the transformation $x \Rightarrow \left\langle \Phi(x^j)\Phi(x^j)^T \cdot x \right\rangle$. Then, the main objective consists of finding the solution to an eigenvalue problem that satisfy $\lambda v = Cov\ v$, without working explicitly in the feature space. By analogy to the input space analysis, all solutions v with $\lambda \neq 0$ are encountered in the sub-space generated by $\Phi(x^1), \ldots, \Phi(x^p)$. This includes two helpful implications:

1. The following equation can be used:

$$\lambda \left\langle \Phi(x^n) \cdot v \right\rangle = \left\langle \Phi(x^n) \cdot Cov\ v \right\rangle \qquad \forall n = 1, \ldots, p \tag{3.10}$$

2. Provided that $\lambda \geq 0$ are found subject to the existence of non null eigenvectors $v \in \mathcal{H} \backslash \{0\}$; and given that coefficients belonging to $\alpha_i (i = 1, \cdots, p)$ are determined

by linear combinations of $\Phi(x^n)$, υ can be written as:

$$\upsilon = \sum_{i=1}^{p} \alpha_i \Phi(x^i) \tag{3.11}$$

These expressions can be merged by substituting both into $\lambda\upsilon = Cov\,\upsilon$ and multiplying both sides by $\Phi(x)^T$ in order to express them as kernel terms $K(x^i, x^j) = \Phi(x^i)^T \Phi(x^j)$:

$$\lambda \sum_{i=1}^{p} \alpha_i \left\langle \Phi(x^n) \cdot \Phi(x^i) \right\rangle = \frac{1}{p} \sum_{i=1}^{p} \alpha_i \left\langle \Phi(x^n) \cdot \sum_{i=1}^{p} \Phi(x^j) \left\langle \Phi(x^j) \cdot \Phi(x^i) \right\rangle \right\rangle \qquad \forall\, n = 1, \ldots, p \tag{3.12}$$

which in terms of the matrix (Gram $p \times p$) notation, integrated by the elements $K_{ij} = \left\langle \Phi(x^i) \cdot \Phi(x^j) \right\rangle$, the equation for all n are consolidated in:

$$p\lambda K\alpha = K^2\alpha \tag{3.13}$$

where α represents the column vector integrated by elements $\alpha_1, \cdots, \alpha_p$. Finding solutions to the previous equation requires to solve an eigenvalue problem

$$p\lambda\alpha = K\alpha \qquad \forall\lambda \neq 0 \tag{3.14}$$

It can be demonstrated that this simplification (removing K from both sides) led to (3.14) without those K that showed zero eigenvalues, not affecting the projection of principal componets and bringing all useful solutions from (3.13). So if $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eigenvalues of K ($p\lambda$ solutions) and $\alpha^1, \cdots, \alpha^p$ the whole corresponding eigenvectors set, being λ_q the last non-zero eigenvalue (assuming that Φ is not identically 0). The condition of unitary norm ($\langle \upsilon^n \cdot \upsilon^n \rangle = 1$) for corresponding vectors in the feature space leads to the following solution of normalisation over $\alpha^1, \cdots, \alpha^q$ when (3.11) and (3.14) are used:

$$\begin{aligned} 1 &= \sum_{i,j=1}^{p} \alpha_i^n \alpha_j^n \left\langle \Phi(x^i) \cdot \Phi(x^j) \right\rangle = \sum_{i,j=1}^{p} \alpha_i^n \alpha_j^n K_{ij} \\ 1 &= \left\langle \alpha^n \cdot K\alpha^n \right\rangle = \lambda_n \left\langle \alpha^n \cdot \alpha^n \right\rangle \end{aligned} \tag{3.15}$$

The principal components projections can be calculated by projecting a x test point with an image $\Phi(x)$ onto eigenvectors υ in the feature space with $n = 1, \cdots, q$; and expressing them in kernel notation using (3.11); that way principal components can be extracted:

$$\left\langle \upsilon^n \cdot \Phi(x) \right\rangle = \sum_{i=1}^{p} \alpha_i^n \left\langle \Phi(x^i) \cdot \Phi(x) \right\rangle \tag{3.16}$$

are the non-linear principal components or features corresponding to Φ (Schölkopf et al., 1998; Bishop, 2006).

Comparison of kernel and linear principal components is possible, and the next section is concerned with data and methodological aspects that posed a real-world agricultural problem in perspective using both approaches.

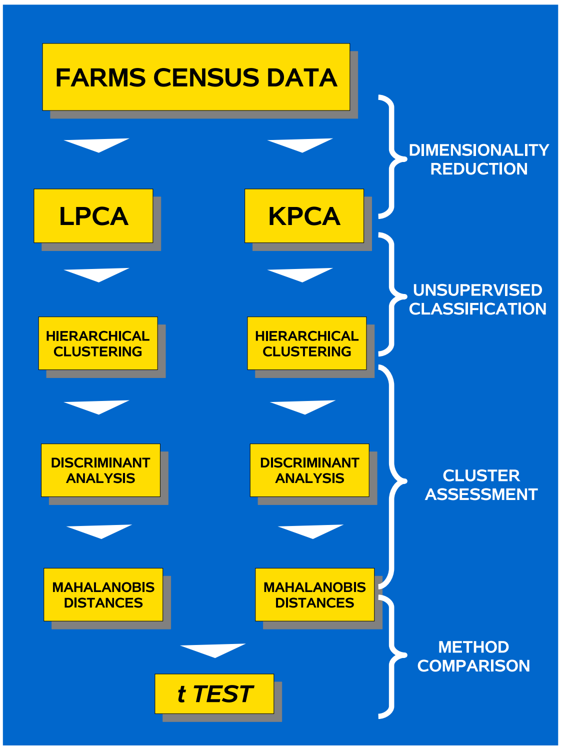


Fig. 3.1. Methodology scheme.

3.4 Data preprocessing and methods

To detect differences between feature extraction paradigms, a repeated-measures design was adopted; all farms were included in each treatment. Fig. 3.1 illustrates a scheme where the methodology used is outlined. As can be seen two methods (independent variables) of feature extraction were compared, linear (LPCA) (Pearson, 1901; Hotelling, 1933a,b) and kernel principal component analysis (KPCA) (Schölkopf et al., 1998). Differences in performance between both methods were examined based on the effectiveness of extracted features to yield meaningful and compact farm groups (dependent variable) within unsupervised classification by hierarchical clustering procedures (Ward, 1963; Johnson, 1967), using as few principal components as possible. For the purpose of this study, meaningful groups were defined as those clusters whose means were significantly different from each other, showing strong similarities within groups and possessing high variability between groups. Such estimations were based on a discriminant analysis approach (Fisher, 1936) using the statistics of Wilks' lambda ($W\lambda$), Hotelling's test (T^2), Pillai's trace test (P); Roy's maximum root (RM); and average squared canonical correlation (r^2).

Of primary importance when evaluating group compactness is the Mahalanobis distance between each sample and their class centroid obtained from discriminant analysis, which were used to perform t-test, in order to compare how closely individuals within different groups were positioned together under linear and kernel approaches. Results from this test were considered significant at a confidence level of 95%.

The database used in this research includes census information from 168 households whose attribute values were centered and normalised before proceeding with the experiments. Sampled farms were divided into two experimental groups: Aragua-Guarico group (168 farms) and Guarico group (103 farms) to gain insight into the performance of both feature extraction methods over data, linearly or non-linearly separable. Linear feature extraction was performed using the software CSTAT (CIRAD, 1989); the kernel approach was done with software developed within the framework of this research following the principles described by Schölkopf et al. (1998). Discriminant analysis was carried out through the 7M routine of the sofware BMDP (Dixon et al., 1981); and means comparisons were performed using the paired samples t-test of the software SPSS (SPSS-Inc, 1999).

3.5 Results and discussion

3.5.1 Aragua-Guarico set

3.5.1.1 Feature extraction

Initially, to gain insight into the minimum information required to assign farm objects to a category using two kernel methods (Gaussian and polynomic), successive hierarchical classifications were performed increasing recursively the number of feature vectors (principal components) extracted in the segmentation process. Table 3.1 shows several statistics pertaining to a linear discriminant analysis considering previously classified farms groups, using from 1 to 6 principal components as feature vectors. As can be seen, the best separation between clusters appears to occur when only the 1st principal direction was used for the classification task for both Gaussian and polynomic kernels.

Table 3.1. Impact of number of feature selected on clustering performance using two different kernel functions (Gaussian and polynomic) after stepwise discriminant analysis for Aragua-Guarico group data, as number of retained principal components (PC) in the hierarchical classification was increased.

	Gaussian ($\sigma = 8$)						*Polynomic* ($order = 3$; $\sigma = 60$)					
PC	$\%C$	$W\lambda$	PT	T^2	RM	r^2	$\%C$	$W\lambda$	PT	T^2	RM	r^2
1	96.4	0.13	0.86	6.33	6.33	0.86	98.2	0.09	0.90	9.7	9.7	0.90
2	96.4	0.15	0.84	5.52	5.52	0.84	98.2	0.09	0.90	9.7	9.7	0.90
3	91.1	0.30	0.69	2.29	2.29	0.69	99.4	0.32	0.67	2.06	2.06	0.67
4	92.3	0.29	0.70	2.40	2.40	0.70	99.4	0.32	0.67	2.06	2.06	0.67
5	93.5	0.31	0.68	2.20	2.20	0.68	99.4	0.32	0.67	2.06	2.06	0.67
6	94.0	0.30	0.69	2.24	2.24	0.69	99.4	0.32	0.67	2.06	2.06	0.67

%C: percentage classified correct; $W\lambda$: Wilks' lambda; PT: Pillai's trace; T^2: Hotelling's test; RM: Roy's minimum root; r^2: squared average canonical correlation

It is noteworthy that slightly more than 10% higher was the squared average canonical correlation for the polynomic kernel. Moreover, the Wilks' lambda statistics was the lowest; and Pillai, Hotelling and Roy's tests achieved their maximun value within polynomic approach, illustrating the superior performance reached with this kernel when compared with Gaussian.

3.5.1.2 Clustering performance

Table 3.2 shows a comparison between profiles of clustering performance using kernel (polynomic and Gaussian) and linear approaches. Overall, the figures would appear to suggest that featured extracted by kernel paradigm yield the highest performance in terms of classification accuracy and explained variance by the model (r^2). Additionally, both kernel methods of feature extraction, after hierarchical clustering, seem to yield groups where the variation explained by classes are higher than those clusters based on feature vectors generated by the linear approach.

Table 3.2. Impact of kernel function on hierarchical clustering performance using linear and different kernel (Gaussian and polynomic) approaches after stepwise discriminant analysis for Aragua-Guarico group data.

Kernel	$\%C$	$W\lambda$	PT	T^2	RM	r^2
Linear	91.7	0.37	0.62	1.65	1.65	0.62
Gaussian	96.4	0.13	0.86	6.33	6.33	0.86
Polynomic	98.2	0.09	0.90	9.7	9.7	0.90

%C: percentage classified correct; $W\lambda$: Wilks' lambda; PT: Pillai's trace; T^2: Hotelling's test; RM: Roy's minimum root; r^2: squared average canonical correlation

A visual approximation of these differences can be appreciated in Fig.3.2, where squared adjusted means of Mahalanobis distance and their respective confidence intervals (95%) are depicted by feature extraction method. As can be observed, clusters segmented from feature vectors extracted by the linear approach and the Gausssian kernel were shown to be comparatively more scattered with respect to the clusters achieved from the polynomic feature extraction method, which showed a higher proximity (minimium distance) between a within-class object and its cluster centroid.

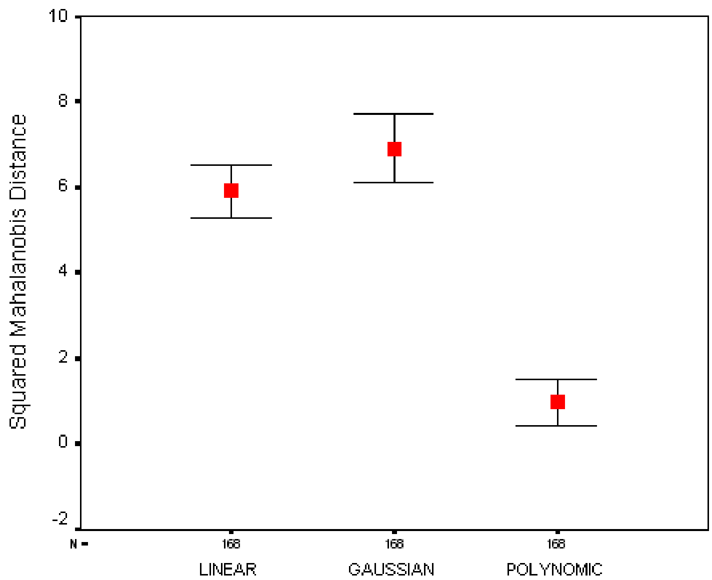


Fig. 3.2. Adjusted means and confidence intervals (95%) of squared Mahalanobis distance by selected feature extraction method (linear, gaussian and polynomic) for Aragua-Guarico group, after stepwise discriminant analysis.

This graphically predicted tendency is ratified by the paired samples statistics of squared Mahalanobis distances shown in Table 3.3. It can be observed that standard errors remained relatively small to the sample mean for the three approaches, suggesting that populations are well represented in all these samples. However, the real differences lie in the fact that cluster segmented from feature vectors achieved with polynomic kernel tended to show smaller within-class mean distances than those clusters whose feature extraction was performed by linear approach or Gaussian kernel method.

Table 3.3. Aragua-Guarico paired samples statistics.

		Mean	*N*	*Std. Deviation*	*Std. Error Mean*
Pair 1	LPCA	5.9268	168	7.48322	.57734
	GKPCA	6.9190	168	7.31799	.56459
Pair 2	LPCA	5.9268	168	7.48322	.57734
	PKPCA	.9833	168	5.09151	.39282
Pair 3	GKPCA	6.9190	168	7.31799	.56459
	PKPCA	.9833	168	5.09151	.39282

LPCA: Linear principal component analysis
GKPCA: Kernel principal component analysis (Gaussian)
PKPCA: Kernel principal component analysis (polynomic)

Linear approach and Gaussian kernel lead to within-group mean distances that exceed polynomic kernel by a ratio of 6.04 and 7.05 to 1 respectively. Statistically these means were different at a high level of significance ($p < 0.001$) as can be noticed in Table 3.4, which shows a summary of the t-test. It was clear that feature extraction through the polynomic kernel created a systematic effect, big enough to lead to mean differences that overcome the possibilities of any random effect.

Mean differences between the linear and Gaussian approach were statistically quite similar ($p > 0.05$). It is interesting to note that although its mean difference value was -.9923 the confidence interval revealed that it could have been zero; while for mean differences where the polynomic kernel was involved, confidence intervals did not contain

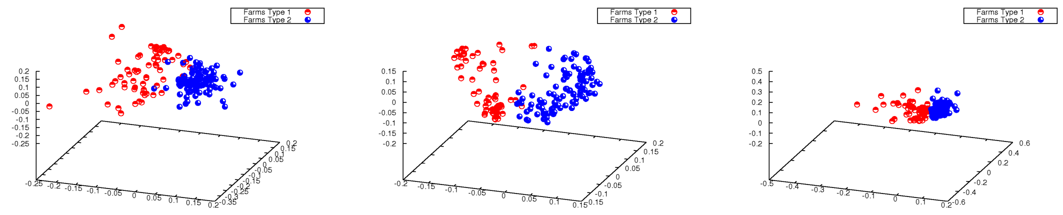
Table 3.4. Aragua-Guarico paired sample test

	Paired Differences							
			Std. E.	*95% conf. interval*				
	Mean	*Std. Dev.*	*Mean*	*Lower*	*Upper*	*t*	*df*	*Sig.*
LPCA-GKPCA	-.9923	8.80237	.67912	-2.3330	.3485	-1.461	167	.146
LPCA-PKPCA	4.9435	5.31537	.41009	4.1338	5.7531	12.055	167	.000
GKPCA-PKPCA	5.9357	8.05168	.62120	4.7093	7.1621	9.555	167	.000

LPCA: Linear principal component analysis
GKPCA : Kernel principal component analysis (Gaussian)
PKPCA: Kernel principal component analysis (polynomic)

zero as a potential value of the true difference. Moreover, given a Pearson's correlation coefficient of $r = .70$ between linear and polynomic approach, the magnitude of the observed effect was large according to the criterion of Cohen (1992).

Fig. 3.3 illustrates the reasons why previously displayed statistics make sense. As can be seen, the three feature extraction approaches lead to different configurations of class-centroids covariance matrix, and to differences between-class variance. Such configurations showed that separation between groups ocurred in the same direction of the high variance principal directions, and yielded different level of between-class overlapping after a discriminant analysis (Hastie et al., 2001). Despite the fact that two groups emerged with the Gaussian kernel, there was problematic overlapping which prevented separation of these groups into two distinct classes. Conversely, with the polynomic approach a clear direction was much more evident with the projected centroids and their respective within-class covariance matrix showing the minimum class overlapping (Jolliffe, 2002).



(a) (b) (c)

Fig. 3.3. Projection onto the three first principal components by farm's class of Aragua-Guarico data, for different approaches of feature extraction: linear (a); Gaussian (b); and polynomic kernel.

Additionally, it is also interesting to note how the polynomic kernel uncovered general features that permited the characterization of distinct groups of farms in such a way that resulted in good class separation while also yielding a more compact (high proximity) clusters-covariance which was confined in terms of least sum of squares around relative small distances to its centroid space. This might explain the higher misclassification rate observed in projections resulting from linear and Gaussian approaches, where the covariance matrix was much more scattered with respect to the mean differences between classes, preventing the definition of a sub-space with the easily separated categories.

3.5.1.3 Summary of farm's classes (Aragua-Guarico set)

At first glance, these results of unsupervised farms' classification after nonlinear feature extraction, seems to confirm that farm classes that were found in Chapter 2 by uing linear PCA procedures, are actually present in the Aragua-Guarico dataset. The predominance of growing-feeding systems (class 1) in Aragua's county, and cow-calf systems (class 2) in Guario's counties appear to be true. The main reason for this is that data were linearly separable (Duda et al., 2001; Schölkopf and Smola, 2002). As a consequence, many of the attributive characteristics of farm types remain the same as described in previous chapter for this experimental group.

3.5.2 Guarico set

3.5.2.1 Feature extraction

The idea of proceeding with further cluster partitions on previously found groups into successively optimal farms sub-groups, responds to the stakeholder necessity of being more focused at local issues for policy decisions, intervention plan formulation and investment. However, going deeply into a particular geographical region incorporates new challenges for feature extraction as objects (farms) from different categories are more difficult to differentiate; particularly when taxonomic classes are linked to spatial entities as farms to land areas, given that objects are more similar each other as they become closer (Burrough and McDonnell, 2005).

Guarico farms constituted a sample of 103 households and its partitioning into sub-categories constituted a very complex task. On the one hand, several shared conditions of topography, vegetation and climate; on the other hand, inner land use and production is fully surrounded by non-linear relationships that made it difficult to select measurement whose values would incorporate the complexity of the group. As a consequence, non-linear procedures were implemented to distinguish features that were invariants within a given category of farm. Table 3.5 shows the results of several statistics tests of a discriminant analysis applied to Guarico data clustered into three classes via hierarchical methods. In this case the objects were grouped based on the definition of similarity supplied by two kernel methods of feature extraction (Gaussian and polynomic); and the feature vectors were recursively increased from 1 to 6 in order to verify the minimum information required for the classification task.

Table 3.5. Impact of selected features on clustering performance using two different kernel (Gaussian and polynomic) approaches of feature extraction, after stepwise discriminant analysis for Guarico group data, as number of retained principal components (PC) in the hierarchical classification was increased.

