The University of Edinburgh

School of GeoSciences

Use of Multispectral Data to Identify Farm Intensification Levels by Applying Emergent Computing Techniques

by

Astrid Márquez

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

November 2011

Contents

List of Figures

List of Tables

Acknowledgements

I wish to express my gratitude to the Venezuelan Ministry of Science and Technology which gave me the financial support, the National Land Institute (INTI) for their contribution with the spatial farm data and the Geographical Institute Simn Bolvar (IGSB) that provided the satellite images; without their support the development of this work would have been impossible.

I would like to thank my supervisors, Dr. Graham Russell and Dr. José Alí Moreno for their continuous encouragement, careful guidance, and friendship throughout the development of this work.

This acknowledgement would not be complete without thanking my friends in Edinburgh: Mathew Palmer, Helen Bain, Alicia Salazar, Carlos Acosta, Roberta Berguero, Kostas Ververedis and the Campanelli family for their unconditional friendship, as well as my family and friends in Venezuela whose constant support gave me the strength to succeed in difficult times.

Abstract

Concern about feeding an ever increasing population has long been one of humankind's most pressing problems. This has been addressed throughout history by introducing into farming systems changes allowing them to produce more per unit of land area. However, these changes have also been linked to negative effects on the socio economic and environmental sphere, that have created the need for an integral understanding of this phenomenon. This thesis describes the application of learning machine methods to induct a relationship between the spectral response of farms' land cover and their intensification levels from a sample of farming of Urdaneta municipality, Aragua state of Venezuela. Data collection like this is a necessary first steep to implement cost-effective methods that can help policy-makers to conduct succesful planing tasks, especially in countries such as Venezuela where, in spite of there being areas capable of agricultural production, nearly 50% of the internal food requirements of recent years have been satisfied by importations. In this work, farm intensification levels are investigated through a sample of farms of Urdaneta Municipality, Aragua state of Venezuela. This area is characterised by a wide diversity of farming systems ranging from crop to crop-livestock systems and an increasing population density in regions capable of livestock and arable farming, making it a representative case of the main tropical rural zones. The methodology applied can be divided into two main phases. First an unsupervised classification was performed by applying principal component analysis and agglomerative cluster methods to a set of land use and land management indicators, with the aim to segregate farms into homogeneous groups from the intensification point of view. This procedure resulted in three clusters which were named extensive, semi-intensive and intensive. The land use indicators included the percentage area within each farm devoted to annual crops, orchard and pasture, while the land management indicators were percentage of cultivated land under irrigation, stocking rate, machinery and equipment index and permanent and temporary staff ratio, all of them built from data held on the 1996-1997 venezuelan agricultural census. The previous clusters reached were compared to the ones obtained by applying the learning machine method known as self-organizing map, which is also an unsupervised classification technique, as a way to confirm the groups' existence. In the second stage, the learning machine known as kernel adatron algorithm was implemented seeking to identify the intensification level of Urdaneta farms from a landsat image, which consisted of two sequential steps: namely training and validation. In the training step, a predetermined number of instances randomly selected from the data set were analysed looking for a pattern to establish a relationship between the label and the spectral response in an iterative process which was concluded when the machine found a linear function capable of separating the two classes with a maximum margin. The supervised classification finishes with the validation in which the kernel adatron classifies the unseen samples by using a generalisation of the relationships learned while training. Results suggest that farm intensification levels can be effectively derived from multi-spectral data by adopting a machine learning approach like the one described.

Chapter 1

Introduction

Producing enough food to feed an ever increasing population has long been one of man's most pressing problems. Scientific concern with the capacity of agricultural systems to support population growth was initially expressed by Malthus (1783), since then, many technological advances have occurred, allowing unthinkable agricultural output increasing in Malthus' days. Particular reference deserves the development of new varieties of wheat and rice led by Borlaug; that allow substantial increases in hectare's yields, which made it possible to bridge the gap between population growth and food production, a contribution that earned him the Nobel Peace in 1970. That breakthrough in agricultural technologies, gave raise in 1968 to the term Green Revolution, accelerating agricultural intensification, which even though has been widely recognised as an effective means to keep pace with global population growth and the increased food requirements, their effects on the socio economic and environmental sphere have been the source of much controversy.

In the socio economic arenas, the debate revolves around the ability of the green revolution to provide and enhance peasants' livelihoods. At one end, detractors arguing that because such technology implies the adoption of high-yielding crop varieties, increased use of fertilisers and pesticides, and investments in irrigation and agricultural machinery, the most favored were those farmers already well endowed, given that not all the inputs required were scale neutral, so that small producers who theoretically were his target were often marginalized (Carswell, 1997). At the other end, supporters argue that yield increases achieved through the green revolution helped to escape from poverty to an important number of farmers in the developing world (Paarlberg, 2009). Prices decline that made possible improvements in human health and life expectancy due the raise in calories intake that benefited consumers around the world have been cited as another success of the green revolution (Evenson and Gollin, 2003; Jewitt and Baker, 2007). Nevertheless, that prices decrease have been also reported as adversely affecting those farmers whose cost reduction through productivity increases were not enough to exceed price decreases (Evenson and Gollin, 2003).

In the environmental field the literature review suggests that the disparity of views is closely related to the matter studied. There are topics on which most researchers agree pointing out the negative effects of the green revolution, such as on biodiversity (Matson et al., 1997; Chamberlain et al., 2000; Smith et al., 2005). Pollution as a result of nutrient imbalances seems to be another area in which consensus prevails (EEA, 2009); indeed fertilization has been cited as responsible of most water eutrophication process (Conley et al., 2009), and also it has be listed as the main source of greenhouse gases emissions, specifically the intensive use of nitrogen fertilizer of high-yield crops and the significant production of greenhouse gases during its synthesis (Galloway et al., 2008; Smith et al., 2008).

However, recently researches have referred intensive farming as an effective means to diminish greenhouse gas emissions, arguing that even though emissions from production and application of fertlizers increased between 1961 and 2005, the net effect of higher yields associated to farming intensification since 1961 has avoided up to 161 gigaton of carbon emissions (Burney et al., 2010). In the same line Wise et al. (2009) concluded that increased farming yields could be as efficient as wind and solar energy technologies as a means to reduce CO_2 emissions.

The impact of intensive farming on forested areas is one of the subject characterized by the absence of consensus, on one hand there is a researchers group arguing that agricultural intensification can contribute to forest conservation (Tachibana, 2001; Shively and Pagiola, 2004) in the other side, the group which holds that intensification encourages deforestation (Pichon, 1996; Bilsborrow and Carr, 2000), and between them, those authors suggesting that the relationship is indeterminate (Angelsen and Kaimowitz, 1999).

Then, the question is whether the intensification of production systems required to satisfy the food requirements of an increasing population can be achieved while meeting acceptable standards of socio-economic and environmental quality. Responding to this question is a difficult task that requires great efforts and commitment by all stakeholders, and especially on the part of those responsible for the definition of agricultural policies, farmers on whose shoulders rests food production, specialists in various disciplines, whose opinion should be the starting point for the definition of policies, plans and government strategies. In the complicated matter ahead, an important step it would have techniques that facilitate the identification and monitoring of the processes of agricultural farm intensification.

Traditionally farm intensification studies have been mainly based on farm attributes gathered in surveys and censuses (Brown and Podolefsky, 1976; Turner et al., 1977; Turner and Doolittle, 1978; Shriar, 2000) that are notoriously difficult to monitor. However, with the availability of data from remote sensing satellites that are specifically tailored for broad-scale observation of the earth's land cover and along with the advent of personal computers that provide scientists with the opportunity to use powerful data processing techniques, the spatial approach to study agricultural intensification is gaining momentum (Duvernoy, 2000; McAlpine and Freyne, 2001; Kerr and Cihlar, 2003; Tappan and McGahuey, 2007). In the field of pattern recognition, there are multiple techniques that have been used in order to identify the essential properties of the categories of interest. Maximum-likelihood classifiers and Bayesian, are the most common parametric methods, the former seeks to find the parameter value that is best supported by the training data, so that the parametric form of the class-conditional probability densities must be known, while in Bayesian method, parameters estimation are considered random variables having a known prior density. For its part, non-parametric techniques, are appropriate for cases in which the shapes of the underlying density functions are unknown.

Following this trend, I applied ideas from the field of machine learning for pattern recognition of the multispectral response of farms with different levels of intensification, taking the benefits of using a satellite-based approach, among these, the coverage of large geographic areas provided by their synoptic view, the digital format of the data and its compatibility with geographic information systems, eliminating the need to digitize the information, facilitating the analysis at a considerable less cost than other methods. In the remainder of this chapter a summary of the agricultural intensification definitions and measurements proposed is presented, succeed by an overview of the agricultural intensification theories; next a general background to Venezuela is given, then the research objectives are summarised and finally the outline of the dissertation is offered.

1.1 Definitions and measurements of agricultural intensification

Frequency of cultivation, inputs such as fertilisers, pesticides, agricultural machinery and outputs, referred as to yields per unit of land, have been the three broad aspects used by most researchers studying agricultural intensification. Let us start with Boserup (1965) classic work, who defined agricultural intensification as "the gradual change towards patterns of land use which make it possible to crop a given area of land more frequently than before". This definition makes it obvious that to her a clear sign of intensification is an increase in the frequency with which a parcel of land is cultivated. This statement has been supported in many research works, some by adopting it just like that (Turner et al., 1977; Ruthenberg, 1980); whilst others by adding some input indicators (Brookfield, 1972; Turner and Doolittle, 1978; Shriar, 2000; Demont et al., 2007). Cropping frequency has also been included in an indirect way (McAlpine and Freyne, 2001; Thapa and Rasul, 2005) even though the formers declare that in their work land use intensity does not hold the meaning given by Boserup (1965).

Another essential aspect included in almost every agricultural intensification definition is the level of production or output per hectare, which has been expressed in some definitions as a monetary value. Tiffen et al. (1994) further defined it as an "increased average inputs of labour or capital on a smallholding, either cultivated land alone, or on cultivated and grazing land, for the purpose of increasing the value of output per hectare". Montilla et al. (2004) emphasizes the necessity of improving the productivity by extending the rational and balanced use of technological inputs as a means to increase the productive response, making land, labor and capital more profitable.

Some writers distinguish between input and output intensification, the former measuring the increases in fertilisers, pesticides, labor, irrigation, mechanization among the most frequently cited, and the later accounting for increases in the production obtained per hectare expressed by weight, caloric or monetary means (Turner and Doolittle, 1978; Lambin et al., 2000). Other authors have made distinctions between "labour-led" and "capital-led" agricultural intensification, the former comprising those situations characterized by more use of labour per unit of land for land preparation, weeding, manure application and harvesting, whereas the later implies more use of inputs such as fertiliser, pesticides, and agricultural machinery (Carswell, 2000). However, as pointed out by Aune and Bationo (2008) find a clear-cut among them in real's farm could be difficult, since in the same farm both intensification pathways can coexist.

Agricultural intensification studies do not always employ similar variables and units, which vary depending on the aim and research context, and also on data availability, such as cited by Turner and Doolittle (1978) who identified the food-ton or number of calories obtained over 20 years as the ideal measure of agricultural intensity. However, they could not use any of the above measurement considered suitable for them given the constraints on the availability of data required for that purpose.

Given the diversity of agricultural intensification definitions and its measurements, it was considered important to clarify the meaning adopted in this thesis. Therefore in this work, agricultural intensification is related to a set of land use patterns and management practices implemented by farmers with the aim to achieve high output per unit of land. It must be noted that output is mentioned in the definition, since achieving higher output as a means to increase productivity is the underlying reason to intensify farms, even though no output indicator was taken into account in clasifying the farms, basically because the 1996/1997 Venezuelan agricultural census, which was the main attributive farm database does not offer it, but the data required to calculate the area of each farm allocated to annual crops, orchards and forage, used as indicators of farm land use patterns, while the area under irrigation, stocking rates, mechanisation, equipment and labour index were used as input indicators. It is important to note that these variables were expresed as percentages to avoid grouping farms based on size.

A summary of some agricultural measurements used by researchers during the last 50 years is offered in Table 1.1.

Table 1.1. Main agricultural intensity measurements proposed

Authors	Measurement	Variables
Boserup, 1965	cropping frequency	forest-fallow, bush-fallow, short-fallow, annual-cropping and multi-cropping
Brookfield 1972	cropping frequency cultivation methods crop segregation	cultivation and fallow frequency clearing method, mounding, composte, terrace, irrigation
Brown & Podolefsky 1976	cropping frequency management	fallow period,enclosure, erosion control water control, ground preparation fertilisitation
Ruthenberg, 1976	cropping frequency	percentage of land under cultivation relate to the area total suitable for arable farming
Turner et al, 1977	cropping frequency	number of years of fallow for every year of cultivation, expresed as percentage
Turner & Doolittle 1978	cropping frequency management	percentage of time in cultivation crop protection, hydraulic control, soil fertility maintenance
Doan, 1995	percentage of cultivated area productivity	percentage devoted to vegetables and fruits production per unit of land or person
Shriar, 2000	proportion of cultivated area and its management	established plots, high value crop production, plowing, fertilizer, ranching intensity, intercropping pesticide, permanent crop,
McAlpine & Freyne 2001	cropping-fallow cycle (indirect way)	percentage of recent anthropogenic compared with primary vegetation
Thapa & Rasul 2005	cropping fallow cycle (indirect way) management	proportion of shifting cultivation horticulture, paddy cultivation, annual cash crops average/ha: fruit and wood trees cattle, pigs, goats and poultry proportion of production used for consumption
Tappan & McGaguey 2007	land-use-cover types	percentage of cropland, tree savannas, savanna woodlands gallery forest, woodland
Demont, 2007	cropping frequency management	cropping and fallow periods fertilisers, herbicides and insecticides
Alvarez, 2008	output, input	milk per hectare and per cow, cows per hectare, feed per cow

1.2 General overview of agricultural intensification theories

There are essentially two main theories used to explain the evolution of agricultural systems, both of which have a number of variants, but always relating agricultural change to population trends, since food production is considered to be the most important motivating force of agricultural activities.

The first theory is the Malthusian, which dates back to Malthus (1783), and which was followed by the neo-Malthusian (Ehrlich, 1968; Meadows, 1972; Dasgupta, 1995) which are based on the idea that population increase has the potential to outstrip agricultural production, thus leading to land destruction. In other words, neo-Malthusians believe that population growth causes people to start moving to other land to avoid starvation but subsequently destroying the new land as well, resulting ultimately in land fragmentation, environmental deterioration, poverty and famine. (Moseley, 2000)

The Boserupian theory, on the other hand, maintains that population growth is the major factor governing agricultural development. Boserup (1965) argued that it is this factor which triggers agricultural intensification by reducing the length of fallow period. In other words, agriculture systems characterised by long fallows are practiced when the population is low enough to allow it, whilst to feed an increasing population, agricultural systems must increase overall production by increasing the frequency of cultivation and by the gradual incorporation of weed control, fertilization, ground preparation and irrigation practices.

Even though these two models are considered to be oversimplifications of reality, they provide the basis for looking into the complexities surrounding agricultural change (Stone, 2001). As these complexities are closely related to the large variation in ecological, cultural and economic conditions that exist, it is hard to find a single model capable of explaining agricultural change as a common worldwide phenomenon (Grigg, 1982). Within current literature there is plenty of evidence both to support (Brown and Podolefsky, 1976; Turner et al., 1977; Ruthenberg, 1980; Netting, 1993) and refute (Conelly, 1992; Padoch, 1985) Boserup's theory of population pressure determining the level of agricultural intensity. Other assumptions made by Boserup have also been criticised. For example, it has been argued that land shortage is not the only stimulus for intensification, which can also be driven by risk reduction (Saunders and Webster, 1987) and social production (Brookfield, 1972).

Later on, the differences of Brookfield with the Boserup theory become profound when he suggests that the role that she gave to population growth make her model "reductionist" and "unilinear", pointing out that there are another variables with great impact, such as diversification of production and livelihood opportunities, investment, and finding new ways of using and managing resources, all of which are important roads to agricultural change.(Brookfield, 2001)

Boserup (1965) can be considered to be the main detractor of neo-Malthusians postulates. In her own words she states: "The neo-Malthusians collect all the evidence on the misuse of land and paint a picture of the world as a place where growing populations are pressing against a food potential which not only is incapable of increase but is even gradually reduced by the action of these growing populations". Based on evidence from Uganda, Carswell (2002) also reject the neo-Malthusian narrative, arguing that empirical findings have showed that the high rates of population growth in the Kigezi District had not had cause depletion of forest or erosion, nor had negative consequences on soil fertility and yields.

The controversy between the Malthus and Boserup theories seems to remain open, the research of Bilsborrow and Carr (2000) suggests that neither neo-Malthusian nor Boserupian can explain the linkages between population dynamics and land use changes in Latin America. On the other hand there is the research of Demont et al. (2007) in northern Cote d' Ivoire arguing that the Boserupian and Malthusian processes coexist, rather than contrast.

Intensification theories based on the existence of a positive link between population and intensification levels have been enriched by including market access as one of the principal driving forces, Pingali et al. (1987) suggest that transportation infraestructure and access to urban markets are as important as population. DeWilde (1967) included not only population and market access, but also the dominance of cereals in the cropping pattern as driving forces that might induce intensification in crop production and crop-livestock interactions.

McIntire et al. (1992) tackled agricultural intensification as an evolutionary process closely linked to the interactions between crop and livestock production. The nature of the relationship between population and crop-livestock interactions were described by him as an u inverted $\cap$, to symbolize that the level of interactions increase as the population does and reach maximum level at an intermediate population density followed by a specialisation process and lower interactions at higher population densities. Baltenweck et al. (2003)

argue that the linear evolutionary process of crop livestock interactions and integration postulated by McIntire et al. (1992) may not equally hold everywhere, because ecological and socio-economic conditions can lead to alternative or sub-pathways for intensification.

Hayami and Ruttan (1971) proposed the ’induced innovation’ thesis, which states that changes in production and productivity in the developing world are due to the adoption of new technologies. The induced innovation hypothesis has been the target of considerable criticism (Brookfield, 1972; Blaut, 1977; Olmstead and Rhode, 1993) as well as much support (Lipton, 1989; Lele and Stone, 1989; Turner et al., 1993). Lipton (1989) has made a distinction between agricultural intensification that results from population growth generating technologies to increase production efficiency and intensification which result from increasing populations inducing rises in labour use per hectare as labour becomes more plentiful relative to other factors. He also argued that as a population grows both types of intensification are required in order to meet the goals of poverty reduction, and to confront food availability and entitlement constraints. Lele and Stone (1989) concluded that agricultural intensification may take place spontaneously or as a result of policies and incentives to shift to crops of higher values, with the former happening as land is cropped more frequently in response to higher population densities.

The previous overview of agricultural change shows the diversity of approaches that have been proposed to study agricultural intensification, which are mainly based on the results obtained whilst studying the phenomenon in specific locations.

1.3 General background to Venezuela

Venezuela is a tropical country located in the American Continent in the north of South America. To the north it borders the Caribbean Sea, to the south Brazil and Colombia, and with Guyana and Colombia to the east and west respectively. It has an area of 916,445 km^2, and an estimated population by the year 2010 of 28.8 million (INE, 2007).

The commercial oil explotation started in Venezuela almost a century ago, since then, practically every aspect of the country have been marked by it. This influence is easy understood by looking at the magnitude of the conventional and non-conventional oil proven reserves, which are the largest in the western hemisphere and in the world respectively (PDVSA, 2007, 2010). Paradoxically, the enormous incomes coming from that advantage, specially those received during the recurring oil booms, instead to become the base to diversify and boost the economy, with the corresponding well-being of its population, have been associated to the increasing external dependence. A retrospective analysis of major macroeconomic aggregates show the speed and impact of the oil industry, which in just the first 8 years of commercial operation (1917-1925) displacement the agricultural exports as the main economic activity (Arias, 1993). Indeed, while agricultural production made up about one third of Venezuela's gross domestic product (GDP) in the 1920s, it shrank to less than one tenth by the 1950s. This downward trend has also characterized the development of the venezuelan agriculture during the last three decades, when agriculture had making up about 6% of GDP, as can be inferred from the analysis of Fig. 1.1.

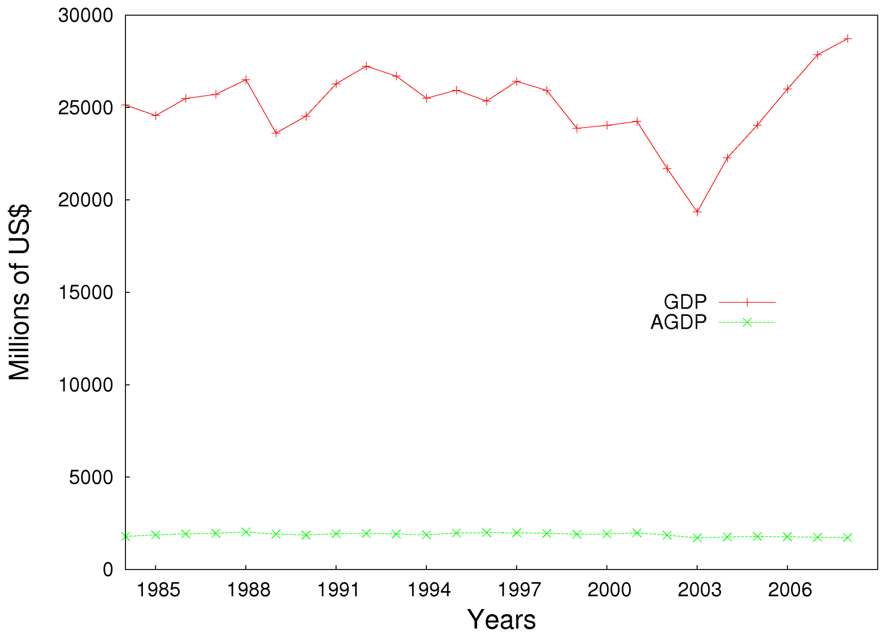


Fig. 1.1. Evolution of gross domestic product (GDP) and agricultural gross domestic product (AGDP); source BCV (2006).

In an attempt to explain the extent to which the increase in oil production and income was followed by a corresponding decrease in agricultural production delaying industrialization Rodríguez (1997) among others, uses the theory developed by Corden and Neary (1982) according to which, whenever a commodity brings a sudden increase of income in one sector of the economy, which is not matched by increased income in other sectors of the economy, causes a distorted growth in services and other non-tradables, which cannot be imported, while discouraging the production of tradables, which are imported, as was the case on the 1960s in the Netherlands after the discovery of natural gas, which is where the name of the problem comes from.

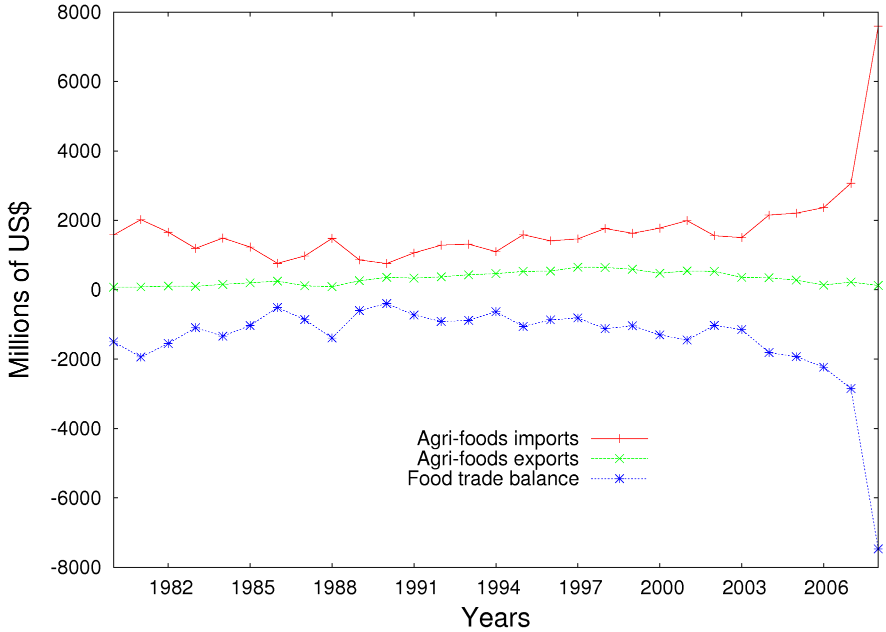
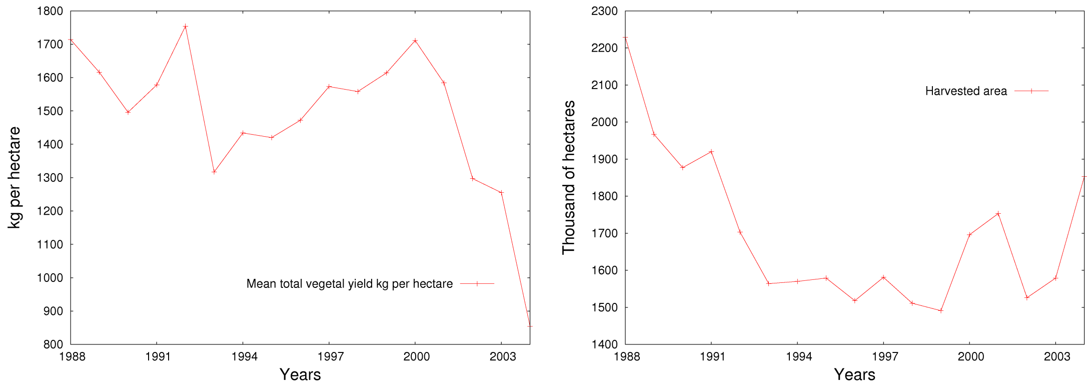


Fig. 1.2. Venezuelan agricultural trade balance; source INE (2009)

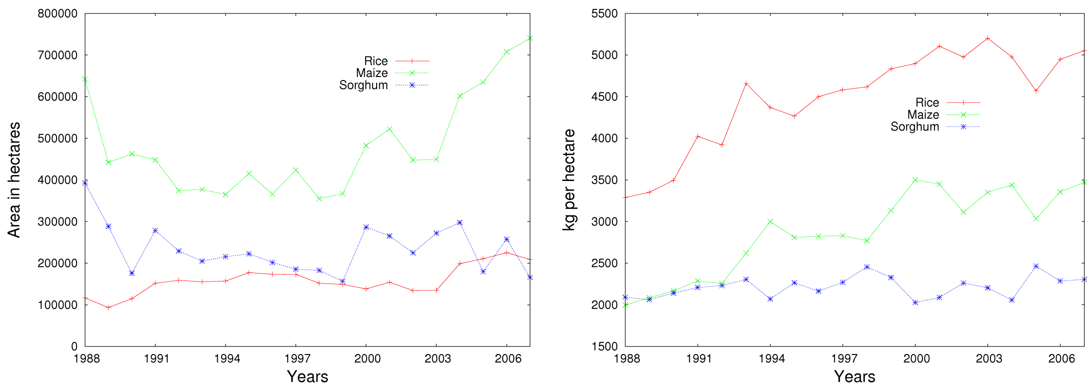
The lack of capacity of the Venezuelan internal supply to meet food demand occurs even though apparently the agricultural frontier of the country has increased by an average of 2,888 thousands ha per year, according to the reported figures of deforestation, making Venezuela one of the ten countries worldwide registering dramatic forest lost (UN2005). Nevertheless, the deforestation figures calls attention since these appear not be reflected in the trend of the area devoted to agricultural production, as apparently is from Fig 1.3 suggesting that both agricultural extensification and intensification have ocurred in Venezuela.



(a) (b)

Fig. 1.3. Trends in plant total production (a) and harvested area (b) from 1988 to 2006; source INNOVA (2009).

However, by analyzing the trend discriminated by items, intensification seems to prevail. Such is the case of maize and rice, which showed during the study period significant increases in yields (Fig. 1.4). Similar behaviors have roots and tubers, observing a slight increase in harvested area accompanied by a considerable increase in yields.



(a) (b)

Fig. 1.4. Harvested area (a) and yield per ha (b) of selected cereals from 1998 to 2007; source INNOVA (2009).

The evolution of the area harvested and the respective yield per hectare of fruits shows oscillations that prevent conclude whether its production is marked by the intensification or otherwise have been extensify. (Fig. 1.5).

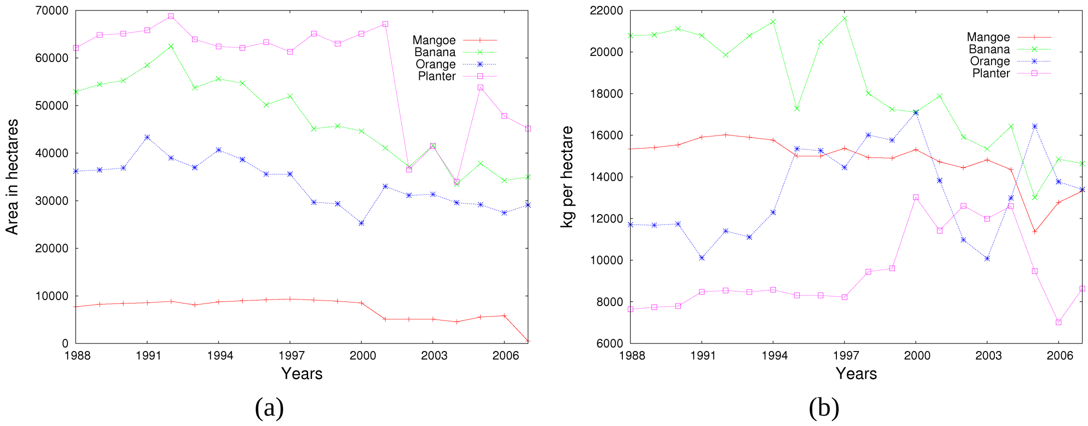


Fig. 1.5. Harvested area (a) and yield per ha (b) of selected fruits from 1998 to 2007; source INNOVA (2009).

In Venezuela there are a wide number of farming systems, which show great variability not only due to differences in agro ecological conditions but also with the surrounding areas. Table. 1.2 summarises Venezuelan farms classified by the number and area of holdings. Here the dominance of the latifundio is clearly evident with 46% of the total agricultural land area over 1000 ha in the control of 1%, whilst 75% of the up to 20 ha share not more than 5% of the total agricultural land.

Table 1.2. Distribution of exploitations according to size 1937-1997

	% of total holdings		% of total area	
	1937	1997	1937	1997
Under 1 ha	1.8	8.53	0.02	0.07
1 and under 5 ha	33.6	39.81	0.28	1.52
5 and under 10 ha	22.4	15.03	0.36	1.61
10 and under 20 ha	14.8	11.85	0.58	2.49
20 and under 50 ha	10.5	10.31	0.87	5.05
50 and under 100 ha	4.3	5.45	0.89	6.02
100 and under 500 ha	5.9	6.82	4.00	23.54
500 and under 1000 ha	1.9	1.20	4.00	13.25
1000 and over	4.8	0.99	89.00	46.45

Source: MAC (1997)

By looking at the results of the first and the last Venezuelan agricultural censuses carried out in 1937 and 1997 respectively, we can observe a steady diminishing in the % of total land area occupied by holdings bigger than 1,000 ha resulting in an increase in the % of the remaining holding intervals size.

Nevertheless, agricultural land is still clearly dominated by holdings over 1,000 ha, providing evidence of the scarce effectiveness of the land reform, which has mostly remained as a rhetorical exercise against latifundios, which have been pointed out as the main cause of land inequity.

Concern about the potential negative environmental impacts which according to some researchers accompany the processes of agricultural intensification, are echoed in several studies conducted in Venezuela, such in the one conducted by Fernández et al. (1998) who reported high vulnerability to water erosion due to inadequate mechanisation practices and strong rainfall in the venezuelan central high plains. In other areas, such as in the basin of Valencia lake and in the western plains, most of the land degradation was found to be due the indiscriminate use of chemicals. In both cases degradation seems to mainly be connected to land management practices associated with the intensification of farming systems in those areas.

Another pressing problem facing Venezuela is the level of poverty amongst its inhabitants. According to the 2001 population census, about 45% of the total Venezuelan population currently lives in poverty, and almost half of these people are living in extreme poverty. This phenomenon is even more dramatic among the rural population where nearly 70% of the population lives in poverty compared to 40% in urban areas.