	Gaussian ($\sigma = 8$)						*Polynomic* ($order = 2$; $\sigma = 80$)					
PC	$\%C$	$W\lambda$	PT	T^2	RM	r^2	$\%C$	$W\lambda$	PT	T^2	RM	r^2
1	75.7	0.27	0.78	2.45	2.36	0.39	86.4	0.20	0.92	3.35	3.16	0.46
2	81.6	0.22	0.95	2.54	2.16	0.47	91.3	0.13	1.26	3.52	2.13	0.63
3	82.5	0.20	1.02	2.06	2.13	0.51	88.3	0.14	1.24	3.38	2.06	0.62
4	74.8	0.24	1.00	2.08	1.27	0.50	91.3	0.15	1.21	3.15	1.84	0.60
5	90.3	0.09	1.38	3.50	2.43	0.69	91.5	0.11	1.31	3.83	1.94	0.65
6	81.6	0.26	0.96	1.91	1.19	0.48	83.5	0.16	1.16	2.94	1.93	0.58

%C: percentage classified correct; $W\lambda$: Wilks' lambda; PT: Pillai's trace; T^2: Hotelling's test; RM: Roy's minimum root; r^2: squared average canonical correlation

As can be appreciated, the maximum variability explained by the model was slightly more than 70% for the Gaussian kernel and 65% for polynomic. Overall, accuracy was quite good for all the experiments but the maximum systematic variability explained by farm classification was achieved with the Gaussian kernel given the comparatively low value of Wilks' lambda tests and the high value of the Pillai, Hotelling and Roy test. In summary, the best class separation and the maximum variability explained by classes is achieved using five feature extracted vectors in both kernel functions.

3.5.2.2 Clustering performance

A comparison between the best performing configuration of kernel methods and the linear approach whose feature extraction required six principal directions (see Chapter 2) is presented in Table 3.6. As can be seen, the profiles of clustering performance after discriminant analysis for the Gaussian kernel denote that means of farm classes on the selected variables were different in the population given the closeness of Wilks' lambda statistic to zero; and the comparatively higher value of the Pillai, Hotelling and Roy tests with respect to the linear and polynomic approaches. Also, classification based on Gaussian feature extraction, showed the higher average squared canonical correlation (r^2) supporting the idea of well separated groups accounting for a high percentage (69%) of the total variance explained.

Table 3.6. Impact of kernel function on clustering performance using linear, Gaussian and polynomic approaches of feature extraction, after stepwise discriminant analysis for Guarico group data.

Kernel	$\%C$	$W\lambda$	PT	T^2	RM	r^2
Linear	88.3	0.15	1.21	3.36	2.30	0.60
Gaussian	90.3	0.09	1.38	3.50	2.43	0.69
Polynomic	91.5	0.11	1.31	3.83	1.94	0.65

%C: percentage classified correct; $W\lambda$: Wilks' lambda; PT: Pillai's trace; T^2: Hotelling's test; RM: Roy's minimum root; r^2: squared average canonical correlation

The percentage of farms classified correctly was slightly higher when feature extraction was performed by polynomic kernel compared to insert linear and Gaussian kernel approaches. However, this feature extraction method did not provide enough information to find directions in the feature space along which farm groups were as well separated as with the Gaussian kernel. Even though, its performance was much better than classification based on linearly extracted feature vectors.

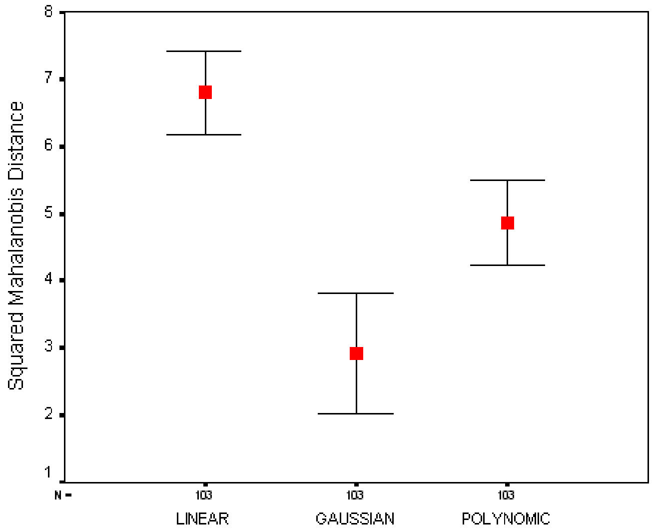


Fig. 3.4. Adjusted means and confidence intervals (95%) of squared Mahalanobis distance by selected feature extraction methods (linear, gaussian and polynomic) for Guarico group, after stepwise discriminant analysis.

Within canonical discriminant analysis, if a farm belongs to a particular class, it must fulfill some distance constraints with respect to this class' centroid, and it is expected that projections of these groups onto some discriminant direction are compact and show minimum overlapping. Hence, an easy way to asses the compactness of a given class is to look at the proximity of an observation set to its class-centroid. As reference, Fig. 3.4 displays the adjusted means and their respective confidence intervals of squared Mahalanobis distances for farm objects with respect to their class-centroid in clusters yielded under different methods of feature extraction.

Table 3.7. Guarico paired samples statistics.

		Mean	*N*	*Std. Deviation*	*Std. Error Mean*
Pair 1	LPCA	6.7971	103	6.10915	.60195
	GKPCA	2.9136	103	2.54947	.25121
Pair 2	LPCA	6.7971	103	6.10915	.60195
	PKPCA	4.8553	103	5.95678	.58694
Pair 3	GKPCA	2.9136	103	2.54947	.25121
	PKPCA	4.8553	103	5.95678	.58694

LPCA: Linear principal component analysis
GKPCA: Kernel principal component analysis (Gaussian)
PKPCA: Kernel principal component analysis (polynomic)

If mean distances and their confidence intervals between methods are considered, it can be seen that the linear approach is several orders of magnitude bigger than means achieved with Gaussian and polynomic kernels. This contrasts with the idea of well separated groups given that highly spread clusters tend to overlap each other. Feature extraction performed with the Gaussian kernel, in turn produced a cluster whose mean distance described more compact groups topology with low risk of overlapping; while polynomic kernel showed means at an intermediate position within this distance scale. A paired comparison statistics of these means distances is summarised in Table 3.7 along with their respective spread of the average variability and standard error of the mean.

Overall, all standard error resulted reasonably low for the three methods. Nevertheless, it can be observed that the resulting average variation of the sample mean for the Gaussian kernel was the lowest, suggesting that the means (observed by classes) yielded with the information provided under this method were very close to their population means. Differences between means by feature extraction method are displayed in table 3.8 where, additionally, results from t-test are presented.

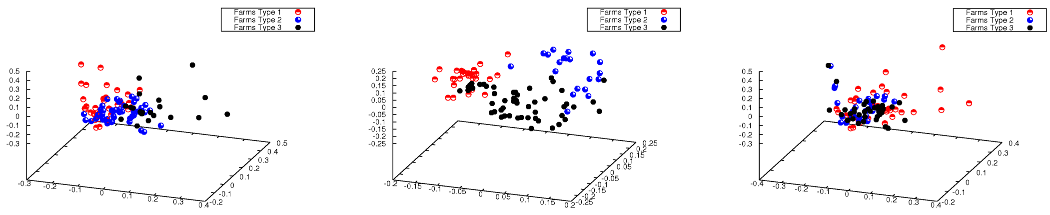
Table 3.8. Guarico paired sample test

	Paired Differences							
				95% conf. interval				
	Mean	*Std. Dev.*	*Std. E. Mean*	*Lower*	*Upper*	*t*	*df*	*Sig.*
LPCA-GKPCA	3.8835	7.16050	.70555	2.4840	5.2829	5.504	102	.000
LPCA-PKPCA	1.9417	4.56512	.44981	1.0495	2.8340	4.317	102	.000
GKPCA-PKPCA	-1.9417	7.24327	.71370	-3.3574	-.5261	-2.721	102	.008

LPCA: Linear principal component analysis
GKPCA : Kernel principal component analysis (Gaussian)
PKPCA: Kernel principal component analysis (polynomic)

It can be appreciated that cluster segmentation based on Gaussian kernel feature extraction showed the highest mean differences ($p < 0.001$) when compared with linear and polynomic approaches. It is obvious that the Gaussian kernel extracted features brought out the underlying semantic content of the classes that led to compact clusters characterised by very small mean distances between cluster-objects and their centroids.

A graphic representation of this effect is illustrated in Fig. 3.5, where farm objects were projected onto their first three principal directions with different levels of class overlapping for different feature extraction methods used. It can be appreciated that just one of the three algorithms (Gaussian kernel) lead to a classification model that described in a suitable way (without overlapping) the groups sugested by the instances cloud. Results from linear and polynomic-kernel methods presented no desirable characteristics for cluster separation. This is mainly due to the topology of the sample covariance



(a) (b) (c)

Fig. 3.5. Projection onto the three first principal components by farm's class of Guarico data, for different approaches of feature extraction: linear (a); Gaussian (b); and polynomic kernel (c).

matrix as a result of the effect that the feature extraction method had on class-objects components coordinates.

Linear and polynomic-kernel methods produce regions with low density of instances surrounding a central high densed region where all classes converge, generating a poor quality cluster representation. Conversely, the model obtained with the Gaussian kernel showed a sample covariance matrix structure where labelled clusters can be perceived and the representation scattered along several directions in such a way that permitted classes' separation. Of course, some unexpected items from one class appeared wrongly located into the hypothesized space of another; however, this probably arose from the fact that some information was lacking, since this feature extraction was originally performed with five principal directions and only three were used for this plot.

3.5.2.3 Summary of farm classes (Guarico set)

In contrast to results for clustering under linear feature extraction for Guarico group in Chapter 2, which showed poor accuracy and highly overlapping classes, clustering after nonlinear feature extraction yielded clear class separation and high classification accuracy. Three farm classes were identified within the data structure: farms type 1 (cow-steer system), farms type 2 (cow-store systems), and farms type 3 (cow-calf systems). All these categories share some common attributes according to the classification of Seré and Steinfeld (1996). For instance, the three farm classes occupy a humid-subhumid region of tropical lowland, and livestock production is based on mixed systems. On the other hand, there are several attributes that make them integrate different groups.

To describe each farm category from an input space viewpoint after a classification based on nonlinear mapping, could lead to certain inaccuracies; nevertheless, in this case such an approximation resulted suitable. In this sense, Fig. 3.6 and Table 3.9 show box plots and a statistics summary respectively. As can be seen, variables are rendered by class and the table presents some metrics, that apart from the mean and standard deviation, also include other central tendency and dispersion metrics in order to avoid those biases associated with the presence of outliers. Based in this information the main profile of each farm category is described as follow:

Farms type 1 (cow-steer system):

These farms were characterised by a livestock feeding system based on cash crops, where the proportion of farmers growing sorghum was the most important when related with the other two classes. For example, trimmed means of variable SOR in Class 1, was 76 and 88 % higher than class 2 and 3. Variable MAI showed also higher values within this category. On the other hand, milk production per cow^{-1} and per hectare^{-1}, were also superior for this class; and an important proportion of farmers, given the modality showed by variables PWM and HRC within this category, appear to keep young animal for fattening purposes. In contrast to the tendency observed with the classification achieved after linear feature extraction; the profile of this farm class is consistent with the work of Renard (1997); Ruthenberg (1980) and Entz et al. (2005).

Farms type 2 (cow-store systems):

The livestock feeding system within this typology, appear dominated by natural ecosystems such as forest with low anthropomorphic intervention. The percentage of forest (FOR) within the farm, seems to be one of the variables responsible for clustering together this category, with a trimmed mean for this variable 69 and 54% higher than for class 1 and 3 respectively. Also, these farms showed moderate milk production per cow^{-1} and the lowest per hectare^{-1} with a moderate bovine density (0.40 AU/ha). In this sense, this farm profile agrees with the findings of Scarnecchia (1985) and Coppock (1994) with respect to stocking rate and milk production. It is informative to note that variables SOR and MAI observed lower values for this group respect farm type 1 (61 and 71%) and higher related to farms type 3 (12 and 28%). This indicates that even when cereals contribute more than 10% of the livestock feeding biomass, the cash crop activity is still incipient (Renard, 1997).

Farms type 3 (cow-calf systems):

In this type of farm the livestock feeding system is mostly supported by introduced and native gramineous life forms, with the highest levels of variable PAF, and moderate presence of forest within their inner land cover. These farms also show a strong tendency to sell young animals after weaning, since a high proportion of farms within this group, showed low values for variable PWM, and this is expected given the their low carrying capacity of livestock as a result of the small proportion of inner land-cover dedicted to cash crops, as can be appreciated from modalities showed by variables SOR and MAI in the boxplots.

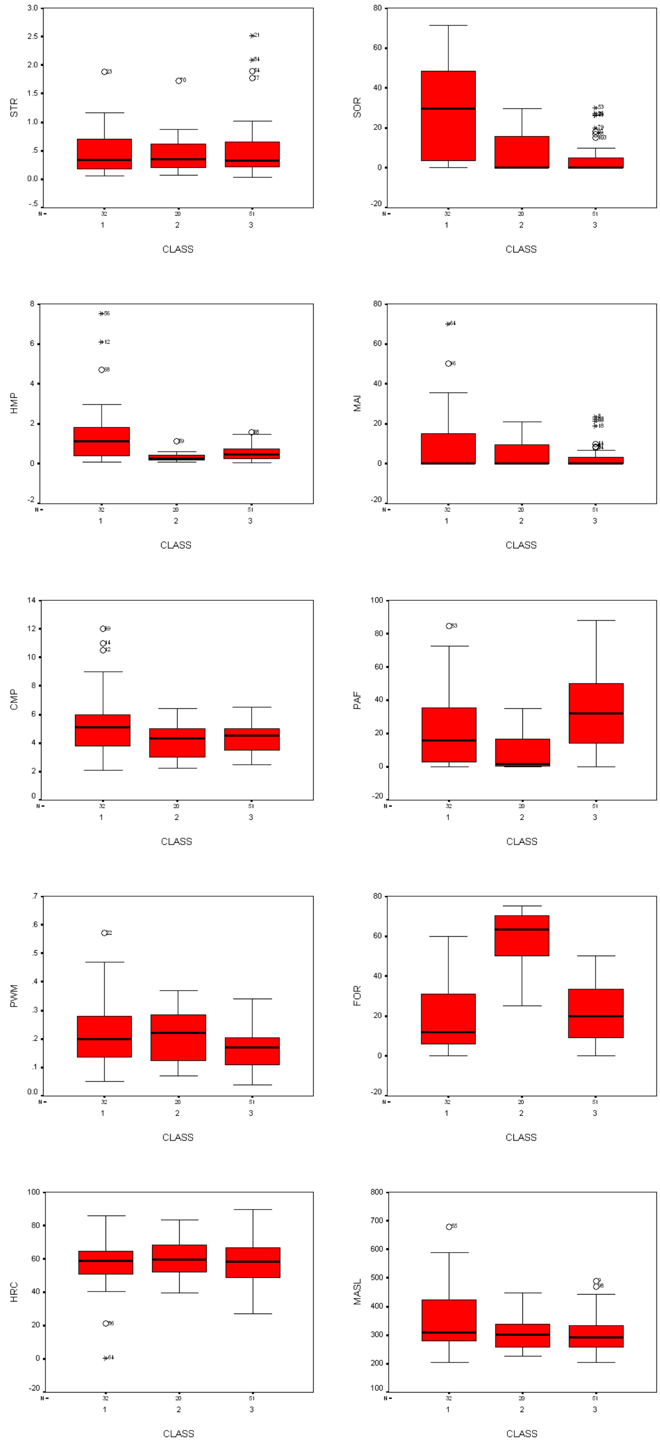


Fig. 3.6. Boxplots of productive and land cover attributes variations by farm class.

Table 3.9. Summary statistics of 138 crop-livestock systems attributes by farm's class.

			STR		HRC		PWM		CMP		HMP		MAI		SOR		PAF		FOR		MASL	
CLASS			Stat.	S. E	Stat.	S. E	Stat.	S. E	Stat.	S. E	Stat.	S. E	Stat.	S. E	Stat.	S. E	Stat.	S. E	Stat.	S. E	Stat.	S. E
1	Mean		.4798	.0708	56.18	2.820	.2216	.0216	5.4294	.4364	1.5328	.2988	9.8478	2.938	29.90	4.159	21.71	3.992	19.32	3.369	358.4	19.59
	95% Confidence Interval for Mean	Lower Bound	.3354		50.43		.1775		4.5393		.9234		3.8563		21.42		13.57		12.45		318.4	
		Upper Bound	.6242		61.93		.2657		6.3195		2.1422		15.84		38.39		29.85		26.19		398.3	
	5% Trimmed Mean		.4388		57.35		.2134		5.2681		1.3101		7.4698		29.29		19.67		18.13		350.2	
	Median		.3419		58.66		.2000		5.1250		1.1350		.0000		29.50		15.72		11.84		310.1	
	Variance		.160		254.4		.015		6.095		2.857		276.2		553.5		510.0		363.3		12280	
	Std. Deviation		.40061		15.95		.12235		2.469		1.690		16.62		23.53		22.58		19.06		110.8	
	Minimum		.06		.44		.05		2.07		.08		.00		.00		.00		.00		203.69	
	Maximum		1.88		85.71		.57		12.00		7.50		70.00		71.43		84.58		60.00		679.19	
	Range		1.82		85.27		.52		9.93		7.42		70.00		71.43		84.58		60.00		475.50	
	Interquartile Range		.5591		14.19		.1475		2.1925		1.4575		15.42		48.41		33.81		25.87		144.3	
	Skewness		1.712	.414	-1.477	.414	.952	.414	1.133	.414	2.281	.414	2.211	.414	.181	.414	1.115	.414	1.025	.414	1.074	.414
	Kurtosis		3.532	.809	4.279	.809	.942	.809	1.086	.809	5.442	.809	5.199	.809	-1.081	.809	.791	.809	-.187	.809	.912	.809
2	Mean		.4605	.0864	59.48	2.797	.2085	.0212	4.1490	.2577	.3290	.0536	4.8125	1.471	7.6780	2.272	8.3640	2.446	58.92	2.940	307.9	14.02
	95% Confidence Interval for Mean	Lower Bound	.2796		53.62		.1641		3.6096		.2168		1.7330		2.9221		3.2437		52.76		278.5	
		Upper Bound	.6413		65.33		.2529		4.6884		.4412		7.8920		12.43		13.48		65.07		337.2	
	5% Trimmed Mean		.4116		59.27		.2072		4.1272		.2994		4.1806		6.8906		7.3578		59.91		304.7	
	Median		.3525		59.42		.2200		4.3500		.2550		.0000		.0000		1.6200		63.18		302.5	
	Variance		.149		156.5		.009		1.328		.057		43.294		103.3		119.7		172.9		3932	
	Std. Deviation		.38646		12.51		.09483		1.153		.23969		6.580		10.16		10.94		13.15		62.71	
	Minimum		.07		39.44		.07		2.25		.07		.00		.00		.00		25.00		227.11	
	Maximum		1.73		83.33		.37		6.44		1.12		21.00		29.53		34.84		75.00		446.67	
	Range		1.66		43.89		.30		4.19		1.05		21.00		29.53		34.84		50.00		219.56	
	Interquartile Range		.4452		17.17		.1700		2.0075		.2625		9.7725		16.25		16.42		20.44		86.24	
	Skewness		1.946	.512	.264	.512	.092	.512	.164	.512	2.013	.512	1.137	.512	.945	.512	1.188	.512	-.946	.512	.886	.512
	Kurtosis		5.303	.992	-.488	.992	-1.254	.992	-.741	.992	5.466	.992	.315	.992	-.591	.992	.269	.992	.626	.992	.663	.992
3	Mean		.5253	.0736	58.83	2.072	.1663	.0100	4.2757	.1592	.5475	.0510	3.0380	.8543	4.4592	1.192	33.82	3.178	21.64	2.152	308.3	9.382
	95% Confidence Interval for Mean	Lower Bound	.3776		54.67		.1461		3.9559		.4450		1.3222		2.0644		27.44		17.31		289.5	
		Upper Bound	.6731		62.99		.1864		4.5955		.6499		4.7539		6.8541		40.21		25.96		327.2	
	5% Trimmed Mean		.4572		58.86		.1642		4.2508		.5219		2.1221		3.3915		32.89		21.28		304.3	
	Median		.3274		58.33		.1700		4.5000		.4800		.0000		.0000		32.18		20.00		293.1	
	Variance		.276		219.0		.005		1.293		.133		37.220		72.503		515.2		236.2		4489	
	Std. Deviation		.52537		14.80		.07161		1.137		.36426		6.101		8.515		22.70		15.37		67.00	
	Minimum		.04		27.18		.04		2.50		.06		.00		.00		.00		.00		205.18	
	Maximum		2.52		89.66		.34		6.50		1.56		23.53		30.00		88.00		50.00		489.14	
	Range		2.48		62.48		.30		4.00		1.50		23.53		30.00		88.00		50.00		283.96	
	Interquartile Range		.4636		18.02		.1000		1.5000		.5000		3.3300		5.0000		35.88		25.33		83.28	
	Skewness		2.234	.333	.013	.333	.351	.333	.212	.333	1.007	.333	2.358	.333	1.893	.333	.474	.333	.162	.333	1.022	.333
	Kurtosis		5.197	.656	-.582	.656	-.456	.656	-.791	.656	.544	.656	4.895	.656	2.403	.656	-.460	.656	-1.072	.656	.388	.656

3.6 Conclusion

A kernel based generalisation of principal components analysis was applied in this research in order to accomplish the feature extraction of real crop-livestock data. The main results show that the kernel approach exhibited a superior performance with respect to the linear method for farms within the Guarico set. In particular, it has been demonstrated that such superiority could be obtained even with limited training data availability and in highly similar geographical conditions shared by all the farms within this subset. Additionally, the segmentation achieved from features extracted by Gaussian kernel, was superior with respect to polynomial in terms of classification accuracy and explained variance. It is noteworthy that compared to linear and polynomial functions, the expressivity of cluster representation from features generated by Gaussian kernel, showed more clearly defined decision regions, with wider separation of gravity centers by class, and much more compact sets, from within-group distances point of view; overcoming, in that way, the limitations ecountered in Chapter 2, with which farm labelling generated in this chapter can be used to perform supervised classification of farms' spectral response.

Chapter 4

Kernel Based Supervised Classification of Crop-Livestock System from Multi-Spectral Images

4.1 Abstract

The main focus of this chapter is to classify the spectral-response[1] of land cover as seen in a Landsat7 ETM image, using several farm categories proposed in chapters 2 and 3. This supervised farm classification method is based on the kernel-adatron (KA) algorithm, which can produce the separation of two farm classes by an optimal decision boundary. This decision function is defined by a linear separating hyperplane in a general feature space. Nonlinearities are handled by mapping the input data into a multidimensional feature space induced by a kernel function. Results suggest that accurate farm classification based on spectral characteristic recorded in a satellite image is possible, using a small training set (20 instances). Accuracy on classification achieved for nonlinearly separable data by KA algorithm, was higher than discriminant analysis (89 vs 55 %). The use of Gaussian led to much more accurate classification with minimun number of instances required when conpared to polynomial kernel. These findings also reveal that repeatable relations between biophysical and spectral features can be derived from abstractions as difficult to observe as farms.

[1] A general term referring to the amount of light reflected by a surface at different wavelengths as seen in an image (Drury, 2001).

4.2 Introduction

Farm classification is one of the most important approaches to retrieving information about agricultural production systems. The use of remote sensing technology has been gaining popularity for studying these systems, due to 1) the possibility of acquiring data for very large areas in short time periods, 2) its good performance in recording land cover effectively, 3) the provision of data that are spatially explicit and 4) the probability of repeating surveys periodically.

Spatial land cover classification has been mainly approached through the following paradigms: maximum likelihood classifier (MLC) (Strahler, 1980); fuzzy clustering (Kosko and Isaka, 1993); and artificial neural networks (ANN) (Miller et al., 1995). However, farm classes are abstractions which are sometimes difficult to observe directly, and this leads to a number of limitations of these methods. For instance, MLC methods are not free from distribution assumptions, given their parametric premises. Fuzzy clustering represents the solutions in terms of probabilities, where both fuzzy rules and membership functions are subjected to the bias of the interpreter. The ANN method has theoretical weaknesses because of its black box character, preventing the proper repeatability of the results. The presence of local minima and of the time-consuming training process (referred to as lack of convergence) are also significant limitations.

Recently some attention has been paid to the use of linear machines (Vapnik, 1995). This has eliminated many of the limitations mentioned above by introducing efficient learning algorithms which identify a linear optimal hyperplane that maximizes the separation between any two classes and addresses nonlinearities mapping input data into a multidimensional feature space induced by a kernel function (Aizerman et al., 1964). Applications of these algorithms have been described by García and Moreno (2004b) on magnetic resonance image segmentation in the sphere of medicine; and by Huang et al. (2002) and Zhu and Blumberg (2002) on the land cover classification domain, showing the method to be robust.

The present chapter extends the use of the linear machine known as the kernel-adatron algorithm (Friest et al., 1998), with the following objective: firstly, to classify the spectral-response of land cover as seen in a Landsat7 ETM image, using several farm categories; secondly, to test the effect of different kernel functions and their parameters on the accuracy of farm classifications; and thirdly, to compare this methodology with standard procedures for supervised classification as discriminant analysis.

The plan of this chapter is as follows. First a brief overview of the kernel-adatron algorithm is given. Secondly a description of the data and experimental design is provided, followed by a brief results and discussion section and finally, the conclusions are summarized.

4.3 The kernel-adatron method

The "hybrid" algorithm known as the kernel-adatron, first proposed by Friest et al. (1998), can be elaborated by examining the development of linear classifiers. Generally speaking, a linear classifier is based on a linear decision function whose estimated output is given by $y = f(\vec{x} \cdot \vec{w}) = f(\sum_i x_i w_i)$, where $\vec{x}$ represents the input feature vector to the classifier; $\vec{w}$ is the vector of weights defining the separating boundary and f is a function that projects input values x on w. In this way input patterns are linearly separated by dividing the input space with a hyperplane (Fig. 4.1).

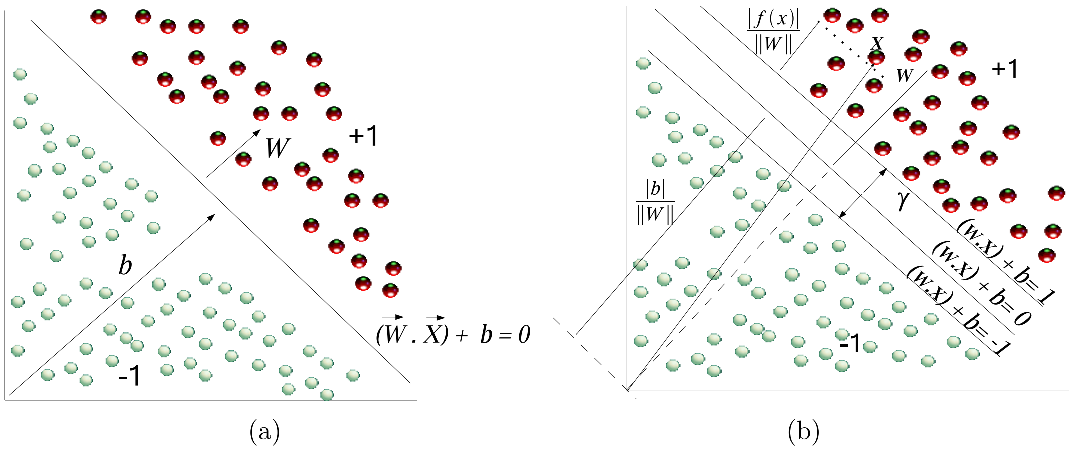


Fig. 4.1. Toy example of decision boundaries between classes; modified from Cristianini and Shawe-Taylor (2000).

There are two main practical approaches to induce linear classifier parameters, on the one hand those methods based on modelling conditional density functions (generative models) such as: linear discriminant analysis (Fisher, 1936; Lachenbruch, 1975) and Naive Bayes Classifier (Domingos and Pazzani, 1997). On the other hand, there are those that pursue the maximization of the outputs quality over a training set (discriminative models). These devices include: logistic regression (Hosmer and Lemeshow, 2000); perceptron (Rosenblatt, 1958) and support vector machine (Vapnik and Chervonenkis, 1974; Vapnik, 1995).

The main characteristics of support vector machines, is that they find a maximal margin hyperplane (Fig. 4.1). This is achieved using optimization procedures that in some cases place severe demands on the computational task. These problems were central to developing the kernel-adatron method which takes advantage of the adatron simplicity (Anlauf and Biehl, 1989), generalizing it to operate in a high dimensional feature space by the introduction of kernels functions. It solves the optimization problem of the Lagrangian formalism performing the margin-maximization through the application of a gradient ascent algorithm, resulting in an enhanced capability to learn nonlinear boundaries with a rate of convergence that is exponentially fast.

4.4 The Landsat data

The use of multispectral data to distinguish one type of land cover from another, has been an effective way of linking anthropomorphic intervention to a physical environment; particularly whithin the agricultural sector (Campbell, 2002). For instance, Wylie et al. (2002) combine optical and thermal data to estimate biophysical properties in vegetation. Other approaches use the land cover mosaic, to induct farm typologies based on their relative spectral similarities, as in the case of Duvernoy (2000).

The popularity of using visible and near infrared (VIR) imagery on the classification of areas covered by agricultural activities, is associated with the fact that plant cell structures, morphology, chlorophyll and other pigments have a marked effect on this wavelength range (Drury, 2001); and also on the temperature brightness of incident thermal infrared (TIR) radiation upon living plants (Rees, 2007).

The configuration of Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor, is particularly well suited to perceive the energy field, in the form of VIR and TIR radiation emanating from vegetation covers (Richards and Jia, 2006). This peculiarity makes Landsat data sensible to spatial patterns tied to crop calendar, and vegetative growth-lessening as a result of phenophases (Campbell, 2002; Richards and Jia, 2006). The spectral bands per pixel in Landsat sensor, account to 7, delineated by six VIR bands, where band 6 is split into two channels defined by filters that control the radiance that reach the sensor; and a panchromatic band (Barsi et al., 2003; Heckenlaible et al., 2007).

Another aspect that presents Landsat 7 as a good choice within the context of this research, is its radiometric resolution. The precision at which this sensor registers the radiation power, for a particular pixel in a given wavelength is 8 bits (256 levels) (Richards and Jia, 2006). This feature enhances the ability to distinguish the spectral responses from different materials, when human-scale factors such as agriculture need to be adressed (Campbell, 2002; Landgrebe, 2007).

Like radiometric resolution; the spatial resolution of Landsat 7, which ranges from 15 to 60 meters per pixel across all the spectral bands, is rich (small or fine) compared to farms, which are the objects under study in this research; and in farming, a pixel samaller than the agicultural field to be studied is usually preferred (Landgrebe, 2007). To these spatial characteristics of landsat, should be added its scanning features, whose cover swath is 185 km^2, which means that each scene sample observes an area of 34.225 km^2. Such an overlay represents an advantage for this research given the scale of the study area (7.730 km^2), and because of the fact that the whole data can be extracted from one image.

Traditionally multispectral data such as Landsat series, have been used on photogeology; geobotany, forestry, agriculture, soil and land surface mapping. They have rarely been used to recognize continuous pixels groups integrated in a class such as a farm, which is a mosaic of land covers. However, due to the use in this research of kernel methods coupled to a maxim margin classificator, where the resultant representation is flexible, uniform over the pattern presented and preserves the topology of the input space, the use of multispectral data like Landsat seems natural, on the problem of discriminate farm types using as indicators their land cover spectral response as recorded in a satellite image.

4.5 Data preprocessing and methods

4.5.1 Informational classes

The data used in this study was assembled on the basis of the unsupervised classification described in chapters 2 and 3. This classification was built on a sample of 168 households, from a population of 1321 farms. A description and account of the labeling used for each informational class achieved in that part of the study, is presented in Table 4.1. These categories were delineated by hierarchical procedures, after diferent feature extraction approaches. Attributive data used to this end, comprised 10 variables (Table 2.1, chapter 2). These attributes were selected because they provided information on some of the two more studied constructs in the farming system field: livestock productive-reproductive management, and land cover management.

As in chapters 2 and 3, sampled data were also partitioned into two comparison groups: group 1 included the whole 168 households corresponding to both states (Aragua-Guarico states); and group 2 (Guarico state), comprised a sub-set of 103 farms from the former Aragua-Guarico group (Fig. 2.2, chapter 2). Data separation permitted farm

Table 4.1. Description of the informational classes used for supervised classification on experimental group 1 (Aragua-Guarico set) and 2 (Guarico set)

Class	*Farm System*	*Definition*	*Ground trait*
		Aragua – Guarico Group	
1	growing-feeding	Apart from dairy production, raise or purchase weaned calves that are fattened until a suitable weight to be slaughtered is reached	Crop land Forested land Plowed lands shrubs, trees
2	cow-calf	Main activity is dairy production, and young animals are sold at different stages: during calfhood, or at some point after weaning	Isolated forest Grassland areas Crop land Isolated shrubs, trees
		Guarico Group	
1	cow-steer	Main activity: cash crops and dairy production. Livestock feeding system based on annual crops. Keep young animal fattening until slaughtered	Crop land Plowed land Isolated grassland Isolated Forest
2	cow-store	Main activity: dairy production, and raising calf after weaning to be sold. livestock feeding system dominated by forest and grassland	Dense forest Sparse grasland Isolated cropland
3	cow-calf	Dairies with strong tendency to sell young animals after weaning. feeding system relies on forest, introduced and native gramineous.	Occasional cropland Open grassland Forested land Sparse shrubs

comparisons between two very different administrative and geographical units (Aragua and Guarico states) and also enable a more detailed exploratory analysis within Guarico state.

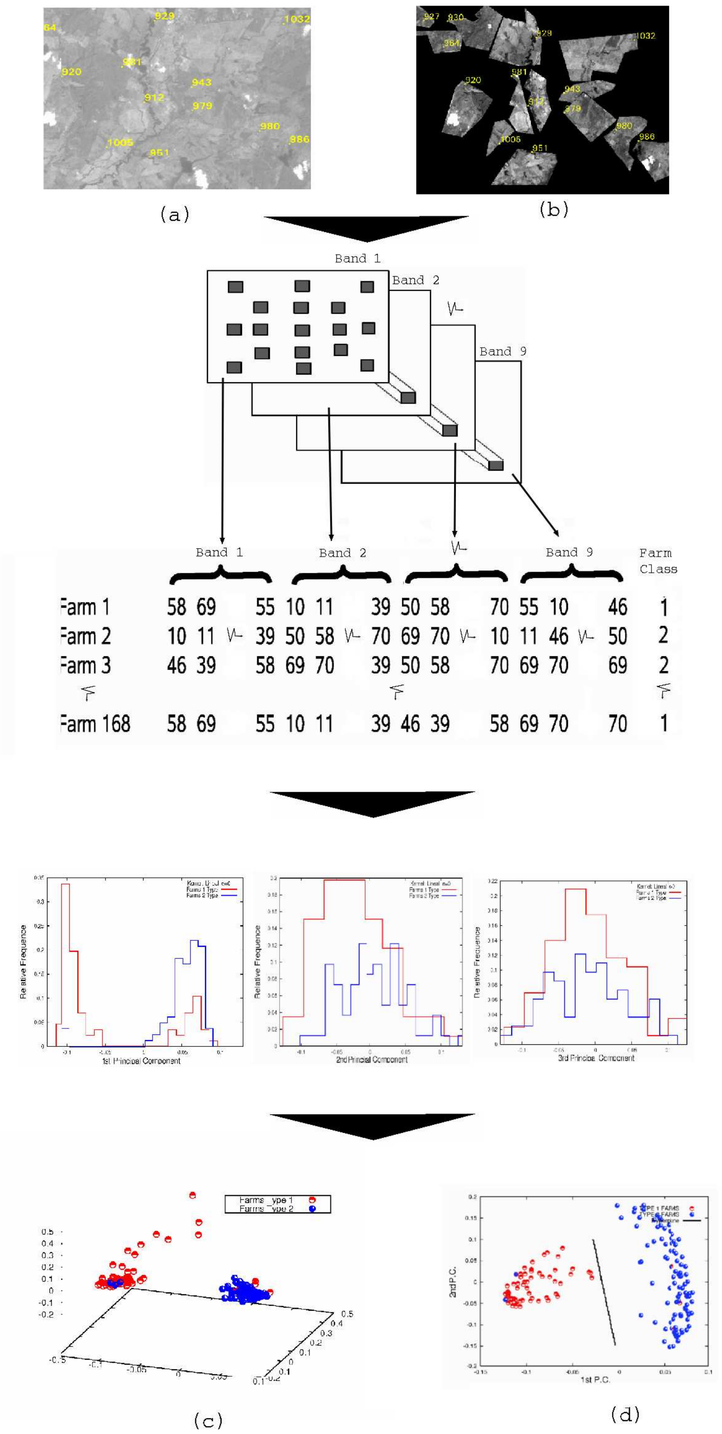
4.5.2 Data preprocessing

Each group of surveyed farms was then labeled according the cluster it belonged to and geo-referenced in order to identify its position on Landsat Enhanced Thematic Mapper plus (ETM+) images, acquired in November 1999. Each image covered an area of 170 km (north-south) x 185 km (east-west). The original formats (FST) were also geo-referenced from header files, associated projection and parameters were validated (UTM, WGS84); layer stack was created; and finally the image was radiometrically enhanced using linear functions via a look-up table.

The nine multi-band raster dataset was sampled producing a collection of pixel values over each band, following an amplified von Neumann vicinity in a pre-selected area of interest within the farm's perimeter (Fig 4.2). The sample for group 1 and 2 amounted to 168 and 103 farm spectral responses respectively, each with 180 component values. This training data set was used as input to a dimension reduction procedure, using principal component analysis with kernel (KPCA).

4.5.3 Data analysis

Each training instance for the learning machine consisted of three-dimensional vectors represented by the coordinates that each instance observed upon the three first principal directions resulting from: a linear (Aragua-Guarico set) and kernel (Guarico set)



From a Landst (ETM+) image (a) subset are created (b), following as clipping criterion the farms' perimeters or boundaries.

The Sampling is performed within each farm subset. The nine multi-band dataset is sampled producing 20 values over each spectral channel, resulting in 180 pixel values per farm instance.

After values are collected, a vector of 180 colums per farm is organised by the concatenation of the 20 pixel values coming from the nine spectral channels

The matrix of 180 columns is centered, and a KPCA performed in order to extract features and reduce dimensionality. Then two or three principal directions with potential for class separation are selected.

If data plotted in the subspace spaned by the selected principal directions (c), show that samples can be linearly separated with minimum errors; a linear machine is then trained with this data in order to find a separating hyperplane (d)

Fig. 4.2. Landsat image segmentation procedure.

Principal Component analysis (PCA). So given a training set $S = | (x_1, y_1), ..., (x_n, y_n) |$, $x_i \in \mathbb{R}^d$ labeled by $y \in \{1, \text{-}1\}$, depending on the class to which each data item belongs. Where x_i corresponds to a vector of instances that lives in the input space, with dimension d. The goal of the learning machine is to linearly separate this pattern set S inducing a hyperplane defined by (w, b), with a decision function $f(\mathrm{x})$ such that:

$$y = f(x) = (\langle w \cdot x_i \rangle + b) \qquad \forall (x_i, y_i) \in S \tag{4.1}$$

For solutions $f(x)$ to be viable, the functional margins[2] of the training data points (Fig. 4.1b), must all be positive:

$$\gamma_i = y_i(\langle w \cdot x_i \rangle + b) \geq 0 \qquad \forall (x_i, y_i) \in S \tag{4.2}$$

As an alternative, by considering normalized parameters w and b $\langle \frac{w}{\|w\|}, \frac{b}{\|\mathrm{b}\|} \rangle$, a more informative Euclidian expression of (4.2) is obtained. Such representation corresponds to what is usually referred to as the geometrical margin:

$$\Gamma_i = \frac{\gamma_i}{\|\mathrm{w}\|} = \frac{y(\langle w \cdot x_i \rangle + b)}{\|\mathrm{w}\|} \geq 0 \qquad \forall (x_i, y_i) \in S \tag{4.3}$$

which represents the geometrical distance of the data points from the separating hyperplane.

[2]Functional margin is the distance between a training example in feature space and the separating hyperplane.