The significant improvement in the performance of the agricultural sector will allow the triple objective of reducing external dependence, thus preventing the drain of financial resources, improve income and living conditions of the rural population as well as prevent the growth of the brutal misery typical of large cities in countries like Venezuela.

In this context, it is obvious that one of the greatest challenges facing the Venezuelan agricultural system for the foreseeable future is to achieve greater levels of food self-sufficiency, to diminish rural poverty and to enhance, or at least maintain, the natural resource base.

1.4 Objectives

The main objective of this thesis is to investigate through a representative sample of farms located in the Urdaneta municipality of Aragua state, whether by applying ideas from the learning machine field for pattern recognition farm intensification levels can be detected from a landsat satellite image. In pursuing it the following two objectives were identified.

1. To derive a farm intensification typology of the Urdaneta municipality farms based on census data.

2. To test the ability of the kernel adatron algorithm at establishing relationships between farm intensification levels and their spectral responses from Landsat imagery, by using the previously intensification typologies derived.

From the two main objectives stated, it is clear that the first one involves an unsupervised classification, since the problem is to group the unlabeled sample of farms into meaningful clusters from the intensification point of view. The aforementioned task was undertaken by applying two different approaches: a traditional hierarchical clustering and a self organizing map (SOM) belonging to learning machines, so that to measure the validity of both unsupervised approaches was the specific objective identified.

As the second pursued goal requires a supervised classification, a sample of the farms categories provided by the clustering previously applied were used as training instances to induct a relationship between the spectral properties of the land cover using landsat images and the corresponding clusters belonging to the identified clusters by applying the kernel adatron algorithm, in doing that to compare the class separation accuracy of the kernel adatron against a linear disciminant analysis, and also to explore the effects of different kernel functions over the kernel adatron algorithm's perfomance were the pursued specific objectives.

1.5 Outline of the dissertation

This thesis is structured in 6 chapters. This first chapter designed as a general introduction, offered an overview of the problem, including a brief review of major agricultural intensification theories and indicators proposed, followed by general information about Venezuela, ending with the presentation of both general and specific objectives pursued in this research.

Chapters 2 and 3 are devoted to the application of unsupervised classification techniques in order to identify the farm intensification typologies of Urdaneta Municipality, Aragua state in Venezuela. The main difference between the mentioned chapters lies in the clustering method employed. Since, in chapter 2 the traditional Ward's hierarchical clustering method was choosen, whereas in Chapter 3 the task was performed using the self-organizing maps which belong to the category of competitive learning network.

In chapter 4 the ability of the kernel adatron algorithm to detect the level of intensification of the farms included in the study based on the spectral characteristics recorded in a Landsat image using as training sets a sub sample of the labels obtained in the unsupervised classification is investigated. Additionally it is going to be used to investigate whether such an approach will attain comparable cluster accuracy as that achieved with traditionally supervised classification methodology as linear discriminant analysis while using only spectral information. Also, the effect of different kernel functions and their parameters on the accuracy of farm classifications from Landsat 7 ETM images will be investigated.

Chapter 5 offers an integrated discussion of the results obtained in the previous chapters and in chapter 6 the overall conclusions of this work are presented.

Chapter 2

Hierarchical Agglomerative Cluster based Segmentation

2.1 Abstract

With the aim to group farms according to their intensification levels, a conventional hierarchical agglomerative clustering using the Ward's criterion of minimum variance and the squared Euclidean distance as a metric was implemented by using the attributive data of 275 farms, randomly chosen from a population of 1,429 holdings located in the Urdaneta municipality of Aragua state in Venezuela, obtained from the Agricultural Census of 1996-1997.

The variables used in clustering farms were the proportion of land in cultivation and under irrigation, stocking rates, machinery, equipment and labour index, built from the data held on the Venezuelan Agricultural Census of 1996-1997. This selection was supported in previous research and was also influenced by the data collected in the aforementioned census, which was the primary source. Yield per hectare was not included, even though it is usually considered to be the best variable representing output intensification, because it was not collected in the census.

After applying principal component analysis and hierarchical clustering methods, three groups were identified and designated as 1 to 3 and rated from low to high, as determined by intensification levels. These were labelled extensive, semi-intensive and intensive. The

labels assigned to each farm are only applicable in the context of Urdaneta municipality, not nationally or regionally.

2.2 Introduction

Classificatory methods are intrinsic to the development of scientific theories (Aldenderfer and Blashfield, 1984). Biology and zoology taxonomy are perhaps the fields for which the first clustering techniques were developed (Jain and Dubes, 1988a). As an unsupervised technique, clustering deals with finding structure in a collection of unlabeled data, seeking to identify groups of objects that satisfy some specific criteria or share some common characteristics. An operational definition of clustering was stated by Jain (2010) as follows "Given a representation of N objects, find K groups based on a measure of similarity such that the similarities between the objects in the same group are high while the similarities between objects in different groups are low".

Clustering ability to describe a large collection of objects in a simple and understandable way has made of it as one of the most used methods of multivariate analysis, in a variety of domains for different types of application, so that, a large body of literature have been reported, such as Kettering (2006) who found over 1,000 papers appearing anually for which cluster analysis has a prominent role, given that cluster analysis allows organizing multidimensional data where visual perception fails. In agriculture usage clustering techniques have been also one the most used to address agricultural typologies, its usefulness has long been recognised (Benedict et al., 1944; Munton and Norris, 1969) since they allow the segregation of farms into simple and clearly recognisable classes, providing the essential picture to design and conduct successful agricultural policies at local, national or whatever other level. Typologies based on such diverse features as: peasant agriculture (Hensall and King, 1966); types of agricultural land (Munton and Norris, 1969); conventional and alternative farming systems (Sampath, 1992); selection of appropriate areas for introduction of new technologies (Hardiman et al., 1990); the commercial orientation of farmers (Makhura et al., 1998) farming system typification (Picazo and Hernandez, 1993; Kobrich et al., 2003); cattle typologies (DaSilva et al., 2003; Urdaneta et al., 2004; Milán et al., 2006); dual purpose cattle productivity (Gómez et al., 2002); and farm segmentation (Iraizoz et al., 2007) can be cited among many others.

This chapter intends to classify a sample of farms located in the Urdaneta municipality of Venezuela Aragua state by applying principal component, hierarchical cluster to a set of land use and land management intensification indicators, taking advantage of recognising the multivariate relationships between variables to find the underlying data structure to form homogeneous groups. These methods have demonstrated to be efficient in dealing with classification issues related to a wide variety of agricultural topics (Rosenberg and Turvey, 1991; Bernhardt et al., 1996; Makhura et al., 1998; Kobrich et al., 2003; Milán et al., 2006; Iraizoz et al., 2007), including agricultural intensification issues, (Thapa and Rasul, 2005; MacLeod and Moller, 2006) among others.

Farm classification by intensification levels is an essential task to gain an integral understanding of the agricultural intensification process which is becoming increasingly important due to its role not only in meeting the food and fibre requirements of a population which, according to the most modest projections, will still reach 9,2 billion by 2050 (United

Nations, 2007), but also due to its potential collateral environmental effects (Shriar, 2000; MacLeod and Moller, 2006). As already indicated in chapter 1, the role of agricultural intensification in increasing food production is uncontroversial but the long list of secondary effects related to it has been the cause of much concern. As a result many studies focus on the nature of the relationships between intensification and issues like deforestation (Pichon, 1996; Meertens et al., 1996; Angelsen, 1999; Bilsborrow and Carr, 2000; Tachibana, 2001; Shively and Pagiola, 2004), rural poverty (Hazell and Ramasamy, 1991; Carswell, 2000; Niazi, 2004) and biodiversity (Matson et al., 1997; Chamberlain et al., 2000; Decaens and Jiméenez, 2002; Smith et al., 2005). Although these studies have been conducted worldwide over the last two decades, controversy over the nature of their relations still remains.

Knowing the nature of these relationship is essential in order for policy-makers to conduct successful planning tasks, especially in countries such as Venezuela where, in spite of there being extensive areas capable of agricultural production, nearly 50% of the internal food requirements of recent years have been satisfied by importations (BCV, 2006). In addition to external food dependency, are the poverty, which affects about 70% of the rural population (IFAD, 2007) and the recurrent degradation of the natural resource base (Pla, 1990; Lozano et al., 2002; Rodríguez et al., 2003)

It is in the described context where the importance of a farm intensification typology as a strategic tool to face the challeges emerges. In that sense the proposed Urdaneta farm intensification typology aims not only to serve as a basis to segment farms in the traditional manner, but also as a first approximation to test the possibility of incorporating remote sensed imagery as a way to study the intensification process.

In the remainder of this chapter a brief description of the study area is provided, then, the methodology is outlined, followed by a summary of the main results, and finally discussions and conclusions are presented.

2.3 Study area.

The area of study is Urdaneta which is one of the 15 municipalities of the Aragua state in Venezuela, occupying 2,024 km^2 and representing 29.25% of the total area of this state. It is located in the so-called central region of Venezuela, which is the most densely populated with about 35% of the entire urban Venezuelan population located in Distrito Capital, Miranda, Carabobo, and Aragua states.

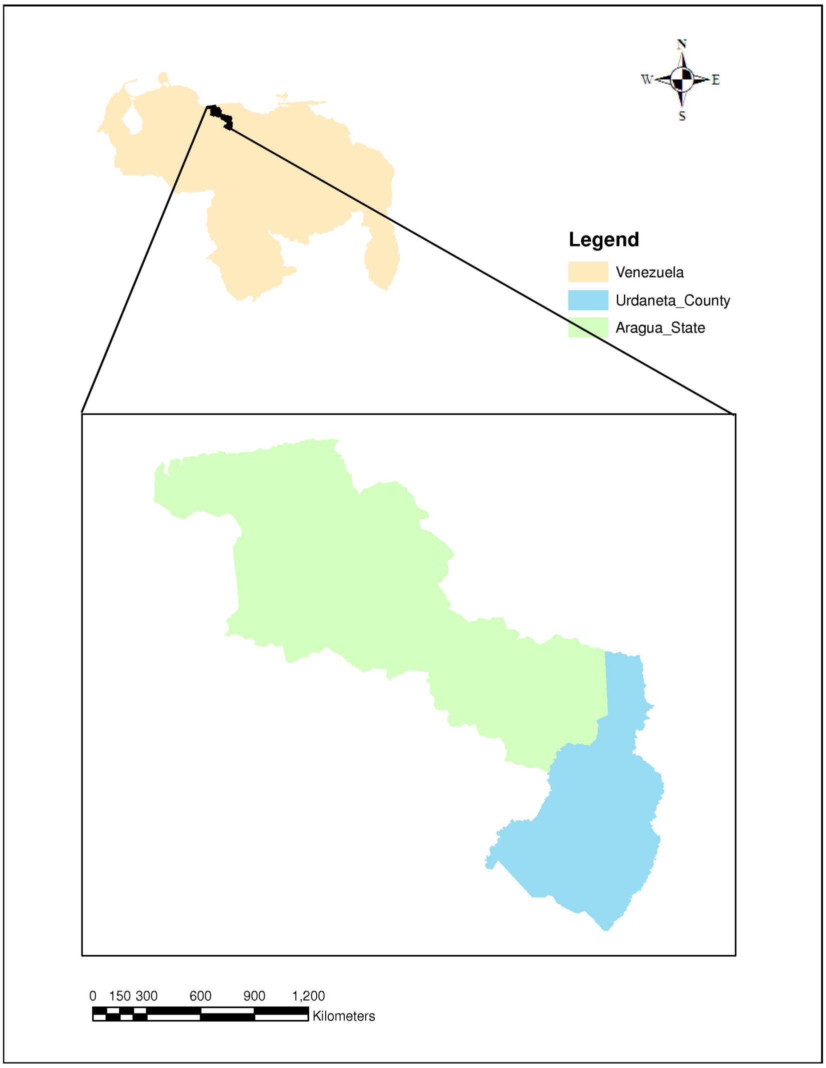


Fig. 2.1. Study area

It is connected to the main cities of the country via the national road network, which is generally in good condition and allows vital communication all year round. The distances from Urdaneta to Caracas, Maracay and Valencia are about 200, 150, and 180 km respectively.

Like most Venezuelan rural areas, it has a low population density, about 9 inhabitants per km^2 concentrated in Barbacoas and Taguay. It is a zone of moderate rainfall (1,200 mm/yr) with an average temperature of 26.7°C. The rainy season extends from May to October and the dry season from November to April. Interannual variability and seasonality of rainfall, makes it one of the main constraints to increased yield per hectare. Soil fertility ranges from medium to low. The indigenous vegetation is typically dry tropical forest. (MARNR, 1983)

This municipality has about 144,463 ha of agricultural land distributed into 1,429 holdings. Table 2.1 shows Urdaneta farms classified by the number and area of holdings. Holdings of more than 100 ha represent just 18% of the total holdings but control about 89% of the total agricultural land, whilst holdings less than 20 ha (62% of the total holdings) occupy less than 3% of the total agricultural land, resembling very much the Venezuelan pattern as a country, shown in chapter 1

Table 2.1. Number and area of holdings by size of Urdaneta county

	Holdings		*Area*	
	Number	%	ha	%
Under 1 ha	33	2.31	24	0.02
1 and under 2 ha	161	11.27	143	0.10
2 and under 5 ha	258	18.05	641	0.44
5 and under 10 ha	301	21.06	1,543	1.06
10 and under 20 ha	130	9.10	1,522	1.05
20 and under 50 ha	174	12.18	5,171	3.58
50 and under 100 ha	116	8.12	7,363	5.10
100 and under 200 ha	101	7.07	12,909	8.94
200 and under 500 ha	88	6.16	26,396	18.27
500 and under 1000 ha	36	2.52	23,686	16.40
1000 ha and over	31	2.17	5,077	45.95

Source: MAC (1997)

The Urdaneta municipality has been selected as an appropriate study case to investigate agricultural intensification since it presents a number of interesting features, such as being a transition between the major urban and rural Venezuelan zones, comprising a wide diversity of agricultural land uses, ranging from annual crop to crop-livestock systems with different managements and a great diversity in holdings areas. These characteristics make Urdaneta a representative case of many of the Venezuelan agricultural areas. Data availability was also an important factor which influenced its selection as study area.

2.4 Data and methodology.

2.4.1 Data

Data used in this research consist of 275 of the total of 1,429 holdings reported in the Urdaneta municipality of Aragua state by the 1996/1997 Venezuelan agricultural census. The census was carried out by the Ministry of Agriculture and Livestock in collaboration with the Central Office of Statistic and Information, the Central Bank of Venezuela and Venezuelan universities, following FAO guidelines (MAC, 1998).

The information was obtained by interviewing landholders directly. The enumeration was based on municipal and parochial maps available from the XII census of population and housing resulting in 5,422 enumeration areas. Holdings were identified by 11 digits; the first eight numbers referring geographical location, which in blocks of two digits indicating state, municipality, parish and sector while the last three digits identify the holding itself (MAC, 1997).

To guarantee the quality of the information collected, prior to the application of the census questionnaries in the whole national territory, an experimental census was organised in Yaracuy state, to test the framework, the questionnaires and operational aspects; after that, a professional team conducted a permanent quality control while the census was carried out (MAC, 1997).

2.4.2 Variable selection

The first step was to identify an appropiate and meaningful set of variables on which to group Urdaneta's farms by intensification levels. Since the agricultural holding has been recognised as the best unit to build an agricultural typology (Kostrowicki, 1977), in the present research the advantage of having detailed data provided by the 1996/1997 Venezuelan agricultural census was exploited. Nine variables were chosen to segment farms by degree of intensification. These were labelled as land use and land management practices as representative intensification indicators in terms of the specific conditions of the study area.

Land use was represented as the percentage area within each farm on which agricultural production is carried out, which was divided into annual, orchard and forage given that these three categories comprise the main productive uses of farm land observed in the study area. Annual cropping percentage (ACP) was calculated by adding up the area under cultivation at the time of the application of the census questionnare plus the area being prepared for annual cropping, mostly represented by maize, sorghum, sweet chili and peppers. Orchard cropping percentage (OCP) accounted for the areas devoted to those crops whose productive cycle exceeds a year, which in the study area are mostly plantations of mandarins, lemons, bananas and mangoes. Forage percentage (FOP) included the pastures areas.

The general category of land management included six indicators accounting for some of the most common management practices used by researchers to reflect farm intensification. The first three, irrigation percentage and machinery and equipment index could be considered as a rough indirect output measurement, since it is supposed that their potential contribution to increase yields per hectare is an important factor inducing most farmers to use them.

Cropping irrigation percentage (CIP) constitutes an important intensification indicator (Turner and Doolittle, 1978; Sampath, 1992; Caraveli, 2000; MacLeod and Moller, 2006), especially for those areas with long dry periods, as water availability allows cropping even during the dry season, and also to minimise the stress caused by drought to plantations and grazing areas, so output per area may increase as a result.

Machinery index (MAI) and equipment index (EQI) as indicators of the capacity to undertake ground preparation tasks have long been recognised as an indicator of farm intensification (Brown and Podolefsky, 1976; Turner and Doolittle, 1978; Caraveli, 2000; MacLeod and Moller, 2006). The first was estimated by adding up the number of tractors of up 80 hp plus tractors more than 80 hp multiplied by two, while the second was the result of adding up the equipment number, in both cases the total number was divided by farm area.

Stocking rate (STR), expressed as animal units per unit of grassland, has been referenced as the index of the intensity of agricultural livestock production (Seré, 1983; Boyazoglu, 1998; Shriar, 2000; MacLeod and Moller, 2006; Milán et al., 2006). Animal units calculations used 1.5 units for bulls, 1.0 unit for cows, 0.75 units for heifer and steer yearling, 0.5 unit for store and 0.25 for calves.

Permanent staff ratio (PSR) and temporal staff ratio (TSR), were included to account for the number of workers per hectare per year, which will be greater for intensive systems, especially for those farms advocated to orchard crops (Doan, 1995).

2.4.3 Statistical Analysis

Since type is essentially a taxonomic concept based on the similarities between the entities under study (Kostrowicki, 1977), the data were analysed by applying principal components and agglomerative hierarchical clusters. Cluster analysis methods are usually complemented with principal components analysis, given the power of the later in reducing dimensions, providing distance measures and detecting cluster structure. These abilities have contributed to making it the most used multivariate analysis technique (Jolliffe, 2002).

Principal component analysis (Hotelling, 1933) is a non-parametric method that works by replacing the original variables with a linear combination of them and expressing it as a set of components which, by definition, are orthogonal and based on statistical variance, so that the components are sorted in descending importance; hence the first component accounts for most of the variance in the data, the second less than the first but more than the third, and so on until the last, explaining all the residual variance. Principal component analysis also forms the basis for hierarchical multivariate methods, since it provides the scores of the observations projected into a new n-dimensional space allowing the calculation of the distance between them, essential in clustering observations by their similarities.

The identification of homogeneous intensification groups was carried out using the clustering method proposed by Ward (1963), which as a hierarchical agglomerative technique works by a progressive fusion of the N entities or cases into groups, until every entity is in the same cluster. The criterion applied by this technique in selecting the entities to be joined seeks to minimise the loss of information resulting from its fusion, based on measuring the squared deviation so that prior to combining any entities in a cluster, the squared deviation is calculated by merging at each step those entities accounting the minimum increase in the error sum of squares from the mean of its cluster (Everitt, 1974).

The general procedure applied can be summarised as follows: first a matrix containing the nine indicators previously selected for each of the 275 farms included as a representative sample of the intensification of farms located in the Urdaneta municipality was created.

This matrix was pre-processed and analysed with CIRAD (1989) software. As a pre-processing task, an analysis of correlation was applied to find out whether there existed variables that should be excluded due to their high correlation coefficients. Then a principal component analysis was performed, using the correlation matrix over standardized data, to avoid distortion as a consequence of diverse and disparate units of measurement of the variables.

The next step consists of deciding the number of components to be retained, in this matter there is not a unique rule. The most common criterion referenced is that of Kaiser (1960) according to which only those components which have eigenvalues of 1.00 or greater should be considered. Jolliffe (1972) reported that Kaiser's criterion is too strict and suggested retaining those components with eigenvalues greater than 0.7 instead. There is also a simple but arbitrary rule of thumb according to which as many components as necessary should be included to be able to explain at least 75% of the variation.

The score of the observations in the components retained were used for cluster analysis by choosing the option maximization of momentum order two of CIRAD (1989) software, which is based on the Ward (1963) method, seeking a maximum variance between clusters while minimising the variance within them, based on Euclidean distance. As an option belonging to agglomerative hierarchical clustering, it starts with including as many cluster as entities decreasing in a stepwise manner, so that at the end there is just one cluster containing all the entities.

The resulting clustering quality was estimated by calculating the coefficient of incidence (Zhao and Karypis, 2004) and silhouette (Kaufman and Rousseeuw, 1990). The first compares the actual and idealized proximity matrices, while the second evaluates the distances of objects within a cluster in terms of its contribution to the overall cohesion or separation of a cluster. Additionally, based on a discriminant analysis (Fisher, 1936), the quality of the groupings is assessed 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). Discriminant analysis is carried out through the 7M routine of the sofware BMDP (Dixon et al., 1981). The appropriate number of cluster was decided by repeating the clustering procedure while varying the cardinality of the number of clusters, in order to choose the number of clusters that yielded the best quality.

2.5 Results.

Table 2.2 shows the correlation coefficient between all pairs of variables, it can be seen that the variables correlate fairly well but not perfectly, which is a desirable condition to perform cluster analysis, so that all of them can be kept. The absence of a correlation coefficient greater than 0.9 indicates that multicollinearity was not a problem, this was confirmed by calculating the determinant of the correlation matrix which being equal to 0.148 allows us to discard it completely (Field, 2005).

Table 2.2. Matrix of correlation

	2 ACP	3 OCP	4 FOP	5 CIP	6 STR	7 MAI	8 EQI	9 PSR	10 TSR
2	1.000								
3	-0.201*	1.000							
4	-0.307*	-0.106*	1.000						
5	-0.033	0.432*	-0.070	1.000					
6	-0.221*	-0.096	0.462*	-0.062	1.000				
7	0.057	-0.002	0.015	0.246*	0.016	1.000			
8	0.033	0.039	-0.009	0.209*	0.098*	0.786*	1.000		
9	0.133	0.084	-0.094*	0.063	-0.065	0.056	0.025	1.000	
10	-0.011	0.270*	-0.053	0.332*	-0.060	0.137*	0.174*	0.052	1.000

The correlation is significant at the .05 level

There is no unique rule for deciding the number of components that should be retained. On the contrary, this is one of the aspects which remains open, so that prior to making the decision the results were analysed by looking at them in light of the Kaiser (1960) and Jolliffe (1972) criteria. In that sense, as can be observed in Table 2.3, the former criterion is fulfilled by considering only the first three components since from the fourth the eigenvalue becomes less than 1, while by taking into account that the fifth component should be retained to be able to explain at least 75% of the variance. Nevertheless it was decided to include the next component, bringing the total variability to 86.40% with an eigenvalue of 0.672, ensuring that any important component is discarded, it is better keep too many rather than too few components (Kobrich et al., 2003).

Table 2.3. Eigenvalues of the matrix of correlation

Component	Eigenvalue	Component	Accumulative
1	2.114	23.49	23.49
2	1.726	19.18	42.66
3	1.490	16.55	59.22
4	0.974	10.83	70.05
5	0.800	8.89	78.93
6	0.672	7.47	86.40
7	0.533	6.15	92.55
8	0.472	5.24	97.79
9	0.198	2.21	100.00

2.6 Determining the number of clusters

Table 2.4 shows the results of several statistical criteria found after a linear discriminant analysis for the purpose of determining the appropriate number of sets in which the input data should be segmented. As can be appreciated, the clustering procedure was repeated to observe how the groups changed as a function of the number of sets used. It is easy to realize that in terms of the variability explained by the model (r^2), the most appropriate number of clusters is three. Likewise, the best separation between clusters appears to be achieved by using three groups Judging from the coefficients of impact and Silhouette. It is also noteworthy that with this number of clusters, it is possible to obtain the lowest value for the statistic Wilks' lambda, while Pillai, Hotelling, and Roy's tests reached their maximum, indicating better performance.

Table 2.4. Impact of cluster's number on cluster quality using discriminant analysis after hierarchical procedures.

$Number\ of\ clusters$	$\%C$	$W\lambda$	PT	T^2	RM	r^2	i	$Slh.$
2 *clusters*	64.1	0.26	0.79	2.40	2.30	0.30	0.27	0.20
3 *clusters*	92.1	0.11	1.31	3.83	1.94	0.71	0.70	0.50
4 *clusters*	88.3	0.15	1.21	3.36	2.30	0.60	0.50	0.32
5 *clusters*	76.2	0.22	0.93	2.50	2.15	0.48	0.30	0.28

%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; i: Incidence; Slh: Silhouette

The two-dimensional representations of the farms in the two first components were analysed as a means to visualize the three farm groups referenced, which are rendered in Fig. 2.2 designated 1 to 3, rated from low to high, as determined by intensification levels and labelled extensive; semi-intensive and intensive.

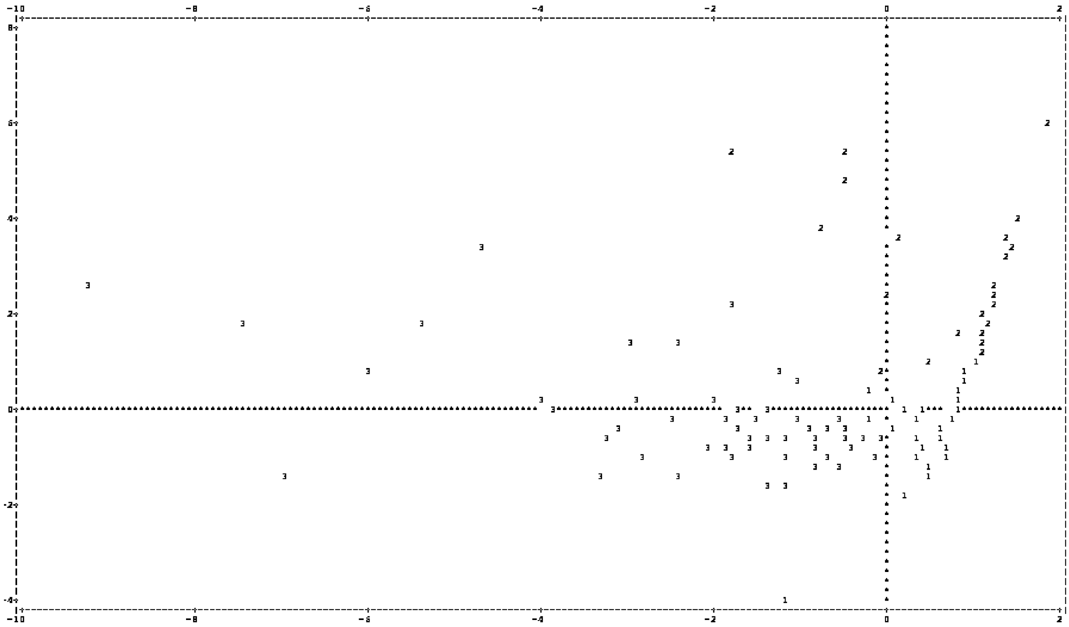


Fig. 2.2. Scatterplot of farms' classes in the first two components

The membership list generated by the CIRAD (1989) programme indicated 167 farms belonging to cluster 1, which, according the descriptive statistics is mostly composed by those farms mainly advocated to short-cycle crops and to a lesser extent to the livestock with low stocking rate, lack of infrastructure for irrigation, and machinery and equipment indices extremely low, with a predominance of regular workforce whose preeminence can be attributed to machinery and equipment lack, based on the features mentioned farms in this group were classified as extensive production systems.extremely low machinery and equipment index, low stocking rates, resembling extensive mixed agricultural production systems.

This group is integrated by 40 farms whose productive activity revolves around livestock and short-cycle crops. By analyzing the descriptive statistics for stocking rates, irrigation infrastructure and machinery and equipment indices in this group and compare them with those of farms included in farms class 1, one can infer that they hold a greater level of intensification that of the latter group, so that for the purposes of this research as a semi extensive production systems.

The salient features of the 68 farms included in this group are related to the provision of irrigation facilities, machinery and equipment, which have been recognized as indicators of Intensification of agricultural production systems, and also having higher permanent and temporary labor ratios. The conjunction of these features make us assume that the aforementioned farms exhibit a higher degree of intensification, so that, for the purposes of this thesis those farms are label as intensive production systems. Within the farm productive activities that make up the third group includes the short-cycle crops and permanent crops.

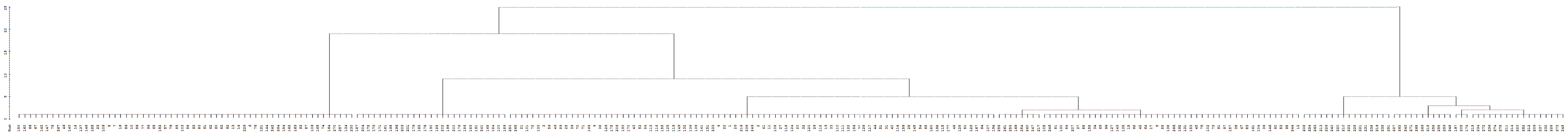


Fig. 2.3. Dendogram.

2.7 Discussion and conclusion

Numerous studies have demonstrated a link between the area devoted to agricultural production, irrigation facilities, stocking rates and machinery index and farm intensification (Brown and Podolefsky, 1976; Turner and Doolittle, 1978; Sampath, 1992; Caraveli, 2000; Shriar, 2000; Thapa and Rasul, 2005; MacLeod and Moller, 2006). In this chapter Urdaneta municipality farms were classified by intensification levels based on the referred variables. The clusters were labelled extensive; semi intensive and intensive. These patterns have been previously reported in the agricultural intensification arenas using unsupervised statistics techniques at farm level (Thapa and Rasul, 2005) and also at national level (MacLeod and Moller, 2006). The intensification groups confirmed in the Urdaneta municipality have also been referenced in earlier studies using ranking methods (Shriar, 2000), output per hectare in monetary terms (Andersen et al., 2006), milk production systems in tropical South America (Seré, 1983). Differences in the number of groups when compared with other intensification typologies such as those dividing farms into only extensive and intensive (Kerr and Cihlar, 2003) or those incorporating low and high extensive and intensive farms (Baltenweck et al., 1998; Urdaneta et al., 2004) can be explained by the intrisic characteristics of the farms under study, and also by the pursued goal, which in the present research, is clustering the sample farms of Urdaneta municipality into meaningful clusters from the intensification point of view, providing the labels to be used as training instances to induct a relationship between them and the spectral properties of their land cover as recorded in landsat image.

Results indicate that extensive mixed farms are by far the predominant systems in the area accounting for 60% of the total farms, followed in importance for intensive systems accounting for nearly 25% of Urdaneta farms and lastly semi intensive systems account for the remaining 15%

Chapter 3

Unsupervised Classification using a Neural Approach

3.1 Abstract

In this chapter have been applied statistical learning ideas for the task of identifying typologies of farms. Previous attempts to group types of farms such as hierarchical and non hierarchical clusters have often found some technical problems. These problems arise because in many cases it is very hard to make unbiased decisions about the appropriate number of clusters. However, when self-organizing Kohonen networks are used as a method of clustering, an objective selection of the number of clusters is provided. While some problems still remain e.g. repeatability of results, this research shows the beginning of a new series of applications of machine learning for unsupervised classification of farms.

3.2 Introduction

As evident in previous studies, one of the main difficulties in designing public policies for the agricultural sector has been the definition of farm typologies. This has been achieved primarily through the development and application of quantitative methods in the field of cluster analysis, which has significantly expanded our understanding about the possible subsets of farm typologies within a given population of farms. However, while these methods have been very useful in the process of selection of representative farms, many of them have the limitation that the decision on the appropriate number of clusters is quite subjective, and most of the time, it is necessary to preprocess the data using feature extraction techniques in order to reduce the dimensionality of the problem.