From a geometrical point of view, the training algorithm, during the learning phase, performs a search on b and w spaces, seeking for values of these parameters fulfilling relations (4.2) or (4.3). It is noteworthy, that in general the main issue here is to find, within infinite options, the vector $\vec{w}$ that maximises Γ_i. Formally this problem is defined as:

$$\begin{aligned} \min \quad & \frac{1}{2}\|w\|^2 \\ \text{subject to} \quad & \gamma_i = y(\langle w \cdot x_i \rangle + b) \geq 1 \qquad \forall (x_i, y_i) \in S \end{aligned} \tag{4.4}$$

The idea of a margin maximization stated in (4.4) consists of an optimization problem that involves a quadratic objective function, subject to a linear constraint. The solution to this problem may be obtained by standard techniques such as the Lagrange method together with Quadratic Programming procedures (Vapnik, 1995). Conversely, the kernel-adatron (Anlauf and Biehl, 1989) stresses an alternative training algorithm using the gradient ascent to maximise the Lagrangian multiplier under linear constraints (Friest et al., 1998), and enabling high dimensional feature space analysis within a kernel framework. This results in a kernel machine as follows:

$$f(x) = \sum_{i}^{n} \alpha_i y_i K(x_i, x) + b \tag{4.5}$$

Where α_i are the Lagrangian multipliers resulting from the solution of the constrained optimization problem; and $K(x_i, x)$ is a problem-specific kernel function that takes account of the non linearities of the data.

4.5.4 Coping with nonlinearities

Using kernels to learn potential nonlinear representation hypothesis based on the function of the form stated in (4.5), essentially involved the simulation of the nonlinear projection of the input data in a higher dimensional space (Schaback and Wendland, 2006):

$$\begin{aligned} \Phi : S \in \mathbb{R}^d &\rightarrow \mathcal{F} \in \mathcal{H} \\ x &\mapsto \Phi(x) \end{aligned} \tag{4.6}$$

where $\mathcal{F}$ denotes a feature space; and, $\mathcal{H}$ represents a dot product space, within which, a learning relationship could be induced between a pattern $\Phi(x)$ and a label y. In this way, having as theoretical context Mercer's theorem (Mercer, 1909; Aizerman et al., 1964); (4.7) represents the kernel matrix, where each entry was a measure of similarity between two objects. Thus, a symmetric function $K(x_i, x)$ was a kernel if it fulfilled Mercer's condition, i.e. the function K is (semi) positive definite. When this is the case there exists a mapping ϕ such that it is possible to write $K(\boldsymbol{x}, \boldsymbol{y}) = \langle \phi(x_i) \cdot \phi(x) \rangle$.

$$K(x_i, x) \triangleq \langle \phi(x_i) \cdot \phi(x) \rangle \Rightarrow \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \dots \\ K(x_2, x_1) & \ddots & \\ \vdots & & \end{bmatrix} \tag{4.7}$$

The kernel represented a dot product on a feature space $\mathcal{F}$ into which the original vectors were mapped (Fig. 4.3). In this way a kernel function defines an embedding of memory patterns into (high or infinite dimensional) feature vectors and allows the algorithm to be carried out in this space without the need of representing it explicitly (Cristianini and Shawe-Taylor, 2000; Schölkopf and Smola, 2002).

Further details on the way this procedure was implemented is outside the scope of this paper. Nevertheless, for those seeking deeper understanding on the ideas behind kernel-based learning theory there are fuller descriptions in Mercer (1909); Aronszajn (1950); Aizerman et al. (1964) and Schölkopf and Smola (2002). Also, applications of kernel methods and learning machines may be reviewed in García and Moreno (2004a,b,c)

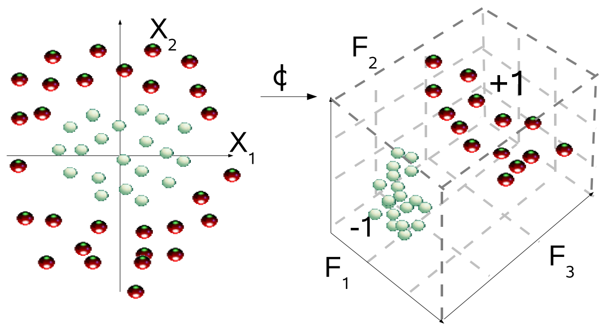


Fig. 4.3. Toy example illustration of the effect of mapping a simple binary problem to a higher dimensional feature space on the ability to separate complex relations.

4.5.5 Experiments

Data were divided into training and test set; the training sets used consisted of a predetermined number of instances randomly sampled from the total data. The sets were balanced, i.e: two (group 1) and three (group 2) classes were equally represented.

The validation set consisted then of the remaining points in the data set unseen by the learning machine during training.

The cardinality of the training set was established experimentally. The KA training procedure for the machines with linear, gaussian ($\sigma = 70$) and polynomial (order= 3; $\sigma = 4$) kernels, was repeated 20 times within a range of training set cardinalities equal to: 10, 20, 30, 40, 50, 60 and 70 instances for group 1; and 4, 10, 12, 14, 16 and 20 for experimental group 2. Then, the trained classifiers were validated and the mean number of inaccuracies (mistakes) and their standard deviations were computed.

After the training process, confusion matrices were built to compare the performance of the kernel machine per class and against the accuracy achieved using discriminant analysis as a standard technique traditionally used for supervised classification purposes.

4.6 Result and discussion

4.6.1 Dimensionality reduction

4.6.1.1 Aragua-Guarico set

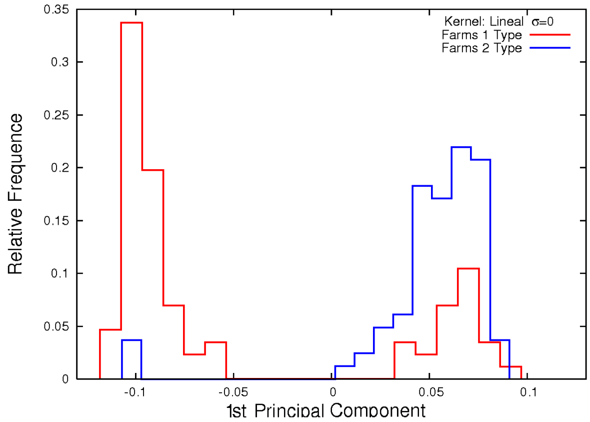
As could be observed in chapter 3, classes for this experimental group resulted linearly separable. In this section, patterns that were present in the analysis of the group 1 domain, showed that features extracted by linear approaches contained the dynamics of interest to achieve an important dimensionality reduction while maximising class separation. This indicate that classes not only were linearly separable from an attributive viewpoint, but also their spectral responses across the spectral channels of a Landsat image. Again as was demonstrated in chapter 3, the feature space provided by linear PCA was rich enough to guarantee appropriate levels of accuracy in spectral classes separation.

In Fig. 4.4 are presented the histograms of the first (a), second (b), and third (c) principal components for the experimental group 1, extracted by using a linear approach. It can be appreciated that both classes occurred in different portions of representation in all these principal directions. It is interesting to note that separation of the two farm categories was not entirely evident; even though an important class discrimation can be observed at the first direction, clusters appear to be overlapped on the second a third components.

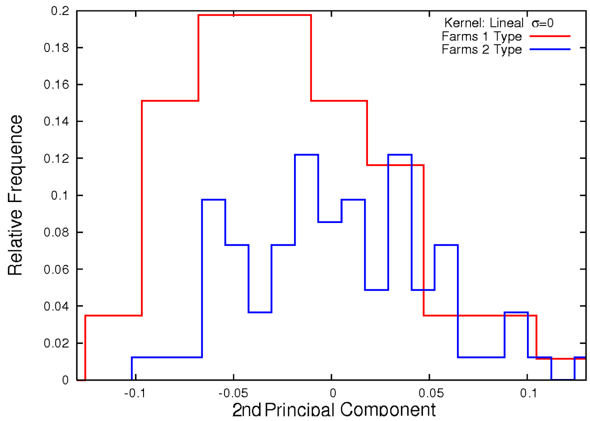
However, the subspace provided by the combination of these three directions (Fig. 4.5), leads to a better discrimination for the given informational classes. As can be seen, the linear transformation (a) of the signal space resulted in the definition of a subregion where a high cluster separation was achieved in the directions of highest variance in this vector space. On the other hand, it seems clear that Gaussian feature extraction (b) did not much improve the discrimination between clusters, showing that the fundamental modes of variation underlying this data categorisation were essentially linear.

4.6.1.2 Guarico set

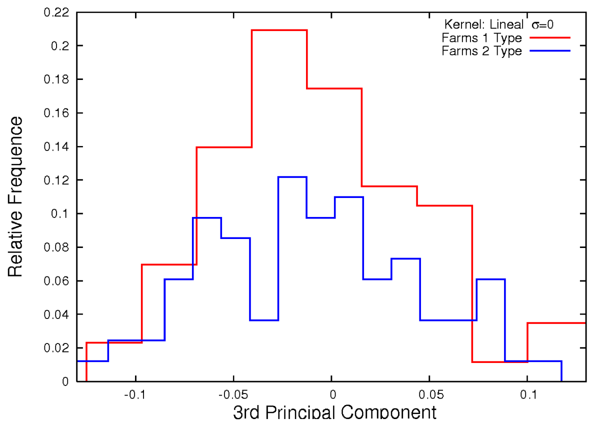
Fig. 4.6 depicts farm class patterns, where the relative frequency for the three first principal components was extracted from data corresponding to experimental group 2, by a nonlinear generalisation of PCA. As can be observed, histograms of the categories living in nonlinear combinations of the signal subspace spanned by these directions, seem



(a)

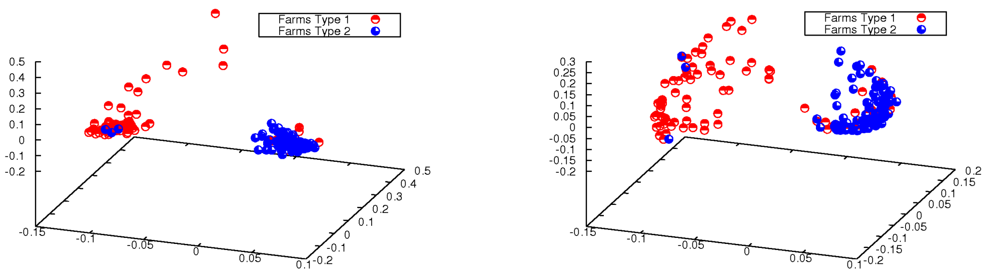


(b)



(c)

Fig. 4.4. Histograms of farm's class relative frequency for the 1st (a), 2nd (b), and 3rd (c) principal components for Aragua-Guarico dataset



(a) (b)

Fig. 4.5. Impact on class separation after linear (a) and gaussian (b) feature extraction preprocessing for Aragua-Guarico group dataset.

to show a high degree of overlapping. Particularly for instance, in the third component where class centroids appear with little differentiation; while in the first PC, the two groups of observations corresponding to classes 1 and 3 appear discriminated in different areas of the diagram, but with class 2 not separated. On the other hand, the separation of classes 2 and 3 is mainly in terms of the second PC; but in this case there was no differentiation for class 1.

It is interesting to note that as in experimental group 1, the subspace created by the combination of these three directions, the nonlinear approach resulted optimal in producing a good class separation as can be observed in Fig. 4.7. In this visual display of data clustering, it also can be appreciated that categories segmented under linear transformation (a) appear not very well separated. Though all three classes occupied different areas of the feature space, there was not clear group structure in the data that led to think of a linear decision surface with minimum misclassifications between farm's clusters.

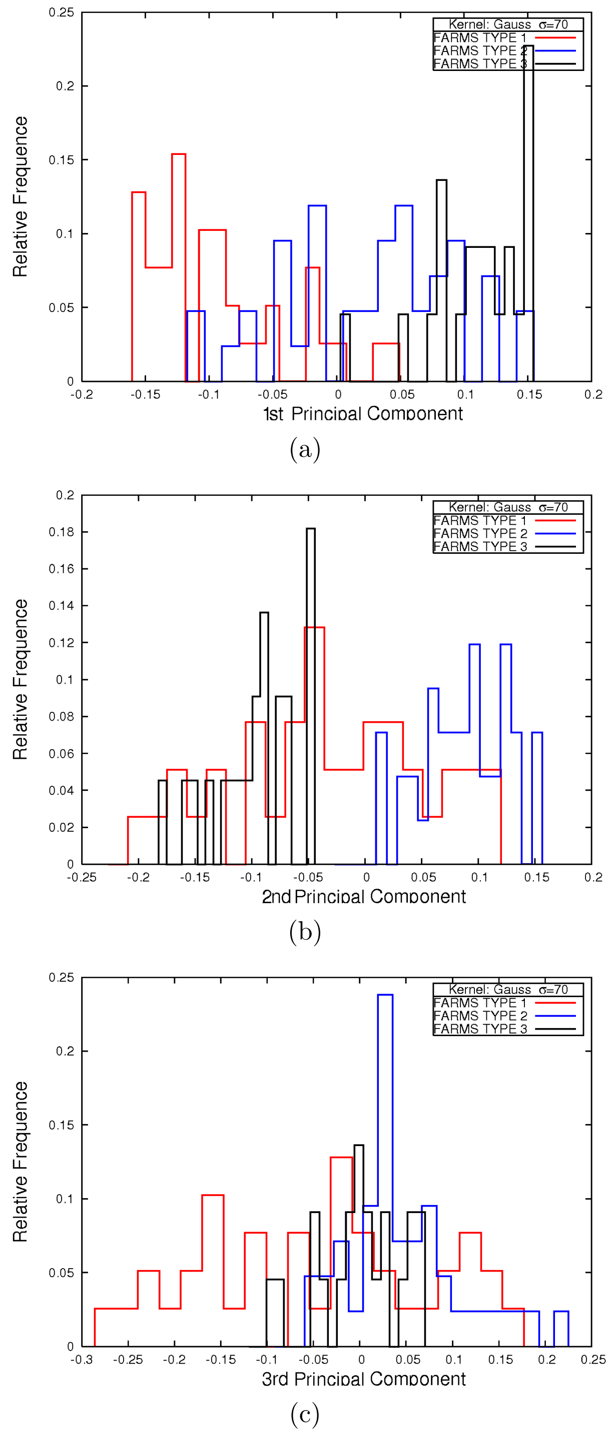
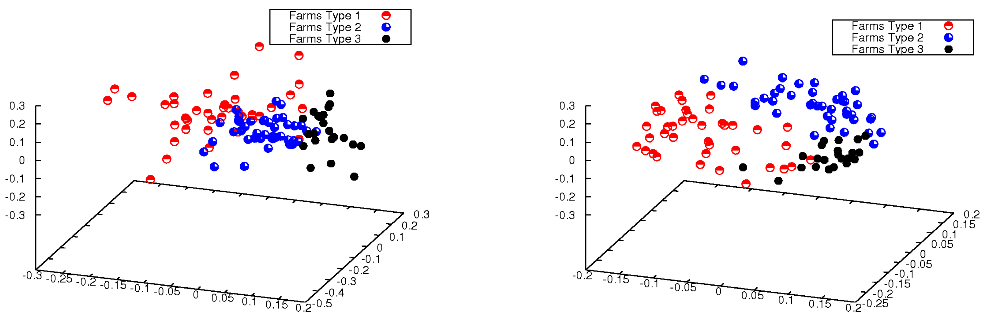


Fig. 4.6. Histograms of farm's class relative frequency for the 1st (a), 2nd (b), and 3rd (c) principal components for Guarico dataset

On the other hand, the nonlinear feature extraction (Fig. 4.7 b) based on a Gaussian kernel ($\sigma = 70$) showed that there was just one clear group structure with different classes representing subsets of a whole spectral group which agrees with what was encountered in experimental group 1 for the class 2. Overall, each informational class dissects the observations into relatively homogeneous areas of the space with little overlapping between class 1 and 3. The spectral likeliness between these two farm classes resembles what was described in chapter 3 with respect to the variations of land cover attributes by farm class; where farms type 3 seemed to share the same proportion of forest, pasture and forages as farms type 1 (Fig. 3.6).



(a) (b)

Fig. 4.7. Farm class scatterplot for the first three principal componets after Linear (a) and Gaussian (b) feature extraction for Guarico dataset.

Additionally, these results are consistent with the idea that diagnostic aspects of these farm classes appear to be concentrated in those principal directions of more variability, which, in other words, means that the spectral vector space of these farm groups contains the dynamic of interest spread across most of the spectral channels used in this study, in which case PCA was a convenient feature extraction technique (Landgrebe, 2007). Moreover, in spite of their lack of class separability in most individual principal directions in experimental groups 1 and 2, their observed complementarity has been reported previously (Guyon and Elisseeff, 2003; Jolliffe, 2002); in the sense that those variables, apparently useless in isolation, might become useful when used in a concomitant way with other variables, leading to good class separation in the feature space.

It follows that from a problem complexity perspective the 180-dimensional vectors can be effectively reduced to only three component values, using linear and Gaussian kernels respectively. The reduced new dimensions, characterized by their consistent high variability captured into these components, can now be used in substitution of the original 180 to train the linear machine; given that nonlinear dependences between objects have disappeared and the resulting data only contain essential information under much simpler restrictions.

4.6.2 Training set impact

The sampling process of multi-spectral data undertaken in this part of the study 4.2, permits a dense collection of spectral signatures which resulted in high dimensional input spaces. In the previous section, it was seen that this problem may be overcome with a minimum loss of information and high rates of discrimination between classes using an improved vector space representation. However, one of the problems that remains open is the lack of sufficient samples to exhaustively describe all the interclass variability present in such a rich source of data. In consequence, the next step is to determine the smallest possible number of training samples to be included in the classification process that guarantee good generalisation capabilities.

4.6.2.1 Aragua-Guarico set

To this end, the impact of training set cardinality on classifier performance for experimental group 1 is depicted in Fig. 4.8. As can be seen on the graph, KA inaccuracies decline as the training set increases in both linear and Gaussian approaches. The differences observed in the performance of both machines ranges from a relatively low mean value of 6 inaccuracies in training groups of cardinality higher than 50 instances, to a value of 18 inaccuracies for sets of 10 instances.

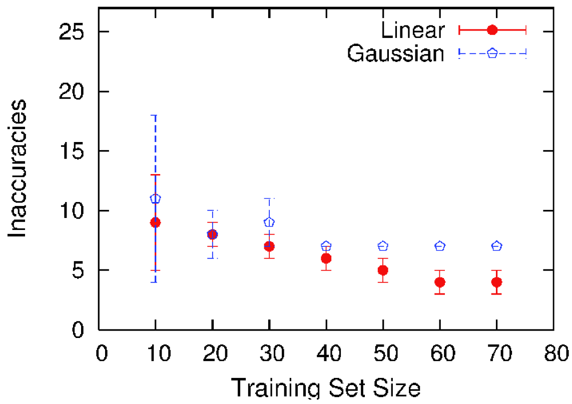


Fig. 4.8. Impact of training set size on generalisation performance of the linear-Gaussian (a) and Gaussian-polynomial classifiers (b)

Generally, the classifier showed a high sensitivity to the training set size; nevertheless, it is interesting to note the high variability in performance observed by the linear machine across all training set size domain. For example, low mean mistakes were committed by the machine even when small training sizes were used (10 instances). In this sense, this result seems to replicate what has been previously encountered by Foody and Mathur (2004), who found that independent of the training set size, if training is performed with instances that lie at the edge of the class distribution, a separating hyperplane may be found even when just few of these cases are available.

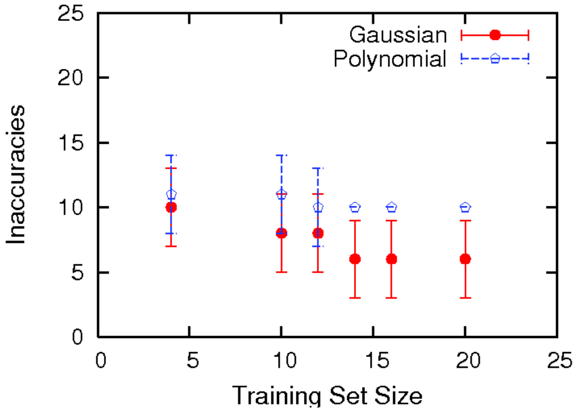
In consequence, the good performance observed for the linear approach in this experimental group would seem to stem, possibly on one hand, from the linear separability of the classes, which is expected given the differences in geographical location between both classes within this group. On the other hand, the presence in the training set of outlying instances might have yielded appropriate instances vectors to find a separating hyperplane with high generalization capacity without requiring an exhaustive training sample (Sanchez-Hernandez et al., 2007; Boser et al., 1992). Given that conversely to traditional approaches, which ignores the outliers, maxim margin classifiers exploit atypical patterns that are at the edge of the decision boundaries rather than those close to the mean centroid (Melgani and Bruzzone, 2004).