Kostov and McErlean (2006) in an attempt to solve the problems described above, suggested to clusters using the technique of fitting mixture distributions, and then apply a likelihood ratio as an objective test to decide the appropriate number of clusters within the population of farms (Everitt, 1993). Although this method is promising, among its disadvantages are included the need to preprocess the input data to extract features, met the basic assumptions of parametric statistics, and due it is a memory-based technique, its use is impractical in real-time applications (Hastie et al., 2001; Bishop, 2006). However, if a neural approach is used, as for instance the Kohonen self-organizing networks, the number of groups present in the data will be naturally identified in a two-dimension weight maps as the network adopts a stable configuration after a process of self-organization or training. Generating without any preprocessing of the input data, a representation which is quite flexible, uniform for patterns presented to it, and also preserves the topology of the input space.

On the basis of these ideas, the aim of this chapter is focused on identifying clusters of farms present in the population of data under analysis, and therefore be able to assign each observation a typology label that tells to what class each observation corresponds. Additionally, it seeks to validate the clustering found and compare the results with the hierarchical method used in Chapter 2. The structure of the chapter includes four sections: the first deals theoretical and methodological aspects of previous researches that support the work. The second describes the methodology used. In the third section analyzes and discusses the results and some conclusions are drawn in the fourth section.

3.3 Self-organizing networks

The self-organizing network refers to a set of methods that are characterized by inducing a non-supervised clustering n-dimensional data under a restricted topology living in an m-dimensional space, $m < n$ (Lippmann, 1987). In other words, the restricted topology structure is designed to present one of the possible m-dimensional similarities between the original patterns.

It is easy to see that the following elements are part of the self-organization:

- An arbitrary set of input patterns n-dimensional
- A set of processing units (PU) n-dimensional
- A network or interconnect structure restricted the PU that somehow represents the ratio of m-dimensional desired similarity
- A measure of distance defined in the space of n-dimensional
- A measure of distance defined in the restricted network of m-dimensional structure

The PU are points in space $\mathbb{R}^n$ defined by their position vector w, usually called weight vector. Each PU is in turn a set of PUs close in the network, those with which it is connected in their immediate environment and make its neighborhood. The structure of the neighborhood depends on the topology of the network and the size of the environment considered. In terms of distance functions, the Euclidean or Manhattan are used.

The process of finding the self-organizing mapping is called the network training. The training used is a competitive one, where only one unit, the winner, and possibly some elements of its neighborhood are adjusted according to the pattern presented. After the restricted topology network structuration with N PUs and define the distance functions, training can be summarized in the following steps:

1. To initialize weight vectors w_i of PUs randomly $\forall i \in 1 \ldots N$
2. To select randomly and uniformly one of the input patterns
3. To find the PU that is closer to the pattern selected. The PU is the winning unit (WU)
4. To make a correction vector w of the WU. The correction makes the PU resembles more the pattern shown
5. If necessary, conduct a weaker correction on w_i belonging to its neighborhood
6. To repeat (2) - (5) a fixed number of iterations or until the corrections are smaller than a preset threshold

The key point in building the mapping is that the degree of closeness between the PU is measured on the structure of the network, while the degree of closeness between the PU and a presented pattern is measured on the space $\mathbb{R}^n$. As training progresses the units are distributed evenly over the input space occupied by the patterns presented. The correction factor decreases with time until eventually the network reaches a stable configuration. At the end of training, the locations of the *PUs* of the network are indicative of intrinsic statistical features contained in the input patterns "(REF)." The network is specialized on the basis of competition in the sense that the units responsible for fragments of related information are close and can interact through short connections. This change by which the *PU* and the neighborhood go from being close only to the network structure, to be close in n-dimensional space, is the process known as self-organization.

Traditionally, networks of this form are used for data clustering, mapping features, vector quantization and compression, applications where features like dimensionality reduction and determination of a bounded number of prototypes are vital. However, because the resulting representation is quite flexible, uniform patterns is presented and preserves the topology of the input space, since there is a similarity between the structure of the network and two-dimensional mosaics of land uses that occur in an agroecosystem, it seems natural to its use in the problem of representation of farms. Next, it describes one of the most widely used self-organizing networks: The Kohonen Network Features.

3.4 The Kohonen Features Network

The Kohonen feature network (Kohonen, 1995) is the most widely used self-organizing method, which typically consists of a layer of processing units connected and forming a restricted network topology. Commonly the *PU* in a line (1D) of N units open at their ends, or closed in a ring, but also can be configured as a rectangular grid (2D) of $p \times q = N$ units.

Learning occurs as described, however, it should make the following points: Both the WU and a number of units within its neighborhood are adjusted so that the weight vector of the units is more like the input vector. Units surrounding the WU fit more smoothly according to the topological distance that separates them from the WU. Weight vectors w_1 are adjusted by the Kohonen learning rule:

$$w_i(t+1) = w_i(t) + \varepsilon(t)\,\Phi_{ij}(t)\,[x(t) - w_i(t)] \tag{3.1}$$

where $w_i(t)$ is the weight vector of the unit i in the istant t, $\varepsilon(t)$ is the learning rate, $x(t)$ is the input vector presented in t and $\Phi_{ij}(t)$ is a function of proximity between units i and j, given by:

$$\Phi_{ij}(t) = exp\left(\frac{-\tilde{b}(ij)^2}{\sigma^2(t)}\right) \tag{3.2}$$

Here, $\sigma(t)$ is the width parameter and the function $\tilde{d}$ measures the distance of units i and j in the network. In a square grid, let (f_i, c_i) position, row and column respectively, which is the unit i and (f_j, c_j) the position of unit j. The distance $\tilde{b}(i, j) = [(f_i - f_j) + (c_i - c_j)^2]^{1/2}$ is the Euclidean distance from positions of units in the network.

The distance function in n-dimensional space, one that measures the closeness between the pattern k and PU i is the Euclidean distance:

$$d\left(x_k, w_i\right) = \left(\sum_{l=1}^{n}\left({x_k}^{(l)} - {w_i}^{(l)}\right)^2\right)^{1/2} \tag{3.3}$$

Both the learning rate $\varepsilon(t)$ as the width parameter $\sigma^2(t)$ are assumed linear functions decreasing in time. Its functional form is:

$$\varepsilon\left(t\right) = m_\varepsilon\left(t - t_0\right) + \varepsilon_0 \tag{3.4}$$

$$m_\varepsilon = \frac{\varepsilon_f - \varepsilon_0}{t_f - t_0} \tag{3.5}$$

$$\sigma\left(t\right) = m_\sigma\left(t - t_0\right) + \sigma_0 \tag{3.6}$$

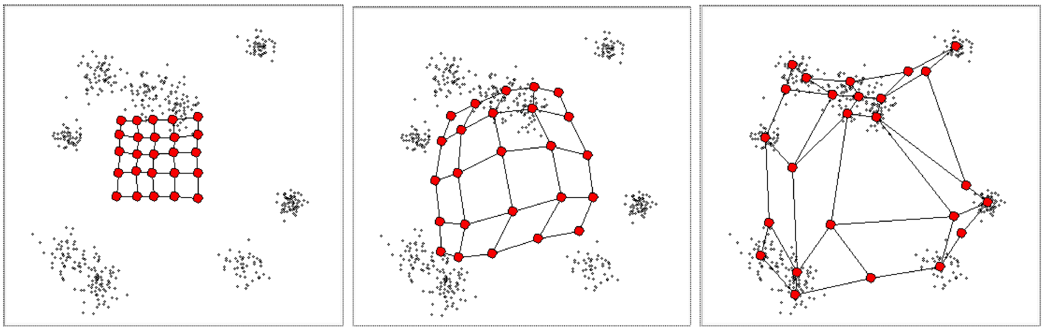
$$m_\sigma = \frac{\sigma_f - \sigma_0}{t_f - t_0} \tag{3.7}$$

The subscripts 0 and f indicate initial and final values, respectively. Note that because both the learning rate as the width parameter are decreasing functions over time, adjustments made on the vectors of weights are increasingly smaller as learning progresses. This indicates that the network is stabilized in the final stages of training.

3.5 Data preprocessing and methods

Before proceeding with the clustering process, a neural network must be trained. To do so is considered a Kohonen self-organizing map (SOM) (Kohonen, 1982) consisting of a set of nodes (N) that are arbitrarily initialized to lie in the plane of a two-dimensional grid (Fig. 3.1a). The training process aims to bend the two-dimensional plane so that the nodes of the grid approximate the training set distribution (black dots) as well as possible (Fig. 3.1b). Once the model fits, the observations can be mapped onto two-dimensional grid (Fig. 3.1c).

The observations are processed one at a time, with the purpose of finding, in terms of Euclidean distance within the grid, the node N closer to the observation x. Thus, each node N and its neighbors are moving toward the observations x with every update of the network. It is important to note that the distances are integers that are defined within the space described by the topological coordinates of each node in the network. Consequently, the net effect of each update is the movement of the nodes to the input data or observations, but keeping the two-dimensional spatial relations between nodes. Thus, preserving the neighborhood topology map, the structures in the input space can be discovered through the exploration of map features.



(a) (b) (c)

Fig. 3.1. Process of training a Kohonen self-organizing map (SOM). The black dots represent the training set drawn from the data distribution. First, the SOM nodes (red dotted lattice) are placed arbitrarily in the input space (a), then the node closest to each training sample is selected and moved towards it, as well (but with less intensity) as their nearest neighbors (b). Later, after an iterative process, as the mesh tends to the distribution of the data set, the training approaches its end (c).

Since the distances between the input data are evenly distributed on the map, the clusters are not easily detected in this. Hence, in order to detect and display clusters is necessary to calculate the U-matrix (Euclidean distance between weight vectors of neighboring neurons) to visualize the topology of the map features by analyzing the weight vector at each point of the grid, with respect to its neighbors and then display the distance between two neighboring units as height. This leads to a three-dimensional Kohonen map containing a geometric approximation of the distribution of vectors in the network. In this type of arrangement is possible to see valleys in those parts of the grid where the vectors are close to each other, and hills where there are large distances between them, indicating similarities and dissimilarities in the input data respectively.

Because the SOM neurons located at the edge of the grid have different mapping qualities that neurons in the center of the map, it is necessary to ensure that the edge effect is removed from the resulting grid of the SOM after training. With this objective, the resulting grid will be embedded in a sphere to ensure a finite space without borders that minimizes the impact of the smaller neighborhoods of the edges (Ultsch and Mörchen, 2005).

To measure the validity of the clustering achieved by both hierarchical and SOM method, two approaches are used: quantitative and qualitative. From the quantitative point of view, the performance of the clustering process will be measured through the procedure described in Chapter 2, while the qualitative evaluation (only for SOM method) of clustering was made through the visual judging of the similarity matrix (Halkidi et al., 2002a,b).

3.6 Results and discussion

Starting with a squared topology and a random initializing with random seed 0, the number of initial processing units was 8 and finished with 600 processing units. The training rate was 400 with a maximum of 400000 iterations for the training phase, with a pruning rate of 4 units. In the Fig.3.2, is represented the local distance structure resulting from the SOM training through unified distance matrix or U-matrix. As can be seen, after training it is clear the presence of a strong block-diagonal pattern of well-separated clusters. Neurons are positioned in the field of different groups during training the model, so that the distances between adjacent neurons are a good approximation of the distances between the patterns that make up the underlying data. Colorimetrically This is manifested by low values for processing units close together (indicated by green, blue and bright blue) or high values revealing large distance between consecutive processing units (in this case is indicated by yellow, purple and bright red).

It is worth reiterating that to avoid edge effects, this grid was embedded in a sphere, ensuring this way a continuous space without boundaries where the first row and column of the grid was connected to the last row and column respectively. In consequence, the cluster observed in the upper right of the diagonal is the same in the lower left corner (Ultsch and Mörchen, 2005).

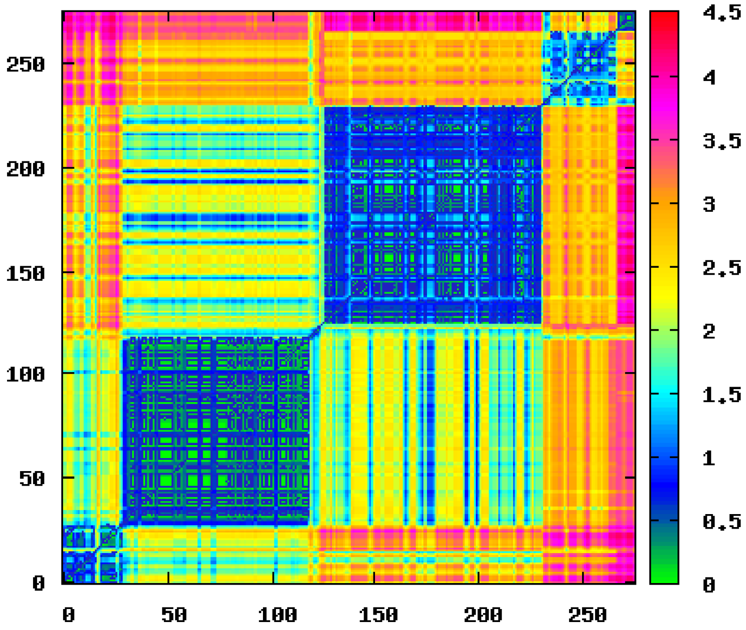


Fig. 3.2. Feature map topology of an unified distance matrix sorted by a Kohonen SOM model fitted to a training sample of 275 farms. The display is coded from higher density areas (bright green) to lower density areas (bright red).

Given that SOM reflects very well the geometric approximations of multidimensional input data in this two-dimensional grid space; because of the ability of these maps to preserve the topology of patterns in the input space. It is easy to interpret in the landscape of Relative distance between meurons in the U-matrix that was generated here, that at least 4 strong clusters appear clearly along the diagonal, indicating the presence of neatly defined structures in the training set. These results resemble those of Tan et al. (2006) when using visual validation to evaluate various methods of clustering.

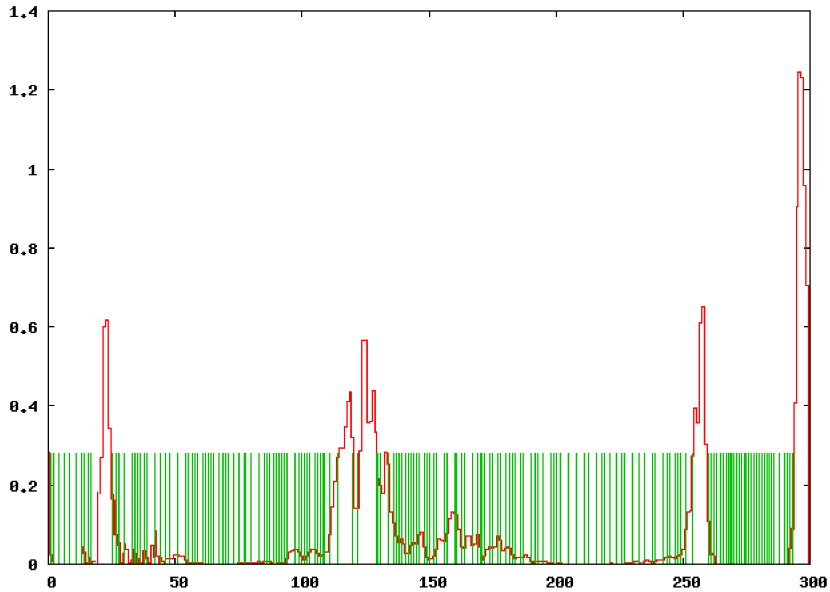


Fig. 3.3. Profile of a unified distance matrix. The display shows the distance between neighboring units as height (red line), giving rise to hills or walls denoting boundaries of partitions created by the weights of the trained SOM. The valleys correspond to those areas populated by farms (green columns) close to each other in the lattice because they share some typical content.

A different perspective of these results can be seen in Fig. 3.3 where a profile of the U-matrix is presented. As can be appreciated, in this cross-section of the U-matrix a red line allows the visualization of the distance between neighboring units as height. In this display are perceived valleys inhabited by vectors of farm's attributes (green columns), or hills, which act as walls that segment the clusters or indicate dissimilarities in the data. Looking at the central part of the same profile of the U-matrix, we can notice that there are also small walls in large valleys, which indicate the presence of small clusters nested within other. These may be considered as outliers or small clusters that can easily be merged with other biger, depending on whether these are groups that tend to be cohesive, not well separated and share some typical content (Ultsch, 2003).

3.6.1 Comparison of hierarchical and self-organizing maps (SOM) clustering performance

The results of comparing three methods of clustering: hierarchical (3 clusters) (see Chapter 2), hierarchical (4 clusters) and SOM (4 clusters) can be seen in Table 3.1. It should be clarified that it was decided to include in the comparison a run with the hierarchical method with 4 partitions, in an attempt to make results comparable between methods, since the SOM method produced 4 clusters.

Table 3.1. Impact of clustering method on cluster quality using hierarchical and self organizing maps approaches.

Method	*%C*	$W\lambda$	*PT*	T^2	*RM*	r^2	*i*	*Slh.*
Hierarchical (3 *clusters*)	92.1	0.11	1.31	3.83	1.94	0.71	0.70	0.50
Hierarchical (4 *clusters*)	88.3	0.15	1.21	3.36	2.30	0.60	0.50	0.32
SOM (4 *clusters*)	94.5	0.09	1.38	3.50	2.43	0.75	0.74	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; *i*: Incidence; *Slh*: Silhouette

As can be appreciated, the performance profiles of the clustering after applying a discriminant analysis to the groups generated by SOM, show that the means of selected variables for each group within this method were different in the population under study given the proximity of Wilks' lambda statistical to zero; and the comparatively high values of the statistical Pillai, Hottelling, and Roy with respect to the two combinations of hierarchical approach. Also, the SOM-based clustering, showed the highest square canonical correlation supporting the idea of well-separated groups with a high proportion of the total variance explained (75 %).

Similarly, in reference to the percentage of classified correct, it appears that this parameter was slightly higher when the grouping was through SOM and hierarchical (3 clusters). This validation is also confirmed after calculating the incidence coefficient. Where this value resulted close to 1 for SOM and hierarchical (3 clusters), and according to the criterion of Tan et al. (2006), this confirms the existence of a high correlation between ideal and actual similarity matrix, pointing out that items belonging to the same group are close to each other. This result is consistent with previous findings of May et al. (2010) who reported that cluster structures showing values within the range from 0.50 to 0.70 are considered reasonably grouped data.

On the other hand, the silhouette coefficient also calculated to validate the performance of this clustering, speaks clearly about the desirable characteristics of the structures found under SOM and hierarchical (3 clusters) aproaches; in the sense that these values indicate that average distance from a given point with respect to its group, is significantly lower as compared to any other group of the structure. This finding agrees with early results of Brun et al. (2007) and Tan et al. (2006), when comparing the performance of different methods of internal validation of clusters. The better performance of the SOM can be attributed to its improved ability to represent the surface of the input space because it can get closer to the original curvature of the data, thanks to self-organization process which inserts units where these are needed (Duda et al., 2001; Bishop, 2006).

At first glance these results reveal that SOM can build a model that is able to generalize properly the structure of input data. However, a closer look can show that the hierarchical method, with three partitions, produced an acceptable model for clustering. Contrary to what happened with the same hierarchical approach, but with four partitions. This particular experiment, running hierarchical with 4 partitions again, was of interest because it had pursued to mach the number of groups used in the hierarchical approach (which is a decision of the investigator) with the number of groups resulting from the self-organization of the network without the influence of the investigator, obtained with SOM. Nevertheless, the performance of this experiment was very poor in terms of the quality of its cluster's quality.

Table 3.2. Confussion matrix for the segmentation achieved by two clustering aproaches: hierarchical and self organizing maps (SOM) trained on 275 cases

		SOM				
		Cluster 1	*Cluster* 2	*Cluster* 3	*Cluster* 4	Σ
Hierarchical (3 *clusters*)	*Cluster* 1	139	16	7	5	167
	Cluster 2	3	34	2	1	40
	Cluster 3	6	5	53	4	68
	Σ	148	55	62	10	275

In order to show the commonalities between the two approaches with better performance (SOM and hierarchical-3 clusters), a confusion matrix is presented in Table 3.2. As can be seen, unlike the hierarchical method, the SOM approach allowed to distinguish four well-defined clusters. Of these four groups, three show significant similarities with the hierarchical method in 92 % of the patterns, equivalent to 226 farms as can be seen on the diagonal of the matrix excluding the cluster 4. The remaining fourth group is composed of 10 farms, whose main use was the leisure and entertainment, so they were discarded from the study. Additionally, 11 farms were excluded because there was no overlap between the two methods and behaved as outliers.

These results are very encouraging, since it had a high proportion of patterns designated in the same groups using two different clustering methods. Therefore, to proceed with the supervised classification in next chapter will be used only those farms and groups where there is overlap between the approaches with better performance (SOM and hierarchical). Consequently, 226 farms will be grouped into three clusters organized as follows: cluster 1 (139 farms); cluster 2 (34 farms) and cluster 3 (53 farms). These groupings will be described in the following section.

3.6.2 Farm clusters summary

Information about typical values and dispersion for each cluster can be found in Table 3.3, where a statistical summary is provided. As can be seen, the statistics are displayed by cluster, where apart from the mean and standard deviation, also include other measures of central tendency and disperssion for the purpose of avoiding the bias associated with the presence of outliers. Based on this information the main characteristics of each type of property described below:

Cluster 1: Comprised of 139 farms devoted mainly to short-cycle crops and to a lesser extent to the livestock with low stocking rate, lack of infrastructure for irrigation and machinery and equipment indices extremely low, with a predominance of regular workforce whose preeminence can be attributed to machinery and equipment lack, based on the features mentioned farms in this group were classified as extensive production systems.

Cluster 2: This group is integrated by 34 farms whose productive activity revolves around livestock and short-cycle crops. By analyzing the descriptive statistics for stocking rates, irrigation infrastructure and machinery and equipment indices in this group and compare them with those of farms included in farms class 1, one can infer that they hold a greater level of intensification that of the latter group, so that for the purposes of this research as a semi extensive production systems.

Cluster 3: The salient features of the 53 farms included in this group are related to the provision of irrigation facilities, machinery and equipment, which have been recognized as indicators of Intensification of agricultural production systems, and also having higher permanent and temporary labor ratios. The conjunction of these features make us assume that the aforementioned farms exhibit a higher degree of intensification, so that, for the purposes of this thesis those farms are label as intensive production systems. Within the farm productive activities that make up the third group includes the short-cycle crops and permanent crops.

Table 3.3. Summary statistics of 226 crop-livestock systems attributes by farm's clusters.

NCM			ACP		PCP		FOP		CIP		STR		MAI		EQI		PSR		TSR	
			Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
1	Mean		44,16	3,455	1,57	,736	,42	,211	1,34	,927	,1940	,06205	,00	,001	,03	,011	,3436	,03109	,20	,073
	95% Confidence Interval for Mean	Lower Bound	37,33		,12		,00		-,49		,0714		,00		,01		,2822		,05	
		Upper Bound	50,99		3,03		,84		3,17		,3166		,00		,05		,4051		,34	
	5% Trimmed Mean		43,51		,11		,00		,00		,0691		,00		,00		,3033		,06	
	Median		33,33		,00		,00		,00		,0000		,00		,00		,2000		,00	
	Variance		1814,202		82,377		6,791		130,743		,585		,000		,018		,147		,819	
	Std. Deviation		42,593		9,076		2,606		11,434		,76505		,014		,135		,38325		,905	
	Minimum		0		0		0		0		,00		0		0		,00		0	
	Maximum		100		100		25		100		8,00		0		1		2,00		10	
	Range		100		100		25		100		8,00		0		1		2,00		10	
	Interquartile Range		98,67		,00		,00		,00		,0000		,00		,00		,4377		,02	
	Skewness		,247	,197	9,075	,197	7,633	,197	8,620	,197	7,754	,197	10,361	,197	5,846	,197	1,574	,197	9,106	,197
	Kurtosis		-1,690	,391	93,894	,391	63,155	,391	73,323	,391	73,489	,391	116,019	,391	36,587	,391	2,407	,391	94,134	,391
2	Mean		17,62	4,584	4,11	2,456	40,74	5,629	12,64	4,785	,6702	,19945	,00	,004	,03	,012	,8281	,45262	,04	,021
	95% Confidence Interval for Mean	Lower Bound	8,38		-,83		29,40		3,00		,2682		,00		,01		-,0841		-,01	
		Upper Bound	26,86		9,06		52,09		22,28		1,0722		,01		,06		1,7403		,08	
	5% Trimmed Mean		14,02		,93		39,73		8,49		,4238		,00		,02		,2898		,01	
	Median		,00		,00		40,00		,00		,0000		,00		,00		,1667		,00	
	Variance		945,395		271,431		1425,897		1030,353		1,790		,001		,007		9,219		,020	
	Std. Deviation		30,747		16,475		37,761		32,099		1,33795		,024		,084		3,03626		,140	
	Minimum		0		0		0		0		,00		0		0		,00		0	
	Maximum		100		100		100		100		6,84		0		0		20,00		1	
	Range		100		100		100		100		6,84		0		0		20,00		1	
	Interquartile Range		27,15		,00		74,50		,00		,8109		,00		,01		,2568		,00	
	Skewness		1,844	,354	5,017	,354	,214	,354	2,367	,354	3,592	,354	6,653	,354	3,060	,354	6,041	,354	4,344	,354
	Kurtosis		2,407	,695	27,424	,695	-1,584	,695	3,979	,695	14,014	,695	44,484	,695	9,419	,695	38,291	,695	18,205	,695
3	Mean		42,02	4,376	14,72	3,284	11,50	3,263	54,83	5,387	,5359	,13327	,13	,031	,25	,047	,6167	,07535	2,58	,653
	95% Confidence Interval for Mean	Lower Bound	33,30		8,18		5,01		44,10		,2705		,07		,16		,4667		1,28	
		Upper Bound	50,73		21,26		18,00		65,56		,8012		,19		,35		,7668		3,88	
	5% Trimmed Mean		41,13		10,91		7,23		55,37		,3719		,09		,19		,5439		1,58	
	Median		28,89		,00		,00		90,00		,0000		,00		,00		,3875		,38	
	Variance		1493,688		841,074		830,287		2263,663		1,385		,073		,173		,443		33,245	
	Std. Deviation		38,648		29,001		28,815		47,578		1,17705		,271		,416		,66546		5,766	
	Minimum		0		0		0		0		,00		0		0		,00		0	
	Maximum		100		100		100		100		4,48		1		2		4,00		40	
	Range		100		100		100		100		4,48		1		2		4,00		40	
	Interquartile Range		80,17		6,85		,00		100,00		,0000		,16		,38		,6356		2,54	
	Skewness		,394	,272	1,915	,272	2,399	,272	-,178	,272	2,040	,272	2,666	,272	2,218	,272	2,407	,272	4,389	,272
	Kurtosis		-1,427	,538	2,410	,538	4,358	,538	-1,927	,538	2,897	,538	7,255	,538	5,389	,538	8,115	,538	23,912	,538

ACP: percentage of annual crop, PCP: percentage of permanent crops, FOP: percentage of forest, CIP: percentage of irrigated crops, STR: stocking rate, MAI: machinery index, EQI: equipment index, PSR:rate of permanent staff, TSR: rate of temporary staff,

3.7 Conclusion

This chapter has addressed the unsupervised classification of farms based on their productive attributes using Kohonen self-organizing maps. The main results from clusters quality validation and its comparation with the hierarchical method used in Chapter 2 showed that clusters obtained with SOM exhibited a higher quality with respect to the clusters obtained by hierarchical methods. It has been demonstrated that although the segmentation achieved by SOM was an improvement over the hierarchical approach in terms of accuracy in classification, explained variance and the degree of cohesion among groups, both SOM and hierarchical method agreed in a high percentage of farms in similar groupings. It is remarkable that clusters in which both methods agreed, showed clearly defined decision regions, with wide separation of their centers of gravity and much more compact sets. These results allow us to overcome the limitations encountered with the application of the hierarchical method, and therefore those clusters where coincided both methods will be used at the stage of supervised classification of multispectral data (Chapter 4).

Chapter 4

A kernel based methodology to identify farm intensification levels in Urdaneta municipality of Aragua state, Venezuela

4.1 Abstract

This chapter deals with a new methodology to induct a relationship between agricultural intensification patterns and farms spectral response. This methodology include the uses of a kernel adatron machine (Friess et al., 1998), census data and remote sensing derived land cover images (Landsat ETM) to model such a relationship. Findings suggests that effective farm intensification detection based on spectral characteristic recorded in a satellite image is possible; and reveals that repeatable links between biophysical and spectral features can be derived from abstractions that are difficult to observe as farms. The accuracy on classification performance shows that the spectral complexity of remote sensed images can be effectively handled without sacrificing the simplicity of linear hypothesis of representation within this methodology.

4.2 Introduction

One principle of agricultural intensification is that there is no single blueprint applicable worldwide to understand this phenomenon. Boserup (1965) proposed a general paradigm that has been acknowledged. Basically this approach states the driving role of demography on changes over farms land uses spatial organization. Such pattern occurs as farm management responds seeking fulfill social and environmental challenges. As a result, intensification processes contributes to the landscape and its attained land cover dynamic due to the resulting influence of their inner land use arrangements. The most basic approach to link land cover dynamic and the study of agricultural intensification is through farm classification; and within this field, the pattern recognition of remotely sensed data has resulted in one of the most effective ways to make land cover data periodically available over large areas in an spatially explicit fashion.

Traditionally the classification of land cover has been addresed through fuzzy clustering (Wang, 1990; Odhiambo et al., 2004); maximum likelihood classifier (MLC) (Strahler, 1980); Bayesian (Forbes and Raftery, 1999; Jeon et al., 2004), and artificial neural networks (ANN) (Chiuderi et al., 1994; Miller et al., 1995; Gleriani et al., 2004). Nevertheless; these methods are highly dependant on data distribution assumptions and solutions of fuzzi logic paradigm, given its probabilistics premises, is not resistant to the bias introduced by users in terms of membership rules. On the other hand, artificial neural networks is plagued of local minima and results are not deterministics. Much of these limitations have been addresed within the foundations of statistical learning theory (Vapnik, 1995), in which pionering application is represented by maxim margin classifiers (Boser et al., 1992); which have been overcome much caveats of traditional methods.

Part of the robustness of this approach is an efficient separation between any two classes by learning algorithms based on the identification of a linear optimal hyperplane that maximises the distance between both informational categories; and addresses non-linearities by mapping input data into a multidimensional feature space induced by a kernel function (Aizerman et al., 1964). The ability of these algorithms to produce an optimal separating hyperplane has been tested on the land cover classification domain showing important levels of classification accuracy (Huang et al., 2002; Zhu and Blumberg, 2002). In the medicine sphere the kernel adatron has shown be an effective mean helping physician at early diseases diagnosis, such as breast cancer (Land and Bryden, 2003), brain human tumor (García and Moreno, 2004) and myocardial infarction (Conforti and Guido, 2005). The algorithm classificatory power has also been proven at scene recognition (Le-Saux and Amato, 2004).

The primary aim of the present study was to evaluate the learning machine approach by using the kernel-adatron algorithm (Friess et al., 1998) to classify farms by intensification levels based on census and remote sensed data. Additionally it is going to be used to investigate whether such an approach will attain comparable cluster accuracy as that achieved with traditionally supervised classification methodology as linear discriminant analysis while using only spectral information. Also, the effect of different kernel functions and their parameters on the accuracy of farm classifications from Landsat 7 ETM+ images will be investigated.

The plan of this chapter is as follows. Firstly a brief overview of kernel methods with emphasis on kernel adatron and kernel principal component algorithm is given. Secondly, a description of the data and experimental design is provided, followed by a brief section of results and discussion and, finally, the conclusions are summarised.

4.3 An overview of learning machines

Support vector machines are a type of machine learning, developed relatively recently based on statistical learning theory introduced by Vapnik and Chervonenkis (1974), and quite successful in resolving basic problems of supervised learning: classification and regression (Shawe-Taylor and Cristianini, 2006). Part of the success is because these are linear machines with an enormous capacity for representation. The solutions are not built into the input space, but in a higher dimensional space, the feature space, where it is possible that a simple linear function is sufficient to solve the problem given the high-order correlations of the data made explicit (Schölkopf et al., 1999).