4.6.2.2 Guarico set

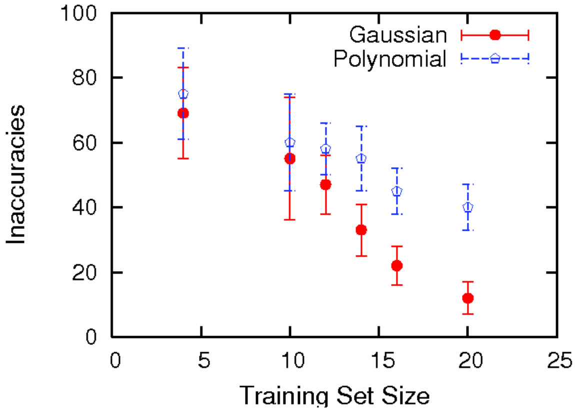
As in experimental group 1, the learning machine showed high sensitivity to training set size with data from experimental group 2. However, in this case the linear approach failed to find a separating hyperplane to discriminate the three classes into which input data in this group was mapped. As a result, Gaussian and polynomial kernels machines were used to fit the labelled spectral response in an accurate way. Fig. 4.9 depicts the performance of both learning algorithms under different training set sizes for the three informational farm classes. As can be appreciated, both machines used for class 1 converged at a relatively low size of training sample (4 instances, 10 inaccuracies). It can be appreciated that with the Gaussian kernel the machine showed a better performance than the polynomic approach; only 4 inacuracies ocurred with a training set size of 16 instances.

Like machine performance for class 1, the KA algorithm observed a high response to training set size for class 2 (Fig. 4.9 b). Nevertheless, finding a separating hyperplane for this class proved more difficult than for class 1; as rate of sampling was lower the inacuracies were several orders of magnitude higher than for class 1, and it was not until 20 instances were used that the minimum level of inaccuracies (8 instances) was obtained. As can be appreciated, for this class also the Gaussian kernel performed better than the polynomic, keeping a clear advantage of 9 % vs. 38 % inacuracies respectively.

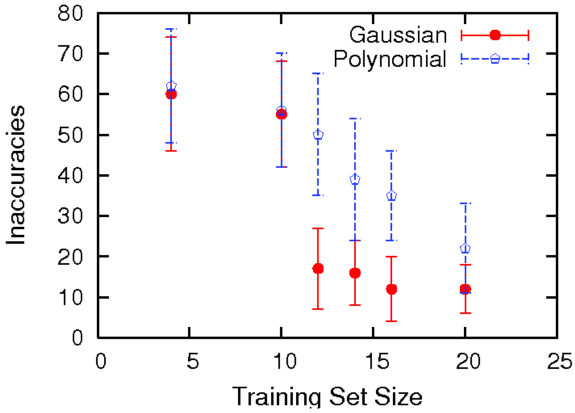
The last experiment addressed the evaluation of performance of the KA for the class 3 (Fig. 4.9 c). It can be observed that both kernel functions performed similarly until 10 instances were incorporated into the training process; and after including 12 instances Gaussian kernel showed the best performance, reaching its minimum level of error at a training set size of 16 where only 5 instances were allocated in the wrong class. It is interesting to note that the lowest error was not reached with the highest training set size; this is possibly due to the same fact as for the experimental group 1, where it was clear that a large training sample is not necessarily required (Sanchez-Hernandez et al., 2007; Melgani and Bruzzone, 2004); although a big data set may offer more possibilities of including key samples (Foody and Mathur, 2004).



(a)



(b)



(c)

Fig. 4.9. Impact of training set number on the classification acccuracy for the separation of farm class 1 (a), 2(b), and 3(c).

These experiments provided a minimum reference sampling size to choose the training machine for further analysis. In this sense, the training sample size for experimental group 1 was 20 instances; while for experimental group 2 the sampling rate was 14, 20 and 16 instances for class 1, 2 and 3 respectively. This level of sampling is quite usual within the field of learning machines (Schölkopf and Smola, 2002), and particularly for those practitioners working on the land use-land cover domain (Huang et al., 2002; Camps-Vals and Bruzzone, 2005; Bazi and Melgani, 2006; Pal and Mather, 2003). It has to be recalled that in this study for each instance involved there were 20 pixels per spectral channel used.

The present study is consistent with the idea of low sensitivity of maxim margin algorithms to the critical problem of scarce and noisy sample availability (Bruzzone and Melgani, 2005), given the particular behaviour of class distribution in hyperspaces reached by nonlinear mappings to find separating hypersurfaces focusing on those samples closer to the decision boundaries (Melgani and Bruzzone, 2004).

4.6.3 Classification

The main objective was to induct the relationship existent between one specific collection of land cover and a targeted farm class label, exploiting a classification strategy that takes advantage of a margin-based geometrical approach rather than a statistical criterion. This makes it possible to handle large input spaces, with high degree of efficiency and dealing robustly with samples characterised by noise, significant uncertainty degrees, and high cost of sample collection and labeling.

4.6.3.1 Aragua-Guarico set

As an intrinsically binary classifier, the KA algorithm was applied straigthforwardly on the feature space selected from the preprocessing stage for experimental group 1 given that only two classes were involved.

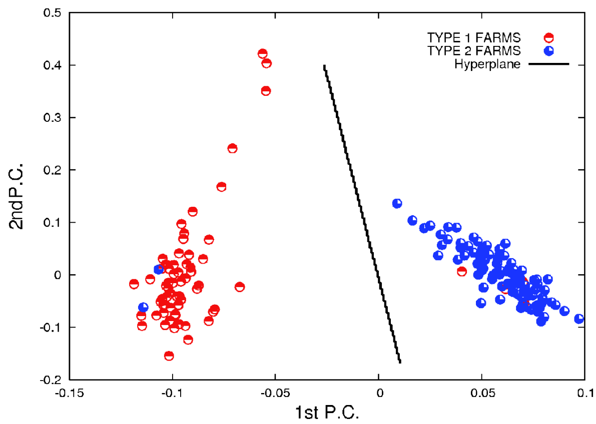
Fig. 4.10(a) and (b) show decision boundaries reached by the KA algorithm for the linear and Gaussian ($\sigma = 70$) kernel functions respectively for group 1. As can be seen, the use of no more than three principal components features allows the learning of very good classifiers. Finding these separating decision functions on the segmentation of farm classes is particularly significant given the nonstationary spatial behaviour of the spectral response of this kind of object; and because of the small training sets size with respect to the dimensionality of the input space.

Table 4.2. Confusion matrix for the accuracy on the segmentation of two farm classes trained on 20 cases using the kernel adatron algorithm.

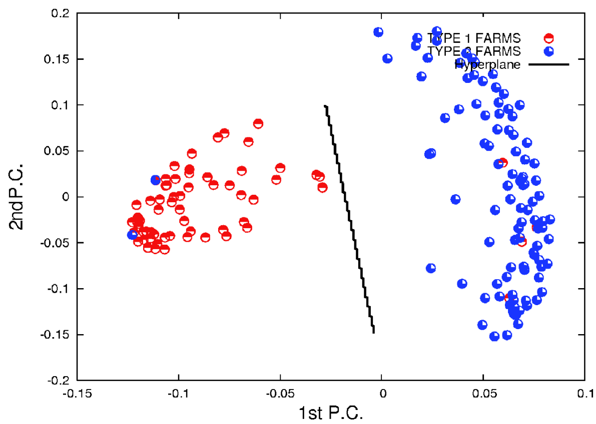
	KA	*Predicted*			
		Class 1	*Class* 2	Σ	Accuracy (%)
Actual	*Class* 1	63	2	65	96.92
	Class 2	4	99	103	96.11
	Σ	67	101	168	
	Accuracy (%)	94.02	98.01		*Overall Accuracy*(%) 96.42

KA: Kernel Adatron

Table 4.2 shows the accuracy levels observed in both cases, employing information only from Landsat images. As can be seen, a satisfactory generalization over the unseen



(a)



(b)

Fig. 4.10. Separating hyperplanes found by a linear (a) and Gaussian (b) KA machine from data of experimental group 1.

instances was observed, with just a few inaccuracies in both classes, suggesting that underlying differences between farm classes were effectively recognised. It is interesting to see that the linear approach showed a superior performance, indicating the intrinsically linear nature of the problem posted within this experimental group. This can be explained by the relative geographic separation between both informational classes.

Table 4.3. Confusion matrix for the accuracy on the segmentation of two farm categories trained on 65 and 103 cases for classes 1 and 2 respectively using the linear discriminant analysis algorithm.

	LDA	*Predicted*			
		Class 1	*Class* 2	Σ	Accuracy (%)
Actual	*Class* 1	62	3	65	95.3
	Class 2	24	79	103	76.6
	Σ	86	82	168	
	Accuracy (%)	72.09	96.34		*Overall Accuracy*(%) 89.88

LDA: Linear discriminant analysis

Results in this study have indicated that the kernel adatron machine may attain a comparable level of generalization as the linear discriminant analysis (LDA), when farms' multispectral response is used (Table 4.3). However, the main drawback of LDA is the requirement of exhaustive description of informational classes. For example, while for LDA it was necessary to use the whole dataset, the KA machine just required 10 instances per farm class, because this approach only focuses on extreme samples for its training making possible the derivation of comparable level of performance at a lower cost. This fact would confirm the argued advantages of previous applications of kernel methods in the land use domain, in which decision functions have been induced without any other *a priori* knowledge about the land cover than labels (Huang et al., 2002; Zhu and Blumberg, 2002). This implies a considerable resource saving in practical application to livestock systems monitoring.

4.6.3.2 Guarico set

The basic KA algorithm, is a binary classifier that makes use of an optimization procedure based on the descent gradient to find the maxim margin hyperplane that separates two groups. For the classification of farms from experimental group 2 a multi-class problem was faced given the existence of three informational categories. To deal with this problem a "one against the rest" strategy was adopted despite its well known suboptimal performance. Basically, three machines (one per each class) were trained organized in such an assemble that the class of interest is compared against the other two (Fig. 4.11).

Table 4.4 presents the performance accuracy of the three KA machines trained for this experimental group. As can be seen the KA appears to be more sensitive for class 1, given the highest accuracy reached, and apparently its degree of confusion seems to be with class 3. This may be explained by the levels of farming intensification observed in farm class 1, with an important degree of fragmentation of the land cover mosaic, which probably facilitated its differentiation from those instances that resemble more natural scenes as less intensive farms classes 2 and 3 (Drury, 2001).

The tendency to wrongly allocate farm type 3 as class 1, might be due to the fact that these group of farms share similar attributes on their proportions of pasture,

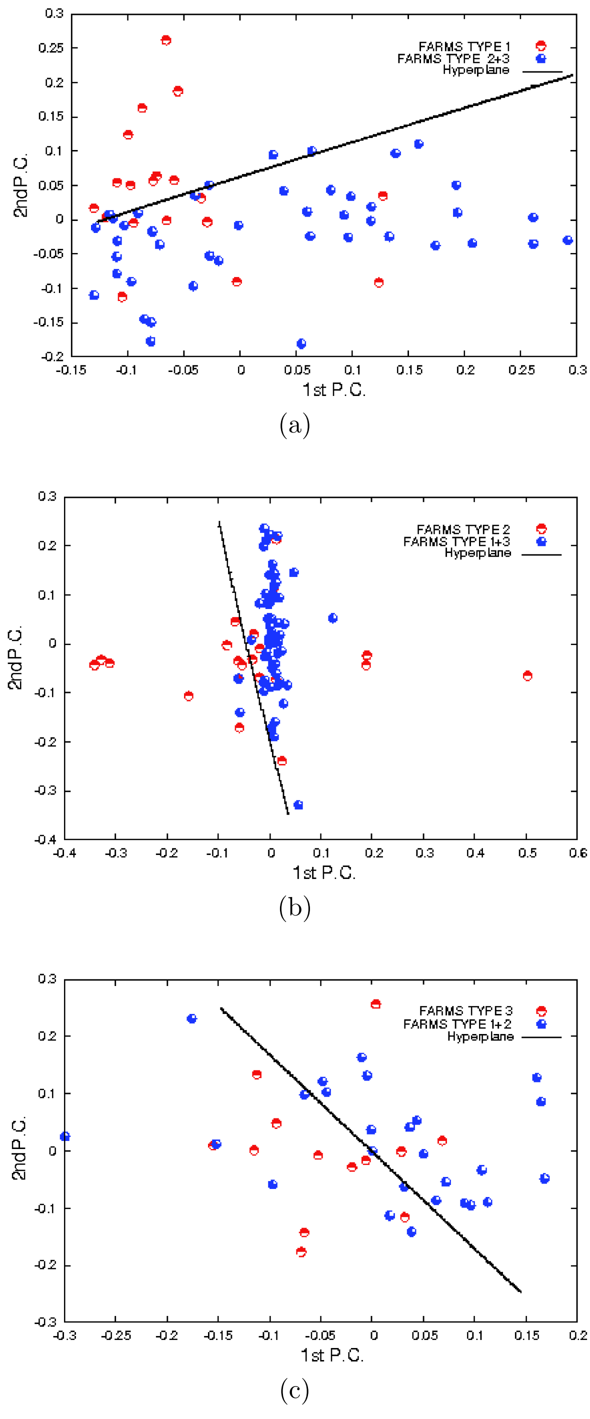


Fig. 4.11. Separating hyperplanes within group 2 for class 1 (a) and 3 (c) using a Gaussian kernel ($\sigma = 200$), and class 2 (b) using a polynomial kernel (order$= 3$; $\sigma = 4$).

Table 4.4. Confusion matrix for the segmentation of three farm categories trained on 14, 20, and 16 cases for class 1, 2, and 3 respectively using the KA machine.

	KA	*Predicted*				
		Class 1	*Class* 2	*Class* 3	Σ	Accuracy (%)
Actual	*Class* 1	35	0	4	39	89.8
	Class 2	2	34	6	42	80.95
	Class 3	3	2	17	22	77.27
	Σ	40	36	27	103	
	Accuracy (%)	87.5	94.44	62.96		*Overall Accuracy*(%) 83.49

KA: Kernel Adatron

forage and forest covertures, which probably are playing an important role in this lack of accuracy. Misclassification observed between classes 2 and 3, can be explained by its lack of anthropogenic modifications leading to occupy less discrete areas of the feature space as a function of the natural environment context (Richards and Jia, 2006; Landgrebe, 2007).

For comparison purposes, Table 4.5 shows the accuracy levels reached by linear discriminant analysis with the same dataset. It is worth noting the poor general performance exhibited by this approach making use of the whole data set. This level of performance might be responding to important properties of the vector space representation used in this study, which probably include a high proportion of nonlinear effects. On the other hand, in this research farms are seen as bags of pixels representing different land covers in a space where each dimension is associated to a spectral channel. Despite the fact that this vector space was sensibly transformed by linear feature extraction to improve representation, and with this to ensure equivalent land covers mapped to similar feature vectors, it was not possible to reach an acceptable level of accuracy with the linear discriminant analysis approach .

Table 4.5. Confusion matrix for the segmentation of three farm categories trained on 39, 42, and 22 cases for class 1, 2, and 3 respectively using the linear discriminant analysis.

	LDA	*Predicted*				
		Class 1	*Class* 2	*Class* 3	Σ	Accuracy (%)
Actual	*Class* 1	26	9	4	39	66.7
	Class 2	3	24	15	42	57.1
	Class 3	5	10	7	22	31.8
	Σ	34	43	26	103	
	Accuracy (%)	76.4	55.81	26.92		*Overall Accuracy*(%) 55.3

LDA: Linear discriminant analysis

Although it has been demonstrated in this research, that using these kernel methods it is possible to pass the critical level of accuracy (70 %) considered as minimum within the remote-sensing specialized literature (Thomlinson et al., 1999; Foody, 2002). The observed classification performance, when compared with preliminary studies in analogous agricultural and forestry applications (Camps-Vals and Bruzzone, 2005; Sanchez-Hernandez et al., 2007), can be criticized on the ground of its misclassification rate (17 %), in the sense that it may result too high for some stakeholders.

One possibility is that such misclassification levels appear essentially from the impact that spatial resolution has on the separability of informational classes (Landgrebe, 2007). Spatial resolution has been established to have a significant influence on spectral class separability, because of the hierarchy that generally characterize informational categories (Campbell, 2002; Rees, 2007); and there is reason to believe that similar effects occur with collections of land cover such as farms (Landgrebe, 2007).

In the present study, spatial resolution of Landsat 7 (ETM+) data, might have been too fine for the purposes of this research, in the sense that sometimes it is desirable to have pixel sizes smaller than the field under study, but not excessively small, because too fine resolution may lead to pixels that spectrally do not represent the field of interest but part of it. In farm classification, most of the time the interest is upon pixels that integrate across what is desired to be called a field, which in this study would be a farm, rather than a tiny part of a particular cover of crop, grassland or forest.

From that viewpoint, an alternative possibility is to use a source of data with a coarse spatial resolution, such as the Moderate Resolution Image Spectrometer (MODIS) (NASA, 2008). This sensor is part of the principal instruments aboard EOS[3] AM-1 (TERRA); and its spatial resolution ranges from 500m to 1 km, with a viewing swath width of 2.330 km. The possibility that the use of this sensor would lead to an improvement in farm classication accuracy, is in line with the reviews of Landgrebe (2007) and Drury (2001); in the sense that compared with Landsat 7, each pixel in MODIS would be made up of a mixture of "Landsat-size" pixels upon categories such as grass, crops, etc; that may lead to an improved representation of a farm, as a field of interest.

The advantages of MODIS would not be circumscribed to the spatial resolution; its spectral resolution, 36 channels covering from visible to thermal infrared spectral regions, also presents some benefit compared to the 7 bands of Landsat. This spectral

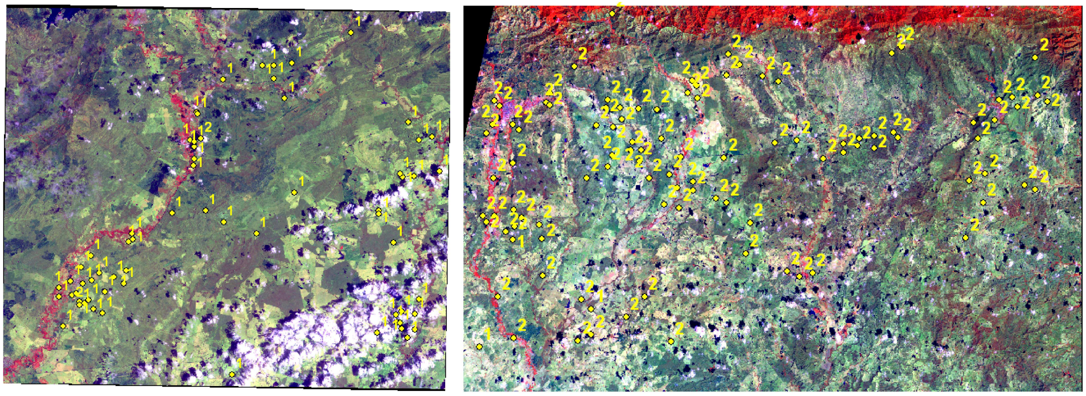
[3]Earth Observation System

richness would increases the accuracy on complex classes discrimination, since high volume space increase such likelihood. Evidence for the significance of spectral resolution on discrimination accuracy comes from Melgani and Bruzzone (2004); Foody and Mathur (2004); Bazi and Melgani (2006) and Muñoz-Marí et al. (2007). They exploit high spectral resolution sources, spreading the data out as much as possible in the feature space to make the most of the spectral richness, that generally results in small classification errors.