The input data are taken to the feature space via a nonlinear transformation which brings diversity richness to the expression of the solution. Additionally, the shape of the solution function is such, that the transformation is not directly involved, it is implicit through kernel functions (Aizerman et al., 1964; Aronszajn, 1950), which are simply inner products in feature space represented as functional in the input space, hence the expression of the transformation is not required. To determine whether a function is a kernel, it must verify compliance with the Mercer theorem (Mercer, 1909). All kernel methods in machine learning are based on this principle, although not all have the same objectives.

Another reason why the linear machines produce such good results is because they care about finding the best possible solution to the problem given. This imposition reduces the degrees of freedom, or in some way, bad conditioning in the approach to learning problems. The quality of the solution is measured through a suitable quantitative criterion, which in the case of classification, involves the finding of a decision function whose margin of separation between classes is the maximum possible. In the case of regression, this usually results in minimizing the squared error that is committed with the approximant function, if it is greater than a specified tolerance. With these additional considerations, the problem becomes a constrained optimization: should optimize a high quality functional under the constraints imposed by the underlying problem. In classification, to produce the correct labels, in regression, to produce interpolation values for each set of input data. This type of problem can be solved by the Lagrange formalism. In fact, it can be shown that the class of problems arising in training the learning machine belong to the field of convex quadratic programming, ie, convex quadratic cost functions with linear constraints, where the existence of a unique optimum is guaranteed (Vapnik, 1995).

Finally it should be mentioned that although there is no assurance that any transformation becomes a complex problem in the input space in a more simple in the feature space, it is certainly possible to increase its feasibility in its formulation by allowing greater flexibility of the restrictions. This relaxation, considered important from the standpoint of learning machines, is achieved by introducing a set of lax variables (Bishop, 2006). The lax variavles will allow the acceptance of small exceptions to the satisfaction of constraints. The size of the deviation allowed will be controlled through an additional term of penalty included as part of the cost function. A parameter known as regularization factor is responsible for weighing the penalty term against the term quality of the solution in the optimization functional. The solution obtained will be the best to a pre-established level of compromise.

4.3.1 Kernel functions

Kernel functions have been referred as a key component to the efficient use of high dimensional feature spaces, in which the input data is mapped looking for a maximal separating hyperplane to obtain a linear boundary of data (Cristianini and Shawe-Taylor, 2000). Kernel functions also have played an important role in the automatic discovery of data regularities, by using computer algorithms that allow data classification into categories based on these regularities.

The introduction of kernel functions in the learning machines arena is acknowledged to Aizerman et al. (1964), even though, Shawe-Taylor and Cristianini (2004) refers Aronszajn (1950) as precursor. However, it was with the publication of Boser et al. (1992) that kernel functions become recognised as a powerful method to find out large-margin classifiers, opening the way to support vector machines, which stand out as one of the most powerful developed algorithms to deal with pattern analysis recognition.

Kernel methods as approach to deal with pattern analysis works by embeding the data into a space where the nonlinear pattern can be discovered as linear relations. This transformation is done by applying the so-called kernel function (Shawe-Taylor and Cristianini, 2004), which is a computational shortcut defined as a function κ for all x,z $\in$ X that satisfies

$$\kappa(x,z) = [\phi\ (x),\ \phi\ (z)], \tag{4.1}$$

Where ϕ is a mapping from X to an feature space F

$$\phi : x \mapsto \phi\ (x),\ \in\ F. \tag{4.2}$$

In other words, kernel functions offer a way of computing the inner product $[\phi(x), \phi(z)]$ in feature space directly as a function of the original input points, by mapping x and z to vectors ϕ(x) and ϕ(z) and then taking their inner product, so that it is possible to train a linear machine in the feature space (Cristianini and Shawe-Taylor, 2000).

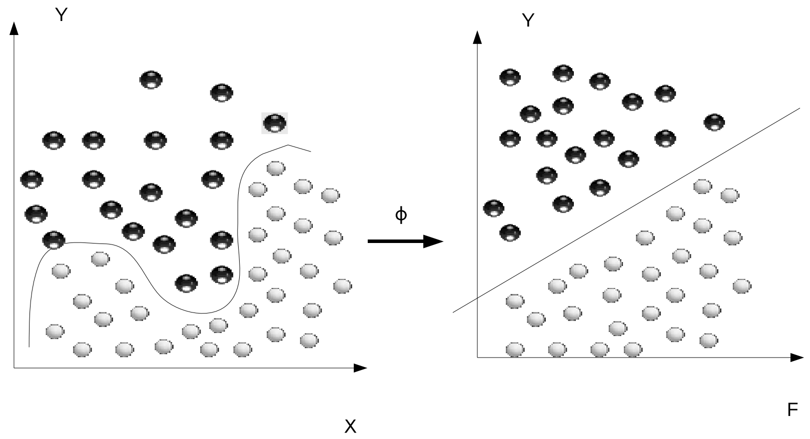


Fig. 4.1. Graphical representation of the classification simplification power of kernel functions to transform nonlinear separable data in input space in linear separable data in the feature space. (Cristianini and Shawe-Taylor, 2000).

The effect of simply computing the inner products between the images of all pairs of data in the feature space as it is done by kernel functions (Fig. 4.1), instead computing the coordinates of the data in that space, has enabled researchers to learn nonlinear relations with a linear machine, and given the kernel firm theoretical foundations, it has also make possible to avoid the typical local minima and incomplete statistical analysis of neural networks and decision trees.

An additional advantage offered by kernel methods is their modularity, allowing the reusability of the learning algorithm, which can work both with any kernel and for any data domain. Since, once "an algorithm procedure is adapted to use only inner products between inputs, it can be combined with a kernel function that calculates the inner product between the images of two inputs in a feature space, making it possible to implement the algorithm in a high dimensional space" (Shawe-Taylor and Cristianini, 2004).

In the context of this research, kernel functions were used as a pattern recognition means by implementing the kernel adatron algorithm to classify farms by intensification levels based on their spectral signature, and also used coupled to principal components analysis to the matrix containig farms spectral response with feature extractions purposes.

4.3.2 Kernel adatron

As a technique belonging the learning machine methodology, the kernel-adatron algorithm (Friess et al., 1998) is capable of learning from examples. It is the result of introducing the kernel function into the adatron algorithm, so that, the data to be classified is projected by a kernel function into a high dimensional feature space where the data can be separate by a maximal margin hyperplane, but the quadratic programming routine usually implemented by support vector machines to find that hyperplane is substituted by an adptation of the adatron algorithm (Analauf and Biehl, 1989), avoiding the expensive computational requirements implicit in quadratic programming routines.

The optimal separating hyperplane pursued in solving a classification problem, has been mathematically expressed by considering a training set $S = \mid (x_1, y_1), \ldots, (x_n, y_n), \mid x_i \in \mathbb{R}^d$ and $y \in \{+1, \text{-}1\}$, where the training vectors are represented by the x_i in the input space of d dimension, and the y_i are the class labels.

The learning machine classification aim is to linearly separate the finite S data set by a decision function $f(\mathrm{x})$ such that:

$$y_1 = f(x_1) \qquad \forall (x_i, y_i) \in S \tag{4.3}$$

To be acceptable $f(x)$, the distance between all trainings point and the separating hyperplane in the feature space must be positive:

$$\gamma_i = y_i f(x_1) \geq 0 \qquad \forall (x_i, y_i) \in S \tag{4.4}$$

This criterion can be extended to the geometrical margins of a linear decision function:

$$\Gamma_i = \frac{\gamma_i}{\|\mathrm{w}\|} = \frac{y(\langle w \cdot x_i \rangle + b)}{\|\mathrm{w}\|} \geq 0 \qquad \forall (x_i, y_i) \in S \tag{4.5}$$

From relation 4.5 it has been stated that the greatest minimum geometrical margin indicates the most stable solution, which is called the maximal margin classifier and also the perceptron with maximal stability.

By assuming the unity as the minimum functional margin, the problem of obtaining the maximal classifier parameters has been stated as:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|w\|^2 \\ &\text{subject to} \quad \gamma_i = y(\langle w \cdot x_i \rangle + b) \geq 1 \qquad \forall (x_i, y_i) \in S \end{aligned} \tag{4.6}$$

The maximal margin classifier can deal with nonlinear generalizations by adding the advantage of the kernel functions, resulting in kernel machine:

$$f(x) = \sum_{i}^{n} \alpha_i y_i K(x_i, x) + b \tag{4.7}$$

Where $K(x_i, x)$ is the kernel function and the α_i are the lagrangian multipliers resulting from the solution of the optimization problem.

The polynomial kernel throughout used in this chapter for a kernel k is defined as: $k(x,z) = p\ (k_1\ (x,z))$, where p (.) is any polynomial with positive coefficients.

Expanding the polynomial kernel k_d using the binomial theorem results in the following polynomial kernel machine:

$$k(x,z) = \sum_{s=0}^{d} \binom{d}{s} R^{d-s} \, (x,z)^{s}. \tag{4.8}$$

The computation followed by the kernel adatron algorithm implemented, expressed in pseudocode are conveyed in Table 4.1.

Table 4.1. Pseudocode for the kernel adatron algorithm

Input	training set S = $\{(x_1,\, y_1), ..., (x_\ell,\, y_\ell\}$
Process	α=0, i= 0, loss = 0
2	repeat
3	for i = 1 : ℓ
4	$\alpha_1 \leftarrow \alpha_1 + (1 - Y_i \sum_{j=1}^{\ell} \alpha_j y_j \kappa\, (x_j, x_i))$
5	if $\alpha_1 < 0$ then $\alpha_1 \leftarrow 0.$
6	end
7	until α unchanged
8	$f(x) = sgn(\sum_{j=1}^{\ell} \alpha_j y_j K\, (x_j, x_i))$
Output	dual variables α, loss and function f

Source: Shawe-Taylor and Cristianini (2004)

4.3.3 Kernel principal component analysis

The kernel methods ability to solve non-linear problems by mapping the data into an usually higher dimensional feature space where the problem can be solved applying linear approaches, has been extended to principal component analysis, given rise to the technique of kernel principal component analysis (Schölkopf et al., 1998). This technique has been recognised as a powerful means to acomplish feature extraction, given its dimensionality reduction capability to extract the smallest set of features able to transmit the essential information contained in the original data (Hastie et al., 2001).

The main difference between kernel PCA and linear PCA is related to the space in which the directions of maximal variance to build the principal components are founded, which in the former case it is represented by feature space, while in the later it is the input space which is used. Consequently, the mathematical and statistical properties of linear PCA does not result modified, so that, the principal components show a descendent order in terms of their variance; the principal components are orthogonal, hence, uncorrelated; and the mean-squared approximation error in representing the observations in the feature space $\mathscr{H}$ by the first principal components is minimal in relation to all possible directions.

Table 4.2. Pseudocode for the kernel PCA algorithm

Input	Data S = $\{x_1, ..., x_\ell\}$, dimension k
process	$K_{ij} = \kappa(x_i, x_j), i, j, = 1, ..., \ell$
	$K - \frac{1}{\ell} jj'K - \frac{1}{y} Kjj' + \frac{1}{\ell^2}(j'Kj)jj'$
	$[\vee, \wedge]$ = eig(K)
	$\alpha^j = \frac{1}{\sqrt{\lambda}} V_j, \ \ j = 1, ..., K.$
	$\widetilde{x_1} = (\sum_{i=1}^{\ell} \alpha_i^j \kappa(x_i, x))^k$
Output	Transformed data = $\widetilde{S} = \widetilde{x_1}, ..., \widetilde{x_\ell}$

Source:Shawe-Taylor and Cristianini (2004)

4.4 Data and methods

To perform the proposed supervised classification, the geo-referenced boundaries of 226 of the 275 farms that made up the sample initially, were delineated in a Landsat 7 (ETM+) image acquired in November of 1999. As it was stated in chapter 3, for purposes of the supervised classification, the sample was made up just by those farms belonging the same group after applying both unsupervised clustering techniques: Ward's hierarchical agglomerative and the Kohonen self-organizing map, so that, each farm instance was labelled as: extensive (class 1, 139 farms); semi-intensive (class 2, 34 farms) or and intensive (class 3, 53 farms) according to one of the three agricultural intensification levels identified in the study area.

The satellite image used in this research was provided by the Instituto Geografico Simón Bolívar, who made the following pre-processing tasks: first, the original format (FTS) was geo-referenced from header files, next the associated projections and parameters (UTM, WSS84) were validated, then a layer stack was created and lastly the image was radiometrically enhanced using lineal function via look-up table.

In collecting the 20 pixels per channel that constitute the spectral farm response a modified von Neumann (1966) vicinity was followed, which having an asterisk ($*$) pattern, offers a greater probability of including within the sample values of the different farm land covers. The collected farm pixels values were concatenated to form a vector of 180 columns, so that, at the end of the pixel collection we have 226 vector with 180 components, because we have a vector for each farm and we took 20 points from each of the nine image channels. These vector were organised in a matrix where an extra column indicating the corresponding farm intensification label was added, resulting in a matrix of order 226 x 181 (Fig. 4.2).

Before proceeding with the classification process a pre-processing by non linear feature extraction was performed using kernel principal component analysis (KPCA) (Schölkopf et al., 1998) in order to reduce the dimensionality of the data, extracting the smallest set of features able to transmit the essential information contained in the original data matrix, consisting of the three principal directions that best separate the classes under study. These three components were retained for training the learning machine using the kernel adatron algorithm (Friess et al., 1998). The training set size was determined experimentally with trials comprising 6, 10, 16, 20, 26, and 30 instances; and the selection of a minimum training set was based on the feed-back from the accuracy achieved by the learning machine.

In the interest of accuracy assesment, a validation data set integrated by unseen samples from the three experimental farm groups was used; and confussion matrices were built to compare actual versus predicted classes by the algorithm. As a result of dealing with three informational groups of farms; the classification task was approached as a multiclassification process; and given the intrinsic binary nature of the kernel adatron algorithm, in which $y \in \{1, \text{-}1\}$, is used to indicate that the input vector belong to a chosen category (Y= 1), or not (Y= -1), the strategy of one against all classes was adopted in order to segment the cases involved in this study. For comparisons purposes the classification was additionally carried out using linear discriminant analysis (LDA) as standard supervised classification method.

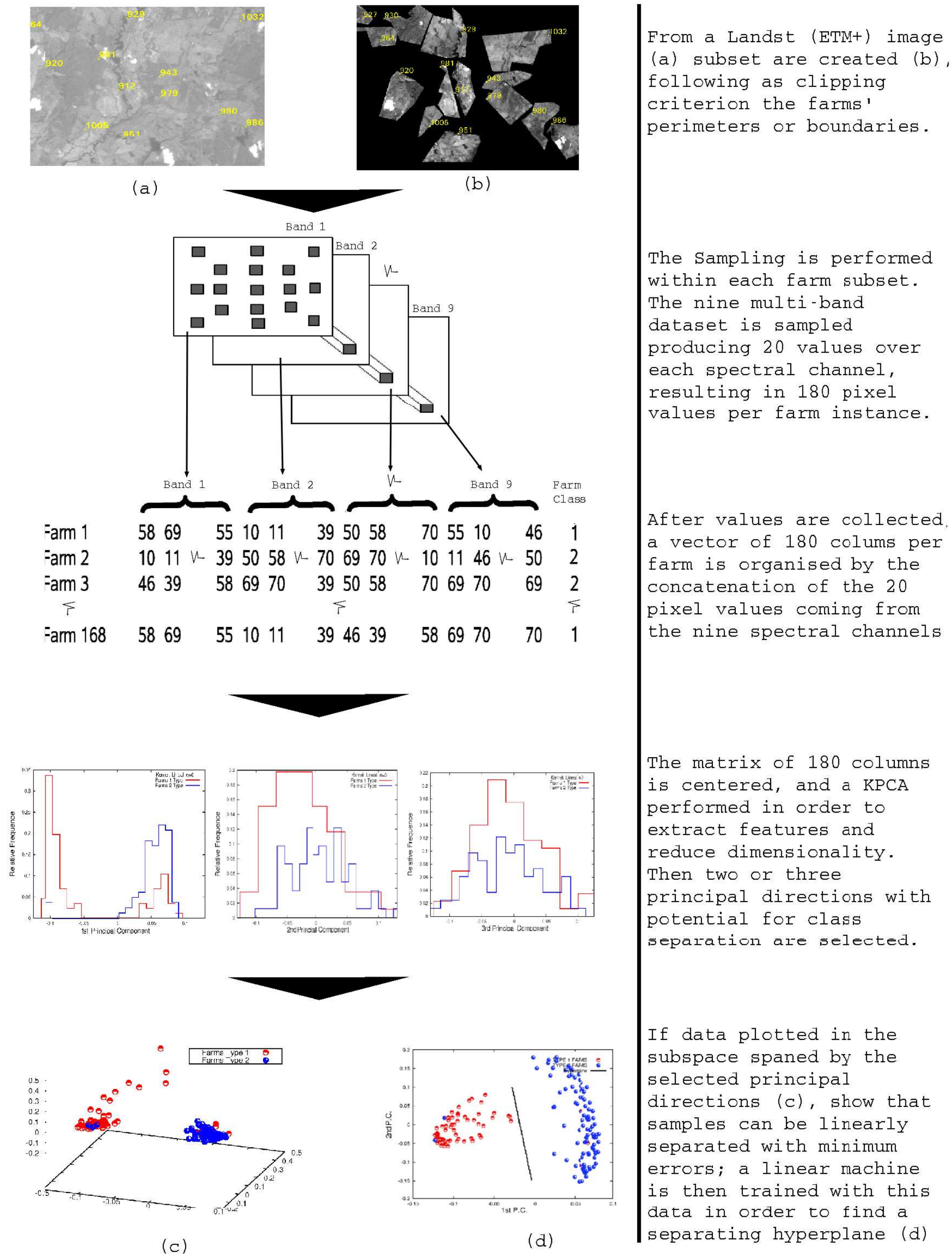


Fig. 4.2. Methodology for the representation of pixel labeling and classification of Landsat images. González (2008).