4.7 Unsupervised classification of forest cover within the study area

Given the manifest importance registered by the variable forest on the discrimination of farm categories in this study, and the increasing concern about forest ecosystems because of the changes led by deforestation with agricultural purposes; an unsupervised classification of forested land is carried out in order to map its spatial distribution and relate different cover categories with farm classes. To this end the two same Landsat 7 (ETM+) scenes that were used in previous analysis (Fig. 4.12), are now classified to provide a complete coverage of the study area.

Prior to proceeding with the classification, a subset of the areas of interest within the image was created (i.e the areas where farms were located) as can be observed in Fig. 4.12. Following this, the isodata algorithm from the software Erdas Imagine was used to first separate the data into 08 distinguishable cluster in the 7-dimensional pixel-value space; and then, these initial classes were then aggregated into two meaningful forest classes: forest cover (tropophilous and riparian forest), shrubs-trees, and an additional



(a) (b)

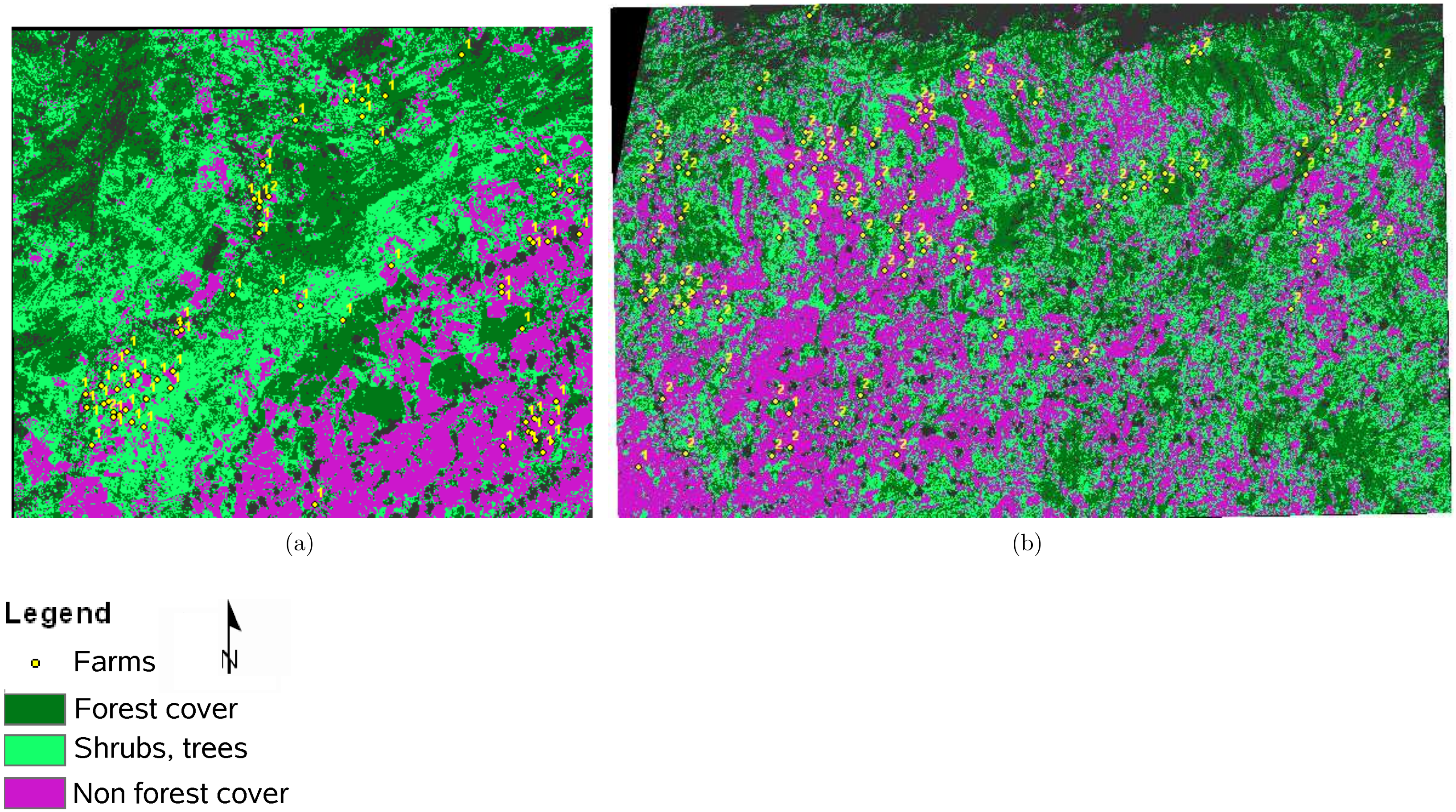
Fig. 4.12. Subsets of original landsat 7 (ETM+) scene upon Aragua (a) and Guarico (b) states, for unsupervised classification. Farms geographical location is indicated by yellow dots, labeled with the class number each holding belongs to.

category where were re-grouped the remaining clusters associated with non-forest covers (crop land, water, clouds, urban, cultivated and native grassland).

Fig. 4.13 shows the results for unsupervised classification for the Aragua (a) and Guarico (b) subsets. It can be appreciated in the images, the spectral clusters grouped in three categories as they were elicit from original images subsets, and ancillary data (Mogollon and Comerma, 1995). As can be seen, forest and shrubs-trees represent the largest part of both subsets (70% Aragua and 50% Guarico); followed by non forest covers, the habitat type most heavily used and intervened by agricultural activities.

In the Aragua subset, non-forest cover becomes more locally concentrated; while in Guarico a decreasing forest cover is observed due to conversion into annual crops and other agricultural like covers. The results of the classification agreed notably well with the ground based vegetation map displayed in Fig. 4.14 .

Fig. 4.13. Classification map of forest cover upon two landsat subsets (a) Aragua and (b) Guarico after unsupervised classification. Geographical position of farms and their typology is indicated in yellow.



Huber and Alarcon (1988) reflect in their map three main patways of cultivated or intervened land that can be identified in the classification. Aditionally, the behaviour ruptures of the covers make it possible to confirm that farm typologies are related to the spatial pattern depicted in the image clustering.

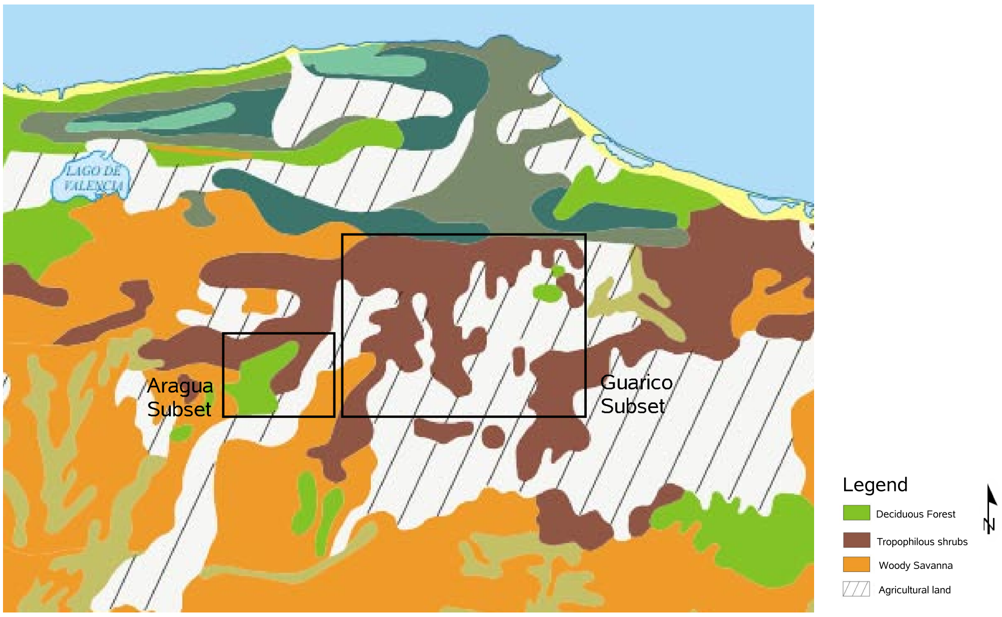


Fig. 4.14. Vegetation map of Venezuela (Huber and Alarcon, 1988). Black squares refer to the relative covering of Aragua (left) and Guarico (right) landsat subsets, used on the unsupervised classification.

Similarly, the results in this study are consistent with another classification undertaken by the USGS (2005)

Based on the feedback from chapters 2 and 3; and having the approximate forest cover for each farm, the relationship observed between forest ecosystems and farm types in this classification can be modeled. Given the binary nature of the farm category (class 1 or 2), a logistic regression approach was adopted rather than simple correlation or linear regression; and in the interest of of generating maximally useful data, apart from the variable forest cover, also the attribute altitude was included based on previous studies of Comerma and Chacon (2002).

Table 4.6. Coefficients of the variables in the equation

							95.0% *C.I.*	
	(β)	*SE*	*Wald*	*df*	*Sig*	*Exp*(β)	*Lower*	*Upper*
Altitude	.045	.007	41.333	1	.000	1.046	1.032	1.060
Forest	-.030	.009	12.523	1	.000	.970	.954	.987
Constant	-9.774	1.610	36.837	1	.000	.000	.000	.000

Table 4.6 shows the parameters that integrate the prediction equations. As can be observed the significance values of the Wald statistic for each predictor indicate that both forest cover and altitude predict farm classes very well. The value of $Exp(\beta) < 1$ for forest percentage indicate that the farm's chances of belonging to class 2 decrease when the value of forest cover goes up. For altitude the situation is different, the $Exp\,(\beta) > 1$ indicate that the likelihood of a farm belonging to class 2 become lower as value of altitude decreases. In other words, Aragua farms (class 1) tend to be located in forested areas at low altitude; while in Guarico farms (class 2) appear to occupy less forested areas, at a higher altitude.

Another parameter that confirms this, is the Hosmer and Lemeshow goodness of fit test, whose non-significant value $\chi^2 = 5.774$, $p = 0.673$ indicates that the model is predicting the real world data fairly well, since it tests the hypothesis that the observations are significantly different with respect to the predicted values from the model.

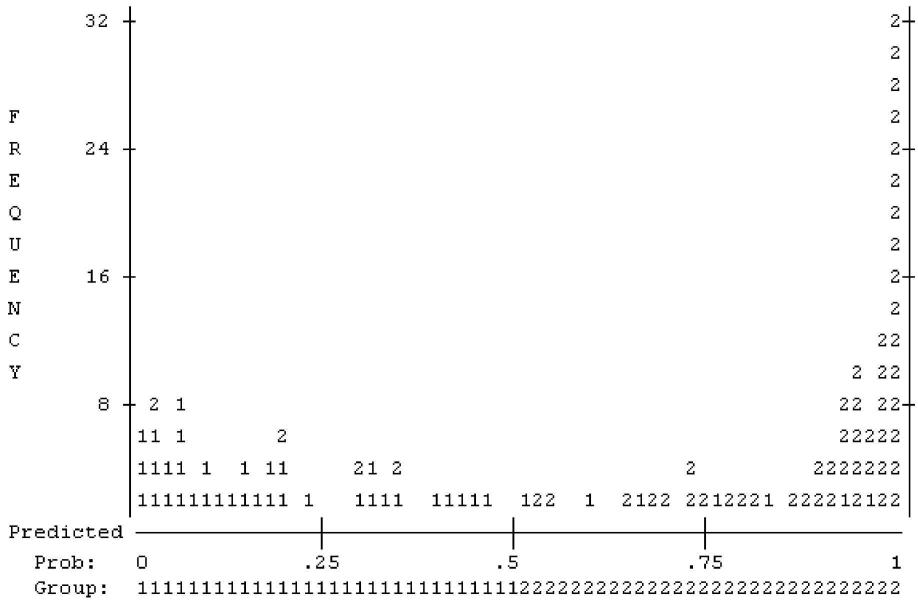


Fig. 4.15. Classification plot for observed group and predicted probabilities.

A visual representation of the model results is in Fig. 4.15. In this classification plot can be seen the predicted probabilities of a farm that having a given forest cover and at a particular altitude can be class 2. It is interesting to note that most of the cases are clustered towards the edges of the histogram, indicating that the model is performing accurately. These results are coherent with what is shown in the classified image (Fig.4.13); which is not suprising, since these two predictors were also used to generate the farm typologies; however, it is worth to noting their concomitant role in isolation from the other 8 attributes previously analyzed in chapter 2 and 3.

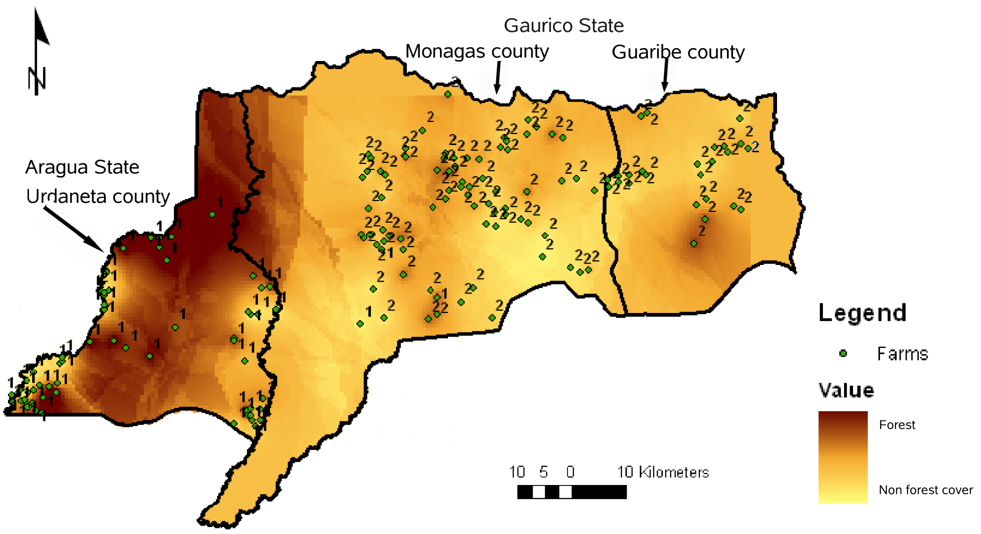


Fig. 4.16. Prediction surface of interpolated forest cover for different farm typologies.

In an effort to provide prediction surfaces of farm forest cover for the study area, an interpolation map by Radial Basis Functions of the within farm forest cover is presented in Fig. 4.16. As can be observed, most farms with high forest covers appear concentrated in Aragua subregion of the field, and there are focus of intensity loss towards the east part (Guarico section) of the study area. This spatially explicit prediction also resembles what is observed in the unsupervised classification, particularly for the farms located in the Aragua subset of the image.

4.8 Conclusion

A supervised farm classification from Landsat multi-spectral information, using the kernel adatron algorithm has been proposed in this research. The experimental results showed, that effective separation of particular groups of farms from others is practically achievable based on multi-spectral characteristics recorded in a satellite image; and revealed that repeatable links between biophysical and spectral features can be derived from abstractions difficult to observe as farms. The unknown equivalence between farm attributes and their spectral response summarized by labels has been used in a direct way to induct a strong hypothesis of representation with high generalisation capacity. The accuracy in classification performance, demonstrated that the spectral complexity of remote sensed images can be effectively handled without sacrificing the simplicity of linear approximations. Furthermore, the satisfactory performance of the Gaussian and polynomic kernel served to increase the explained variation between classes, and suggests that the information required to perform good classifications with this kind of data is relatively reduced. It should, however, be recognised that farm sample numbers were an important limitation and one which might dampen the potential application of this tool beyond the studied geographical area. Nevertheless, new observations may be added to the present sample and incorporated into the training set improving kernel machine generalization.

Chapter 5

General Discussion

This study has shown that the learning machine algorithm known as kernel adatron constitutes an efficient method for building a segmentation of crop-livestock system images. The main advantage is that the segmentation does not require *a priori* information about the theoretic probability models of farm inner land use, or production distribution, since it is guided only by the local information contained in a vector composed by a pixel neighborhood taken within the farm perimeters projected in a Landsat image.

Experimental results have also shown that the separation of particular groups of farms was not affected by the level of linear data separability. At first glance the procedure followed appears not to replicate Huang et al. (2002) and Pal and Mather (2003) methods, in the sense that in the present research the sampling of spectral information occurs within the limits (boundaries) of polygons, which represent farms (mosaics of land cover) rather than particular types of land cover. However, there were many similarities given that part of the success of this alternative is attributed to the enormous richness of representation shown by this methodology, as occurred with multi-spectral information work by Huang et al. (2002). As the previous works cited, the solutions in this research were not built in the input space, but in a higher dimensional space, as they were also based on the general theory of support vector machines.

The classification strategy known as "one against the rest" used in this research, leads to a high classification accuracy partially contrasting with the findings of Pal and Mather (2005). This implies that in order for such training to be effective, input data were taken to the feature space by means of two non-linear transformations - Gaussian and polynomic - whose diversity contributed to the richness of the solution expression despite the fact that some data resulted unclassified. Additionally, the form of the sampling may also have exerted some influence, given that in this research the sample included a vector of randomly selected pixels per image band, which might lead to a quite diverse training vector that gives the machine the opportunity to induct complex relationships regardless of the classification strategy (Huang et al., 2002; Su et al., 2007; Keuchel et al., 2003; Zhu and Blumberg, 2002).

The complexity issue of the training set involved in the supervised classification phase, was addressed through a nonlinear (kernel) mapping. This led to solutions, that make an indirect use of the kernel transformation implicit in the kernel functions, by simple inner products of the vectors composed from the farm pixels to be classified, represented as function in the input space, hence the transformation expression was not required. However, this phase of the research resulted in being highly affected by the precedent unsupervised classification, which mainly involved the feature extraction and clustering procedure on unlabeled samples (Duda et al., 2001).

One of the findings for this part of the research was that feature extraction through kernel methods proved more effective than the linear paradigm in providing principal directions which apart from accounting for most of the data variance, also provided the coordinate systems yielding the projection that best separated farm classes after discriminant analysis. Results seem to replicate what has been found in other fields of study (Schölkopf et al., 1999; Schölkopf and Smola, 2002); and the main idea behind it is that these results were reached by algorithms that sought the best solution (minimum/maximum global) to the given problem. This imposition reduced the degree of freedom, or in other words, the potential wrong conditioning in the learning problem raised.

This part of the study involved finding the direction of maximum variance in the cloud of data by solving an eigenvalue problem (Schölkopf et al., 1998). However, there is no guarantee that the best separation between clusters will be found in the hyperspace created by those principal directions of maximum variation (Jolliffe, 2002). From this point of view these particular results might be regarded as fortunate since some kernel function can deal with infinite dimensions (i.e Gaussian); nevertheless finding the direction of maximum separability also can be seen as the solution to a generalized eigenvalue problem that may be addressed with kernel procedures (Shawe-Taylor and Cristianini, 2006).

As an ancillary methodology for clustering, in this research the activity density of farm attributes by function that distributes their magnitude along a continuous surface was described. This approach resembles much of the element present in Kwan (2000)'s work; and includes several concepts discussed by Silverman (1986); Bailey and Gatrell (1995). One of those concept is radius search, whose cardinality accounts for the smoothing of the referred continuous surface. For instance, given a sample number, low cardinalities of this smoothing factor result in local adaptations of modes concentric around each sample reflecting small attribute details; while this effect is less probable for high cardinalities leading to smoother surfaces. There are several ways of determining the cardinality of this factor (Silverman, 1986; Sheather and Jones, 1991; Bailey and Gatrell, 1995; Levine, 2004). Most of these methodologies are based on dispersion data measurements and correction factor; however for this study Levine (2004) was selected since this approach takes into account the sample size and the area where activity density is considered.

Chapter 6

General Conclusion

This research has described farm typologies of crop-livestock systems where the creation of continuous density surfaces displaying farming feature intensity proved appropriate as an ancillary methodology for eventual unsupervised classification. The achieved visual setting of the data, facilitates the exploration of spatial elements within a multidimensional domain and permits the local examination of attributes interaction in a given geographical context. These elements enable the use of activity density as an accompanying methodology for clustering crop-livestock systems. Thus, this ancillary methodology provided a simple and spatially explicit means of making decisions about the number of clusters into which a farm population may be segmented. It was shown that farm subsets identified through 3D visualisation of attribute density were confirmed by clustering analysis. For the Aragua-Gurico group the intrinsic farm classes in the sampled data resembled the spatial attribute gradients that were observed on density surfaces; and the taxonomy of farms encountered corresponds to meaningful features at ground level. The low accuracy and complexity observed on those farm classes of difficult segmentation, as for Guarico, can be successfully addressed using nonlinear feature extraction procedures. With regard to the representation of the nonlinear feature extraction, it should be taken into account that in crop-livestock pattern recognition this is directly related to the transformation of input data into a reduced representation of farm features. In this aspect Kernel methods are more powerful, particularly for the very complex subset of crop-livestock data (Guarico group), given that they are applied equally to any data structure and show low sensitivity to the sample distribution. In the case of Gaussian and polynomic kernels, the sigma parameter, σ, and the order of the polynomial controls the width of the kernel funtion making it more local as these parameters increase. Although this can contribute to improving detailed representation, it may also increase the model complexity. Conversely, linear procedures cannot be used to describe decision boundary surfaces of very complex data, and a definitively better model can be produced under this approach when linearly-separable data is available. The main advantage of extracting features from kernel methods is the enhancement of clustering results in terms of their discriminative power. Although at first glance there do not seem to be differences for data linearly separated, the polynomic kernel leads to results characterized by clusters with a less sparse covariance matrix, hence much more compact groups. On the complex nonlinearly separable data, the Gaussian kernel leads to adjusted decision boundaries with a high discriminatory power.

On the other hand, the supervised classification part of this research successfully uses the learning machine tools to accomplish the task of classifying multi-spectral farm responses according to labels generated in the unsupervised section of the study. The machines modeled can induct features based on a representative sample of farms that are able to generalize beyond the instances shown. In other words, the algorithm applied achieved the task of distributing a number of vectors in the input space in such a way that this distribution can reflect, in one of several possible ways, the probability density of the signals, which has not been given in an explicit way but uniquely through example vectors. Thus, the relationship present in the original data is described and its representation is simplified by preserving the most relevant features. This study demonstrated that models produced in this way are flexible, expressive and compact. According to the performance of unsupervised and supervised approaches to classifying crop-livestock farms it has been found that kernel methods are effective in assisting this kind of learning task and are also very efficient in achieving a good representation of data, particularly for the complex ones. Now the direction in which future research can be focused might be finding limits for the minimum information required to train a linear machine in order to produce a similar performance; the other direction is regarding an issue that still remains open in this research; that is the choice of parameters to define the kernel complexity. One alternative for the first direction might be to test whether it is possible to gather multi-spectral information within a certain random area around farm centroids, in order to avoid the time-consuming collection and geo-referencing farm perimeters. For the second direction, one option is to carry out experiments with a rapid and systematic mechanism of parameter estimation already available in the literature, to permit objective adjustments to the complexity of the kernel space. Finally, another important direction, and one which must be analyzed from a broader perspective, is to increase the training set in order to gradually enhance the generalization capacity.

References

Aizerman, M., Braverman, E., and Rozonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automations and Remote Control*, 25:821–837.

Amatulli, G., Perez-Cabello, F., and de la Riva, J. (2007). Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. *Ecological Modelling*, 200:321–333.

Andersen, E., Elbersen, B., and Godeschalk, F. (2004). Farming and the environment in the European Union-using agricultural statistics to provide farm management indicators. Technical report, OECD Expert meeting on farm management indicators and the environment, Palmerson North, New Zealand.

Anlauf, J. and Biehl, M. (1989). The adatron: an adaptative perceptron algorithm. *Europhysics Letters*, 10:687–692.

Armstrong, M. (2000). Geography and computational science. *Annals of the Association of American Geographers*, 90:146–156.

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404.

Asrar, G. and Dozier, J. (1994). *EOS-science strategy for the Earth observing system.* AIP Press, Woodbury, NY.

Bailey, T. and Gatrell, A. (1995). *Interactive spatial data analysis.* Longman, New York.

Barsi, J., Schott, J., Palluconi, F., Helder, D., Hook, S., Markham, B., Chander, G., and O'Donnell, E. (2003). Landsat tm and etm+ thermalband calibration. *Canadian Journal of Remote Sensing*, 29(2):141–153.

Baumont, R., Ginane, C., Garcia, F., and Carrère, P. (2005). How herbivores optimise diet quality and intake in heterogeneous pastures, and the consequences for vegetation dynamics. In Milne, J., editor, *Pastoral systems in marginal environments.*, Proceedings of a satellite workshop of the XXth international grassland congress, pages 39–50, Gasgow, Scotland. Wageningen Academic.

Bazi, Y. and Melgani, F. (2006). Toward an optimal SVM classification system for hyperspectral remote sensing images. *IEEE Transaction on Geoscience and Remote Sensing*, 44(11):3374–3385.

Beinroth, F., Jones, J., Knapp, E., Papajorgji, P., and Luyten, J. (1998). Evaluation of land resources using crop models and gis. In Tsuji, G., Hoogenboom, G., and Thorton, P., editors, *Understanding options for agricultural production. Systems approaches for sustainable development*, Dordrecht, The Netherlands. Kluwer Academic Publisher.

Berdegue, J. and Escobar, G. (1990). Conceptos y metodologias para la tipificación de sistemas de finca. In *Tipificación de Sistemas de Producción Agrrícola*, pages 13–43, Santiago de Chile. RIMISP.

Berg, C., Christensen, J., and Ressel, P. (1984). *Harmonic analysis on semigroups.* Springer-Verlag, New York, USA.

Berroteran, J. (1997). Un modelo sostenible de utilización de forraje de sabanas bien drenadas con suelos ácidos en los llanos Venezolanos. In *XIII Cursillo sobre bovinos de carne*, pages 161–191, Maracay, Venezuela. Universidad Central de Venezuela. Facultad de Ciencias Veterinarias.

Berroteran, J. and Zinck, J. (2000). Indicadores de la sostenibilidad agricola nacional cerealera. caso de estudio: Venezuela. *Revista de la facultad de Agronomia de la Universidad del Zulia*, 17:139–155.

Betancourt, K., Ibrahim, M., Villanueva, C., and Vargas, B. (2005). Caracterización del manejo productivo de sistemas lecheros en la cuenca del río bulbul de mataguás, matagalpa, nicaragua. *Livestock Research for Rural Development*, 17(7):80.

Bishop, C. (2006). *Pattern recognition and machine learning.* Springer, Singapure.

Boser, B., Guyon, M., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152.

Bouwman, A., der Hoek, K. V., Eickhout, B., and Soenario, I. (2005). Exploring changes in world ruminant production sytems. *Agricultural Systems*, 84(2):121–153.

Bowman, A. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with s-plus illustration.* Oxford University Press, Oxford, England.

Bray, J. and Maxwell, S. (1985). Multivariate analysis of variance. Sage university paper series on quantitative applications in the social sciences. Sage, Thousand Oaks, CA, USA.

Brown, M., Gunn, S., and Lewis, G. (1999). Support vector machines for optimal classification and spectral unmixing. *Ecological Modelling*, 120:167–179.

Bruinsma, J. (2003). *World agriculture: towards 2015/2030. An FAO perspective.* Earthscan Publications, London.

Bruzzone, L. and Melgani, F. (2005). Robust multiple estimator systems for the analysis of biophisical parameters from remotely sensed data. *IEEE Transactions on Geosciences and Remote Sensing*, 43(1):159–174.

Burrough, P. and McDonnell, R. (2005). *Principles of Geographical Information Systems.* Oxford University Press, New York.

Campbell, J. (2002). *Introduction to remote sensing.* Taylor & Francis, London, UK.

Camps-Vals, G. and Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transaction on Geosciences and Remote Sensing*, 43(6):1351–1362.

Capillon, A. (1985). Connaitre la diversité des exploitations: un préalable á la recherche de réferences techniques regionales. *Agriscope*, 6:31–40.

Casley, D. and Kumar, K. (1990). *Recopilación y uso de los datos de seguimiento y evaluación.* Mundi-Prensa para el Banco Mundial, Madrid, España.

Chacon, E., Baldizan, A., and torres, R. (2004). Manejo sustentable de los recursos alimentarios de las sabanas Venezolanas. In *Primer taller: Perspectivas de las sabanas en el desarrollo agropecuario de Venezuela*, page 54 p. Universidad Simon Rodríguez.

Chicco, C., Plasse, D., and Bodisco, V. (1977). Reproducción del ganado bovino en Venezuela. *Agronomia Tropical*, 27(3):357–386.

CIRAD (1989). *Programa para el procesamiento y análisis estadistico de datos en microcomputadora. Análisis de encuestas (CSTAT). Manual del usuario.* Servicio Informatico, Montpellier, Francia. 151 pp.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1):155–159.

Combellas, J. (1998). Nivel de produccion de leche y necesidad de concentrado en sistemas sustentados en pastos tropicales. *Archivos Latinoamericanos de Producción Animal*, 6(1):45–54.

Comerma, J. and Chacon, E. (2002). Aptitud delos llanos para los principales usos ganaderos. In Romero, R., Arango, J., and Salomon, J., editors, *XVIII Cursillo sobre bovinos de carne*, pages 193–215, Maracay, Venezuela. Universidad Central de Venezuela. Facultad de Ciencias Veterinarias.

Coppock, D. (1994). *The Borana Plateau of Southern Ethiopia: Sysntesis of pastoral research, development and change, 1980-91.* International Livestock Centre for Africa, Addis Ababa, Ethiopia. 393pp.

Crammer, K. and Singer, Y. (2002). Pranking with ranking. In Dietterich, T., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems*, volume 14, pages 641–647. MIT Press.

Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods.* Cambridge University Press, Cambridge, UK.

Delgado, C., Rosengrant, M., Steinfeld, H., Ehui, S., and Courbois, V. (1999). Livestock to 2020. the next food revolution. In *Food, Agriculture and the Environment Discussion Paper 28*, Washington, DC. International Food Policy Research Institute, Food and Agriculture Organization of the United Nations, International Livestock Research Institute.

Diez, J., Coz, J., Bahamonde, A., nudo, C. S., Olleta, J., Macie, S., Campo, M., Panea, B., and Alberti, P. (2006). Identifying market segments in beef: breed, slaughter weight and ageing time implications. *Meat Science*, 74(4):667–675.

Dixon, J., Gulliver, A., and Gibbon, D. (2001). *Farming systems and poverty. Improving farmer's livelihoods in a changing world*, chapter 1st. FAO and World Bank. Malcolm Hall.

Dixon, W., Brown, M., Engelman, L., Fme, J., Hill, M., Jennrich, R., and Toporek, J. (1981). *BMDP statistical sofware:Berkeley.* University of California Press, California, USA.

Dobremez, L. and Bousset, J. (1995). *Rendre compte de la diversité des exploitations agricoles. Une démarche d'analyse par exploration conjointe de sources statistiques, comptables et technico-économiques.* Cemagref, France.

Domingos, P. and Pazzani, M. (1997). On the optimality of the simple bayesan classifier under zero-one loss. *Machine Learning*, 29:103–137.

Domínguez, C. (2006). *Caracterización de fincas, condición corporal y efectos de la suplementación sobre las variables reproductivas en vacas de doble propósito en el estado Guarico*. PhD thesis, Facultad de Agronomía, Universidad Central de Venezuela, Maracay, Venezuela.

Drury, S. (2001). *Image Interpretation in Geology*. Nelson Thornes Ltd, Cheltenham, UK, 3rd edition.

Dubayah, R. (1992). Estimating net solar radiation using landsat thematic mapper and digital elevation data. *Water Resourses Research*, 28:2469–2484.

Duda, R., Hart, P., and Stork, D. (2001). *Pattern classification*. John Wiley & Sons, INC, New York, USA, 2nd. edition.

Duvernoy, I. (2000). Use of land cover model to identify farm types in the Misiones agrarian frontier (Argentina). *Agricultural Systems*, (64):137–149.

Edelsbrunner, H. and Mücke, E. (1994). Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13(1):43–72.

Entz, M., Bellotti, W., Powell, J., Angadi, S., Chen, W., Ominski, K., and Boelt, B. (2005). Evolution of integrated crop-livestock production systems. In McGilloway, D., editor, *Grassland: a global resource*, pages 137–148, Wageningen, The Netherlands. Wageningen Academic.

Espinoza, F., Plaza, N., Díaz, Y., Palma, J., Alderovich, L., Aragort, W., Bracamonte, M., Carrillo, C., Guillén, A., León, E., Mireles, M., Molina, M., Morales, G., Obando, C., Perez, N., Pino, L., and Roa, N. (2005). Situación de la ganadería doble propósito en la altiplanicie de los Llanos Centrales. In *III Jornadas Nacionales de actualización de producción de leche*, Cabudare, Lara, Venezuela.

ESRI (2004). *ArcGis 9.1 User Manual*. Environmental Systems Research Institute, Redland, C.A.

Ewell, J. and Madriz, A. (1978). *Zonas de vida de Venezuela. Memorias explicativas sobre el mapa ecológico*. Ministerio de Agricultura y Cria, Dirección de Investigaciones, Caracas, Venezuela.

FAO (2004). Global livestock production and health atlas. In *Online. Retrieved February 20, 2007 from http://www.fao.org/ag/aga/glipha/index.jsp.*

FAO (2007). adaptation to climate change in agriculture, forestry and fisheries: Perspective, framework and priorities. Technical report, Interdepartamental working group on climate change, Rome, Italy.

Fiedler, F., Frasier, G., Ramirez, J., and Ahuja, L. (2002). Hydrologic response of grasslands: Effect of grazing, interactive infiltration, and scale. *Journal of Hydrologic Engineering*, 7(4):293–301.

Field, A. (2005). *Discovering statistics using SPSS.* SAGE Publications, London, UK, 2nd edition.

Fischer, G., van Velthuizen, H., Nachtergaele, F., and Medow, S. (2000). Global agro-ecological zones. Technical report, FAO-International Institute for Applied Systems Analysis, Rome, Italy.

Fisher, M. and Thomas, R. (2004). Implications of land use change to introduced pastures on carbon stocks in the central lowlands of tropical South America. *Environment, Development and Sustainability*, 6:111–131.

Fisher, R. (1935). *The design of experiment.* Oliver & Boyd.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188.

Fitzhugh, H. (1978). Bioeconomic analysis of ruminant production systems. *Journal of Animal Science*, 46(3):797–806.

Foody, G. (2002). Status of land cover classification accuracy assesment. *Remote Sensing of Environment*, 80:185–201.

Foody, M. and Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93:107–117.

Friest, T., Campbell, C., and Cristianini, N. (1998). *The kernel-adatron: A fast and simple learning procedure for support vector machines.* In: Proceedings of the Fifteenth International Conference on Machine Learning. Morgan-Kaufmann, San Francisco, USA.

Fujiska, S., Castilla, C., Escobar, G., Rodrigues, V., Veneklaas, E., Thomas, R., and Fisher, M. (1998). The effects of forest conversion on annual crops and pastures: Estimates of carbon emissions and plant species loss in a Brazilian Amazon colony. *agriculture, Ecosystems and Environment*, 69:17–26.

Fusagri (2001). Diagnóstico de la ganadería en los Municipios Camatagua y Urdaneta en el Sur del Estado Aragua. In *Papeles de Fundacite Aragua*. Fundacite Aragua, Maracay, Venezuela.

García, C. and Moreno, J. (2004a). *The hopfield associative memory network: Improving performance with the kernel "trick"*, pages 871–880. In: C. Lemaître; C.A. Reyes and J.A. Gonzalez (Eds). IBERAMIA 2004, LNAI 3315. Springer-Verlag, Berlin, Germany.

García, C. and Moreno, J. (2004b). *Kernel based method for segmentation and modeling of magnetic resonance images*, pages 636–645. In: C. Lemaître; C.A. Reyes and J.A. Gonzalez (Eds). IBERAMIA 2004, LNAI 3315. Springer-Verlag, Berlin, Germany.

García, C. and Moreno, J. (2004c). *The kernel hopfield memory network*, pages 755–764. In: P.M.A. Sloot; B. Chopard and A.G. Hoekstra (Eds). ACRI 2004, LNCS 3305. Springer-Verlag, Berlin, Germany.

Gatrell, A., Bailey, T., and Rowlingson, P. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transaction Institite of British Gepgraphers*, 21:256–274.

Gillen, R., Eckroat, J., and McCollum, F. (2000). Vegetation response to stocking rate in southern mixed-grass prairie. *Journal of Range Management*, 53:471–478.

Gillen, R. and Sims, P. (2002). Stocking rate and cow-calf production on sand sagebrush rangeland. *Journal of Range Management*, 55:542–550.

Gleason, A., Prince, S., Goetz, S., and Small, J. (2002). Effects of orbital drift on land surface temperature measured by avhrr thermal sensors. *Remote Sensing of Environment*, 79:147–165.

Goel, N. (1987). Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data. *Remote Sensing Review*, 3:1–212.

Gold, C. (1991). Problems with handling spatial data-the voronoy approach. *CISM Journal ACSGC*, 45(1):65–80.

González, A., Russell, G., Márquez, A., Moreno, J., García, C., Domínguez, C., Colmenares, O., and Machado, J. (2007). Supervised farm classification from remote sensing images based on the kernel adatron algorithm. In *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS07)*, Barcelona, Spain.

Graepel, T. and Obermayer, K. (1998). Fuzzy topographic kernek clustering. In Bauer, W., editor, *Proceeding of the 5th GI Workshop Fuzzy Neuro Systems '98*, pages 90–97.

Green, R. and Hay, S. (2002). The potential of pathfinder avhrr data for providing surrogate climatic variables across africa and europe for epidemiological applications. *Remote Sensing of Environment*, 79:166–175.

Guerra, O. (2004). Valoración del método de densidades focales (kernel) para la identificación de los patrones espaciales de crecimiento de la población de Españ. *International Review of Geographical Information Science and Technology*, (4):136–165.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182.

Hair, J., Anderson, R., Tatham, R., and Black, W. (1998). *Multivariate data analysis*. Prentice Hall, Upper Saddle River, New Jersey, 5th edition.

Hansen, J. and Jones, J. (1996). A system framework for characterizing farm sustainability. *Agricultural systems*, 51:181–202.

Hart, R. (1982). An ecological systems conceptual framework for agricultural research. In Shaner, W., Philipp, P., and Schmehl, W., editors, *Readings in Farming Systems Research and Development*, pages 44–58. Boulder, CO: Westview Press.

Hart, R. (1990). Componentes, subsistemas y propiedades del sistema finca como base para un método de clasificación. In *Tipificación de Sistemas de Producción Agrrícola*, pages 45–61, Santiago de Chile. RIMISP.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The elements of statistical learning. Data mining, inference and prediction*. Springer series in statistics. Springer, New York, USA.

Heckenlaible, D., Meyerink, A., Torbert, C., and Lacasse, J. (2007). Landsat 7 (L7) enhanced thematic mapper plus(ETM+ level zero-r distribution product (LORP) data format control book (DFCB). Technical report, Department of the Interior U.S. Geological Survey, Sioux Falls, South Dakota.

Hermes, L., Frieauff, J., Puzicha, J., and Buhmann, J. (1999). Support vector machines for land usage classification in landsat tm imagery. In *Geoscience and Remote Sensing Symposium*, volume 1 of *IGARSS 99*, pages 348–350. IEEE international.

Hernandez, T. (2007). Enhancing retail location decision support: the development and application of geovisualization. *Journal of Retailing and Consumer Services*, 14:249–258.

Herrero, M., , and Thorton, P. (2007a). The role of multi-scale modelling in assesing levelihood adaptation strategies to climate change. In Donatelli, M., Hatfield, J., and A.Rizzoli, editors, *Farming System Design 2007, International Symposium on Methodologies on Integrated Analysis on Farm Production Systems*, pages 193–194, Catania, Italy.

Herrero, M., Van de Steeg, J., Bermués, A., Karanja, S., Thorton, P., Baltenweck, I., and Staal, S. (2007b). Identifying patways of change in mixed systems by linking regional land use and household models. In Donatelli, M., Hatfield, J., and A.Rizzoli, editors, *Farming System Design 2007, International Symposium on Methodologies on Integrated Analysis on Farm Production Systems*, pages 27–28, Catania, Italy.

Hertz, J., Krogh, A., and Palmer, R. (1991). *Introduction to the theory of neural computation*. Addison-Wesley.

Ho, Y. (1989). Special issue on discrete event dynamic system. In *Proceedings of the IEEE*, volume 77.

Holdridge, L. (1967). *Life zone ecology*. Tropical Science Center, San José, Costa Rica.

Hosmer, D. and Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley, New York, 2nd edition.

Hotelling, H. (1933a). Analysis of a complex of statistical variables into principal components. *The Journal of Educational Psychology*, 24(6):417–441.

Hotelling, H. (1933b). Analysis of a complex of statistical variables into principal components. *The Journal of Educational Psychology*, 24(7):498–520.

Huang, C., Davis, L., and Townshend, J. (2002). An assesment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23:725–749.

Huber, M. and Scheneider, D. (1999). spatial data satandards in view of models of space and the functions operating on them. *Computers and Geosciences*, 25:25–38.

Huber, O. and Alarcon, C. (1988). Mapa de vegetación de venezuela. 1:200.000. In *Online. Retrieved April 25, 2008 from http://library.wur.nl/isric/index2.html?url=http://library.wur.nl/WebQuery/isric/21882*, Caracas. Instituto Venezolano de Investigaciones Cientificas, Ministerio del Ambiente y los Recursos Naturales Renovables.

Hunsaker, C., Nisbet, R., Lam, D., Browder, J., Baker, W., Turner, M., and Botkin, D. (1993). Spatial models of ecological systems and processes: the role of gis. In Goodchild, M., Parks, B., and Steyaert, L., editors, *Environmental Modeling with GIS*, New York, USA. Oxford University Press.

Jain, A., Murty, M., and Flynn, P. (1990). Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323.

Johnson, S. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254.

Jolliffe, I. (1972). Discarding variables in a principal component analysis, I: artificial data. *Applied Statistics*, 21:160–173.

Jolliffe, I. (2002). *Principal components analysis*. Springer, New York, USA.

Kaiser, H. (1960). The application of electronic computers to factor analysis. *Educational and Psicological Measurement*, 20:141–151.

Keuchel, J., Naumann, S., Heiler, M., and Siegmund, A. (2003). automatic land cover analysis for tenerife by supervised classification using remotely sensed data. *Remote Sensing of Environment*, 86(4):530–541.

Kimes, D., Gastellu-Etchegorry, J., and Esteve, P. (2002). Recovery of forest canopy characteristics through inversion of a complex 3d model. *Remote Sensing of Environment*, 79:320–328.

Köbrich, C., Rehman, T., and Khan, M. (2003). Typification of farming systems for constructing representative farm models: two illustrations of the application of multi-variate analyses in chile and pakistan. *Agricultural Systems*, 76(1):141–157.

Kosko, B. and Isaka, S. (1993). Fuzzy logic. *Scientific American*, 271:76–81.

Kostov, P. and McErlean, S. (2006). Using the mixtures-of-distributions technique for the classification of farms into representative farms. *Agricultural Systems*, 88:528–537.

Kostrowicki, J. (1977). Agricultural typology concept and method. *Agricultural System*, 2:33–45.

Kruska, R., Reid, R., Thornton, P., Henninger, N., and Kristjanson, P. (2003). Mapping livestock-oriented agricultral production system for the developing world. *Agricultural Systems*, 77:39–63.

Kwan, M. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C*, 8:185–203.

Lachenbruch, P. (1975). *Discriminant analysis.* Hafner Press, New York, USA.

Landers, J. (2007). *Tropical crop-livestock systems in conservation agriculture. The Brazilian experience*, volume 5 of *Integrated Crop Management*. FAO, Rome, Italy.

Landgrebe, D. (2007). Multispectral thematic mapping of land areas, some fundamentals. *IEEE Geosciences and Remote Sensing Society Newsletter*, (145):11–15.

Levine, N. (2002). Crimestat ii computer program a spatial statistics program for the analysis of crime incident locations (version2.0). Technical report, Ned Levine and Associates, Annandale, VA/The National Institute of Justice, Washington, DC.

Levine, N. (2004). *A spatial statistics program for the analysis of crime incident locations. CrimeStat version 3.0.* Ned Levine & Associates, Houston, TX.

Li, X. and Yeh, A. G.-O. (2004). Data mining of cellular automata's transition rules. *International Journal of Geographical Information Sciences*, 18:723–744.

Loker, W. (1994). Where is the beef: Incorporating cattle into sustainable agroforestry systems in the Amazon Basin. *Agroforestry Systems*, 25:227–241.

Longley, P., Goodchild, M., Maguire, D., and Rhind, D. (2003). *Geographic information systems and science.* John Wiley & Sons, Ltd, Chichester, West Sussex.

MacEachren, A. and Kraak, M. (2001). Reasearch challenges in geovisualization. *Cartography and Geographic Information Systems*, 28(1):3–12.

Maidment, D. (1993). *GIS and hydrological modelling.* In: M. Goodchild; B. Parks and L. Steyaert (Eds). Environmental Modeling with GIS. Oxford University Press, New York, New York, USA.

Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate analysis.* Academic Press, London, UK.

McCracken, S., Brondízio, E., Moran, E., Nelson, D., Siqueira, A., and Rodriguez-Pedraza, C. (1998). Land-use: Example from the agricultural frontier of the Brazilian Amazon. In *Anais IX SimpÓsio Brasileiro de Sensoriamento Remoto*, pages 45–52, Santos, Brazil. INPE.

McIntire, J., Bourzat, D., and Pingali, P. (1992). *Crop-livestock interaction in sub-saharan Africa.* The World Bank, Washintong, USA.

McVicar, T. and Jupp, D. (2002). Using covariates to spatially interpolate moisture availability in the murray-darling basing. a novel use of remotely sensed data. *Remote Sensing of Environment*, 79:199–212.

Melgani, F. and Bruzzone, L. (2004). Classification of hyperpectral remote sensing images with support vector machines. *IEEE Transactions on Geosciences and Remote Sensing*, 42(8):1778–1790.

Mercer, J. (1909). *Functions of positive and negative type and their connection with the theory of integral equations.* Philosophical Transactions of the Royal Socciety of London, London.

Milá, M., Bartolomé, J., Quintanilla, R., García-Cachán, M., Espejo, M., Herráiz, P., Sánchez-Recio, J., and Piedrafita, J. (2006). Structural characterisation and typology of beef cattle farms of spanish wooded rangelands (dehesas). *Livestock Science*, 99:197–209.

Miller, D., Kaminsky, E., and Rana, S. (1995). Neural network classification of remote-sensing data. *Computers and Geosciences*, 21:377–386.

Milne, J., editor (2005). *Pastoral systems in marginal environments*, Proceedings of a sattellite workshop of the XXth international grassland congress, Glagow, Scotland. Wageningen Academic Publisher.

Minsky, M. and Papert, S. (1969). *Perceptrons: an introduction to computational geometry.* MIT Press.

Mogollon, L. and Comerma, J. (1995). *Suelos de Venezuela*. Editorial EX Libris, C.A, Caracas, Venezuela.

Moreno, A. (1991). Modelización cartográfica de densidades mediante estimadores kernel. *Treballs de la Societat Catalana de Geografia*, 6(30):155–170.

Morrison, D. (2005). *Multivariate statictical methods*. Brooks & Cole, Belmont, CA, USA.

Mosquera, O. (2005). Caracterización de las formas de producción bovina de la Región Centroccidental 2000-2003. Venezuela. *Gaceta de Ciencias Veterinarias*, 10(2).

Muñoz-Marí, J., Bruzzone, L., and Camps-Valls, G. (2007). A support vector domain description approach to supervised classification of remote sensing images. *IEEE Transaction on Geosciences and Remote Sensing*, 45(8):2683–2692.

Murray, M. and Illius, A. (2007). Multispecies grazing in the serengeti. In Hodgson, J. and Illius, A., editors, *The ecology and management of grazing systems*, pages 247–272, Oxon, UK. CABI International.

Myneni, R., Asrar, G., and Hall, F. (1992). A three-dimensional radiative transfer method for optical remote sensing of vegetated land surface. *Remote Sensing of Environment*, 41:105–121.

NASA (2008). Moderate Resolution Image Spectrometer (modis). In *Online. Retrieved April 25, 2008 from http://modis.gsfc.nasa.gov/*.

Nhemachena, C. and Hassan, R. (2007). Micro-level analysis of farmer's adaptation to climate change in Southern Africa. Technical report, International Food Policy Research Institute. IFPRI Discussion Paper 00714, Washington, DC, USA.

Nicholson, C., Blake, R., and Lee, D. (1999). Livestock, deforestation, and policy making: intensification of cattle production systems in Central America revisited. *Journal of Dairy Science*, 78(3):719–734.

Ortega, L., Ward, R., and Andrew, C. (2004). Measuring technical efficiency in venezuela: the dual-purpose cattle system. Technical Report EDIS-FE495, Department of Food and Resource Econimics, University of Florida, Gainesville, Fl. USA.

Otte, M. and Chilonda, P. (2002). *Cattle and small ruminant production systems in Sub-Saharan Africa-A systematic review*. FAO, Rome, Italy.

Páez, L. and Jiménez, M. (2000). Caracterización estructural y tipologias de fincas de ganadería de doble propósito en la microregion acequia-socopo del estado barinas, venezuela. *Zootecnia Tropical*, 18(2):177–196.

Páez, L., Linares, T., Sayago, W., and Pacheco, R. (2003). Caracterización estructural y funcional de fincas ganaderas de doble propósito en el municipio páez del estado apure, venezuela. *Zootecnia Tropical*, 21(3):301–323.

Pal, M. and Mather, M. (2003). Support vector classifiers for land cover classification. *Map India conference 2003 Image Processing & Interpretation*.

Pal, M. and Mather, M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5):1007–1011.

Payne, W. and Hodges, J. (1997). *Tropical cattle origins, breeds and breeding policies*. Blackwell Science Ltd, Cambridge, UK.

Pearson, K. (1901). On lines and planes of closets fit to points in space. *Phylosophical Magazine*, 2:559–572.

Pedhazur, E. and Schmelkin, L. (1991). *Measurement, design and analysis*. Erlbaum, Hillsdale, NJ, USA.

Peters, D. and Havstad, K. (2006). Nonlinear dynamics in arid and semi-arid systems: Interactions among drivers and processes across scales. *Journal of Arid Environments*, 65(2):196–206.

Powell, J., Fernadez-Rivera, S., Williams, T., and Renard, C. (1993). Livestock and sustainable nutrient cycling in mixed farming systems of sub-saharan africa. In *Proceedings of an international conference*, Addis Ababa, Ethiopia. International Livestock Centre for Africa.

Powell, J., Pearson, R., and Hiernaux, P. (2004). Crop-livestock interactions in the West African Drylands. *Agronomy Journal*, 96:469–483.

Preston, T. and Leng, R. (1987). *Matching ruminant production systems with available resources in the tropics and sub-tropics*. Penambul Books, Armidale, Australia.

Quiroz, R., Pezo, D., Rearte, D., and Martín, F. S. (1997). Dynamics of feed resources in mixed farming systems of Latin America. In Renard, C., editor, *Crop residues in sustainable mixed crop/livestock farming systems*, pages 149–180, Oxon, UK. CAB International, ICRISAT, ILRI.

Rees, W. (2007). *Physical principles of remote sensing*. Cambridge University Press, Cambridge, UK.

Renard, C., editor (1997). *Crop residues in sustainable mixed crop/livestock farming systems*. CAB International, ICRISAT, ILRI, Oxon, UK.

Richards, J. and Jia, X. (2006). *Remote sensing digital image analysis*. Springer-Verlag, Berlin, Germany.

Rodríguez, J. and González, D. (2001). Estudio ambiental de la cuenca del río Unare, las lagunas de Unare y Píritu y sus áreas de influencia. In *Cuadernos CENAMB*, number 5. Vicerrectorado Académico, UCV, Caracas, Venezuela.

Rodríguez-Iturbe, I., Poporato, A., Ridolfi, L., Isham, V., and Cox, D. (1999). Probabilistic modeling of water balance at a point: the role of climate soil and vegetation. *Proceeding of Royal Society of London*, 455:3789–3805.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.

Ruthenberg, H. (1980). *Farming system in the tropics. (third edition ed)*. Clarendon Press, Oxford, UK.

Sahitoh, S. (1988). *Theory of reproducing kernels and its applications*. Logman Scientific & Technical, Harlow, UK.

Sanchez-Hernandez, C., Boyd, D., and Foody, G. (2007). Mapping specific habitats from remotely sensed imagery: support vector machine and support vector data description based classification of coastal saltmarsh habitats. *Ecological Informatics*, doi:10.1016/j.ecoinf.2007.04.003.

Scarnecchia, D. (1985). The relationship of stocking intensity and stocking pressure to other stocking variables. *Journal of Range Management*, 38(6).

Schaback, R. and Wendland, H. (2006). *Kernel techniques: from machine learning to meshless methods*. Acta numerica (2006). Cambridge University Press.

Schölkopf, B., Mika, S., Burges, C., Knirsch, P., Müller, K., Rätsch, G., and Smola, A. (1999). Input space vs. feature space in kernel-based methods. *IEEE Transactions Neural Networks*, 10:1000–1017.

Schölkopf, B. and Smola, A. (2002). *Learning with kernels. Support vector machines, regularization, optimization, and beyond*. The MIT Press, Cambridge, Massachusetts, USA.

Schölkopf, B., Smola, A., and Müller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation.*, 10:1299–1319.

Seré, C. and Steinfeld, H. (1996). *World livestock production systems: Current status, issues and trends*. FAO Animal Production And Health Paper, Rome, Italy.

Shawe-Taylor, J. and Cristianini, N. (2006). *Kernel methods for pattern analysis*. Cambridge Univ. Press, Cambridge, UK.

Sheather, S. and Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B(Methodological)*, 53(3):683–690.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. chapman & Hall, London.

Singh, V. and Woolhiser, D. (2002). Mathematical modeling of watershed hydrology. *Journal of hidrologic Engineering*, 7:270–292.

Song, C., Woodcock, C., Seto, K., Lenney, M., and Macomber, S. (2001). Classification and change detection using landsat tm data: When and how to correct atmospheric effects? *Remote Sensing of Environment*, 75:230–244.

SPSS-Inc (1999). *SPSS base 7.5 syntax reference guide*. Chicago.

Steinfeld, H., de Haan, C., and Blackburn, H. (1997). Livestock-environment interactions: issues and options. Technical report, FAO-World Bank-European Commission, Suffol, UK.

Steinfeld, H., Wassenaar, T., and Jutzi, S. (2005). Livestock production systems in developing countries: status, drivers, trends. *Rev. sci. tech. Off. int. Epiz.*, 25(2):505–516.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates, Mahwah, NJ, USA.

Strahler, A. (1980). The use of prior probabilities in maximun likelihood classification of remotely sensed data. *Remote Sensing of Environment*, 10:135–163.

Su, L., Chopping, M., Rango, A., Martonchik, J., and Peters, D. (2007). Support vector machines for recognition of semi-arid vegetation types using misr multi-angle imagery. *Remote Sensing of Environment*, 107(1-2):299–311.

Tabachnick, B. (2001). *Using multivariate statistics*. Allyn & Bacon, Boston, USA, 4th edition.

Thomlinson, J., Bolstad, P., and Cohen, W. (1999). Coordinating methodologies for scaling landcover classification from site-specific to global: steps toward validating global map products. *Remote Sensing of Environment*, 70:16–28.

Thorne, P. (1998). Crop-livestock interactions. a review of oportunities for developing integrated models (consultant's report, systems analysis and impact assesment project). Technical report, ILRI, Nairobi, Kenia.

Thornton, P. and Herrero, M. (2001). Integrated crop-livestock simulation models for scenario analysis and impact assesment. *Agricultural Systems*, 70:581–602.

Tobler, W. (1979). Cellular geography. In Gale, S. and Olssen, G., editors, *Philosophy in Geography*, Dordrecht, Reidel.

Urdaneta, F., Materán, M., Peña, M., and Casanova, A. (2004). Tipificación tecnologica del sistema de producción con ganadería bovina de doble propósito. *Revista Cientifica de la Universidad del Zulia*, 14(3):254–262.

Urdaneta, F., Reichel, H., Suarez, G., Peña, M., and Casanova, A. (1999). Eficiencia productiva de arreglos tecnológicos en sistemas de producción de doble propósito en los municipios rosario y machiques de perijá, estado zulia, venezuela. *Revista de la Facultad de Agronomía de la Universidad del Zulia*, 16(1):252–258.

USGS (2005). Global land cover characterization program. In *Online. Retrieved May 26, 2008 from http://edcsns17.cr.usgs.gov/glcc/gifs.html*.

Vaccaro, L. (1998). El potencial de distintos genotipos para uso en diferentes niveles de manejo: evidencia de Venezuela. *Archivos Latinoamericanos de Producción Animal*, 6(1):83–90.

Van de Ven, G., de Ridder, N., Keulen, H. V., and Ittersum, M. V. (2003). Concepts in production ecology for analysis and design of animal ad plant-animal production systems. *Agricultural Systems*, 76:507–525.

Vapnik, V. (1995). *The nature of statistical learning theory.* Springer-Verlag, New York, USA.

Vapnik, V. (1998). *Statistical learning theory.* John Wiley & Sons.

Vapnik, V. and Chervonenkis, A. (1974). *Theory of pattern recognition.* Nauka, Moscow.

Vera, R. (1998). Genotipo, nutrición y reproducción en ganado de doble propósito: una revision selectiva de literatura. *Archivos Latinoamericanos de Producción Animal*, 6(1):55–70.

Wade, T., Schultz, B., Wickham, J., and Bradford, D. (1998). Modeling the potential spatial distribution of beef cattle grazing using a geographic information system. *Journal of Arid Environments*, 38:325–334.

Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Warren, S., Blackburn, W., and Taylor, C. (1986). Soil hydrologic response to number of pastures and stocking density under intensive rotation grazing. *Journal of Range Management*, 39(6).

Wegenholt, K. (2004). January cattle inventory and what and it suggests. Technical report, LSU AgCenter Extension Economist, USA.

White, D., Lubulwa, G., Menz, K., Zuo, H., Wint, W., and Slingenbergh, J. (2001). Agro-climatic classification systems for estimating the global distribution of livestock numbers and commodities. *Environmental International*, 27:181–187.

White, R. and Engelen, G. (1993). Cellular automata and fractal urban form: a cellular modeling approach to the evolution of urban land use pattern. *Environment and Planning*, A:1175–1199.

Wijk, M. V. and Rodríguez-Iturbe, I. (2002). Tree-grass competition in space and time: Insights from a simple cellular automata model based on ecohydrological dynamics. *Water Resources Research*, 38:1179.

Williams, C. (1998). Prediction with Gaussian processes: From linear regression to linear prediction and beyond. In Jordan, M., editor, *Learning and inference in graphical models*, Kluwer.

Wilson, A. and Macleod, N. (1991). Overgrazing: present or absent? *Journal of Range Management*, 44(5):475–482.

Wint, W., Slingenbergh, J., and Rogers, D. (2000). Livestock distribution, production and diseases. In *Towards a global livetock atlas*, Rome. Food and Agriculture Organisation of the United Nations. www.fao.org/ag/againfo/resources/en/glipha/default.html.

Wright, S. (1996). Phenological responses to seasonality in tropical forest plants. In Mulkey, S., Chazdon, R., and Smith, A., editors, *Tropical forest plant ecophysiology*, pages 440–460, New York, USA. Chapman & Hall.

Wylie, B., Meyer, D., Tieszen, L., and Mannel, S. (2002). satellite mapping of surface biophysical parameters at the biome scale over the north american grasslands. *Remote Sensing of Environment*, 79:266–278.

Xu, R. and Wunsch, D. (2005). Survey of clustering algorithm. *IEEE Transactions on Neural Networks*, 16(3):645–78.

Yang, F., Ichii, K., White, M., Hashimoto, H., Michaelis, A., Votava, P., Zhu, A., Huete, A., Running, S., and Nemani, R. (2007). Developing a continental-scale measure of gross primary production by combining modis and americaflux data through support vector machine approach. *Remote Sensing of Environment*, 110(1):109–122.

Zhu, G. and Blumberg, D. (2002). Classification using ASTER data and SVM algorithms; the case study of Beer Sheva, Israel. *Remote Sensing of Environment*, 80:233–240.