.

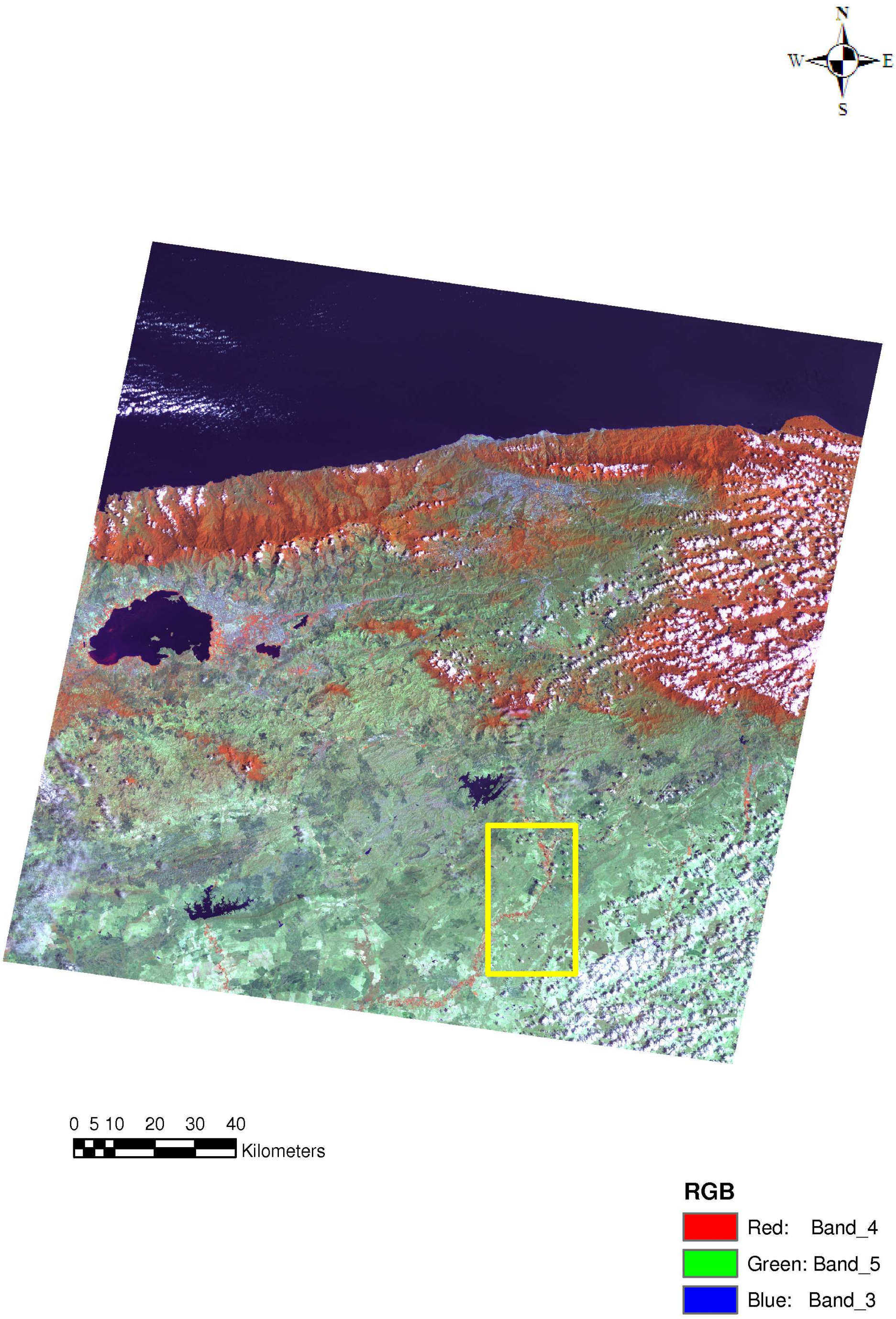


Fig. 4.3. Landsat ETM scene. The yellow box represents the sampling area

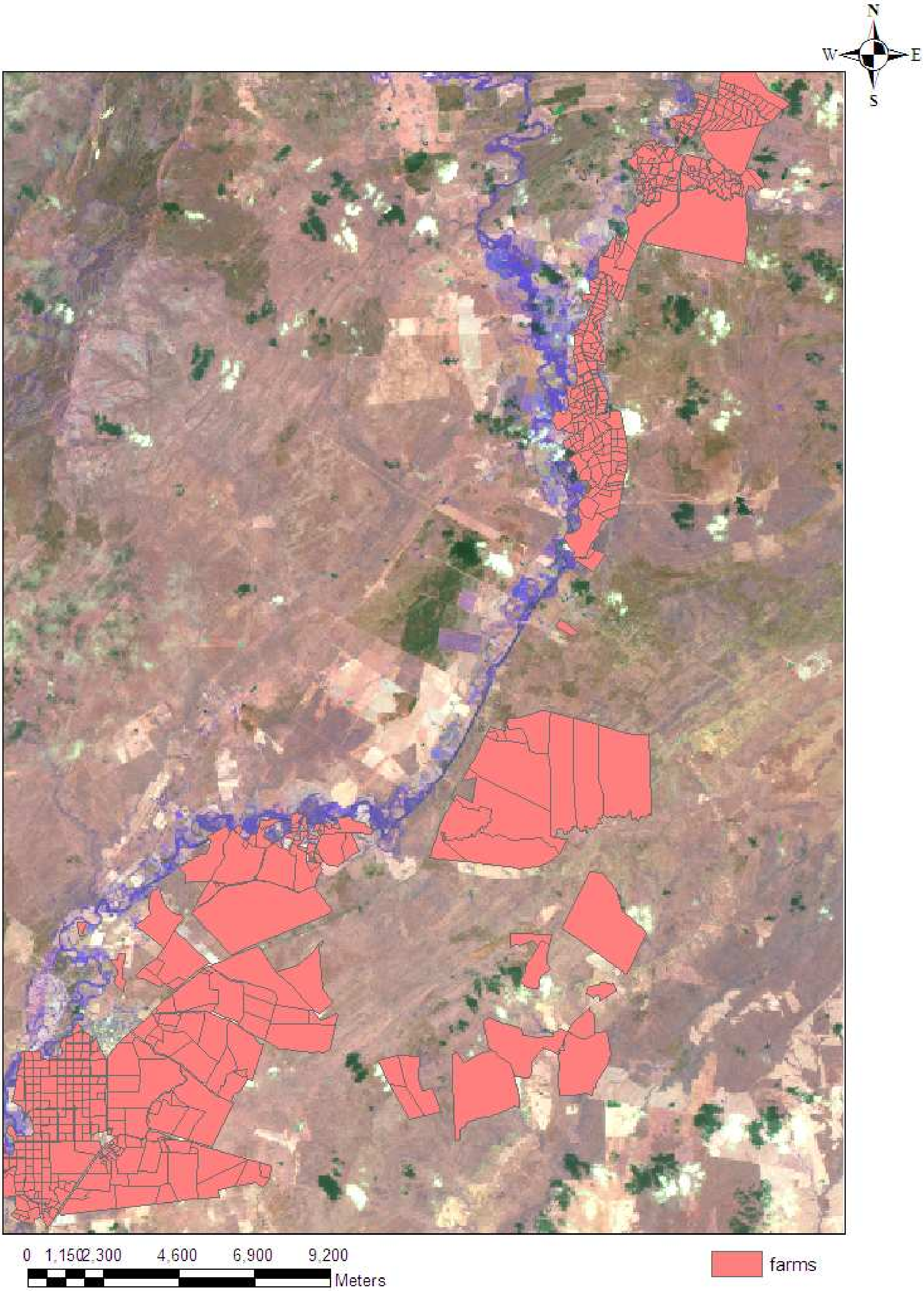


Fig. 4.4. Landsat ETM scene with an overlaping layer of polygons corresponding to the farms used in the linear machine training

4.5 Results and discussion

The kernel principal component analysis procedure carried out to accomplish dimensionality reduction leads to the identification of three principal directions that together best separated the studied classes. Fig. 4.5 depicts histograms by farm categories of the three principal directions (1st, 13th and 23rd) that allowed the best group segmentation. As can be seen for the 13rd principal component, classes appear relatively well separated even though important group overlap is also evident.

Comparatively, the 23rd principal direction showed a better separation between farm categories but there was still some group overlapping among classes 1 and 3. On the other hand, from the perspective spanned by the 13th principal direction, class histograms spread in such a way that projected group centroids appeared almost completely overlapped. However, the space spanned by these three principal directions lead to data projections that maximize the class separation.

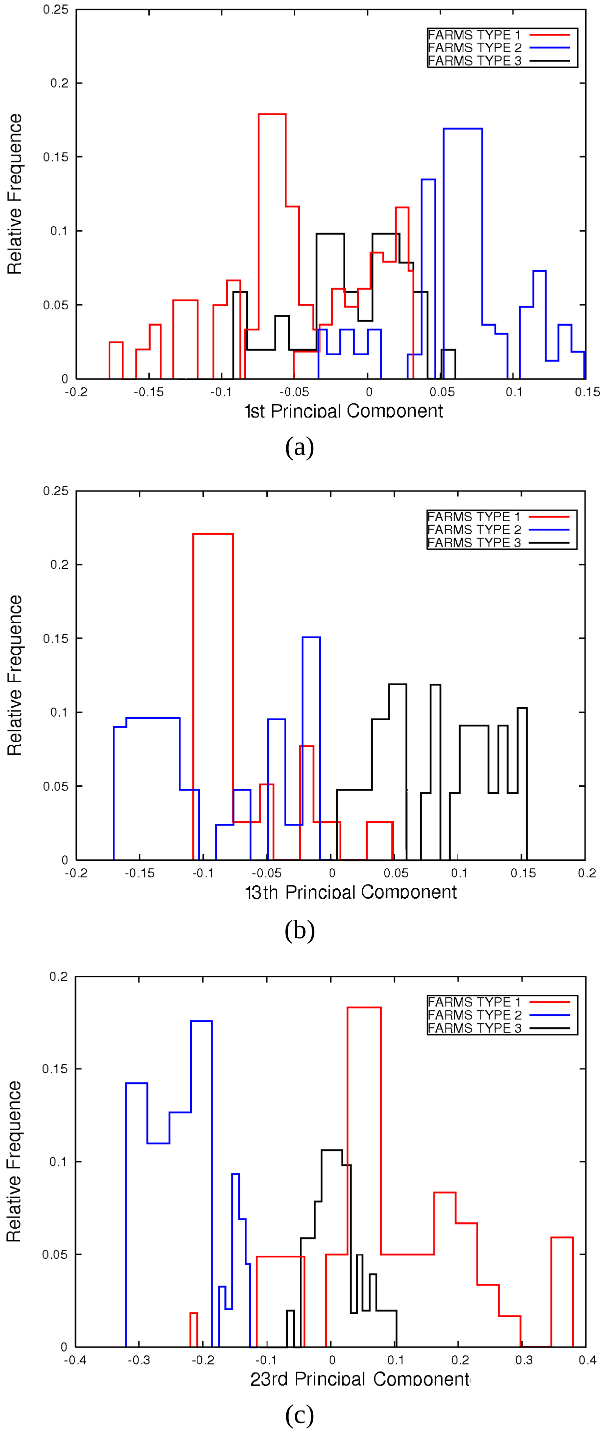


Fig. 4.5. Histograms of the multispectral features for three principal components: 1st(a), 13th(b), and 23rd(c).

Fig. 4.6 illustrates the three dimensional projection of the data onto the referred principal directions. As can be seen, dissected obsevations appear clearly clustered in homogeneous groups showing, with exception of few cases, no overlapping between categories distribution from this perspective. It is also interesting to observe that this class center separation was achieved with a minimum loss of information, in the sense that this projection were involved two non important eigenvectors (13th and 23rd principal component) and in the sense of their relatively low eigenvalues (13th PC: 9.05 and 23rd PC: 2.36) in respect to the 1st PC: 31.25.

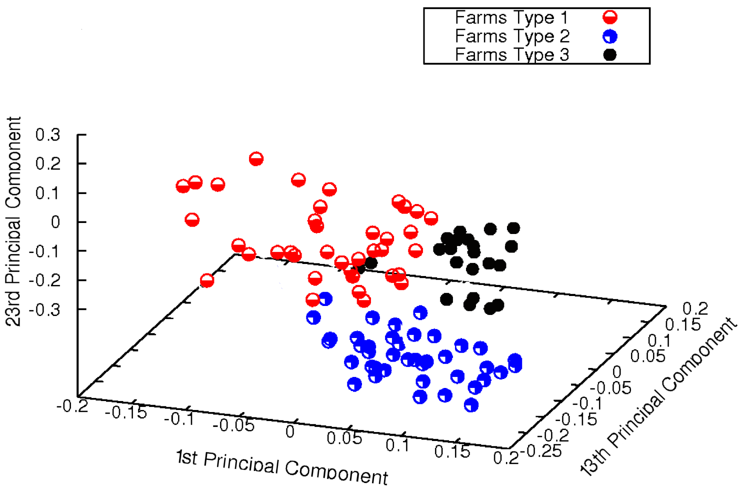


Fig. 4.6. Scatterplot showing the clusters of farms generated by projecting the spectral response on three principal components: 1st, 13th and 23rd after KPCA.

As a consequence, although the classes conditional densities overlap slightly, the three groups occupied different regions of the feature space providing a convenient lower dimensional representation containing the dynamic of the whole input data while optimizing a sum-squared-error criterion. These results appear to replicate what was previously described by Guyon and Elisseeff (2003), who found that variables in the field of feature selection, althoug useless by themselves, may induct improvements in performance when they are suitable combined with others.

On the other hand, the fact that low-variance directions as principal component 23rd also shown good discriminatory power is consistent with the findings of Chang (1983), who observed high-variance components are not always the best separating informational categories. This implies that in order to achieve feature spaces with high discriminative power, low-eigenvalues direction must be explored (Jolliffe, 2002).

In an effort to reduce the required information to perform the training of the learning machine, the impact of the training set size was evaluated. Fig. 4.7 ilustrates the association degree between the innacuracies commited by the classifier (linear and nonlinear versions) per farm class and the size of the training set. As can be observed, the classification accuracy was shown to be highly dependant on the size of the training set, for all informational classes. The most important inaccuracy reductions took place after 10 instances were included in the training process, and stabilized when 20 or more instances were selected.

Due to their general tendency to rely on extreme cases (instances close to decision boundaries) to separate classes, linear machines select the most relevant information from the training set to induct the distribution that generate the data (Foody and Mathur, 2004; Camps-Vals and Bruzzone, 2005). This aspect represents a great benefit for the economy of sampling and explains why only a few training instances were sufficient to permit the convergence of the KA; even though, there were some small differences in the algorithm performance for the separation of each farm class.

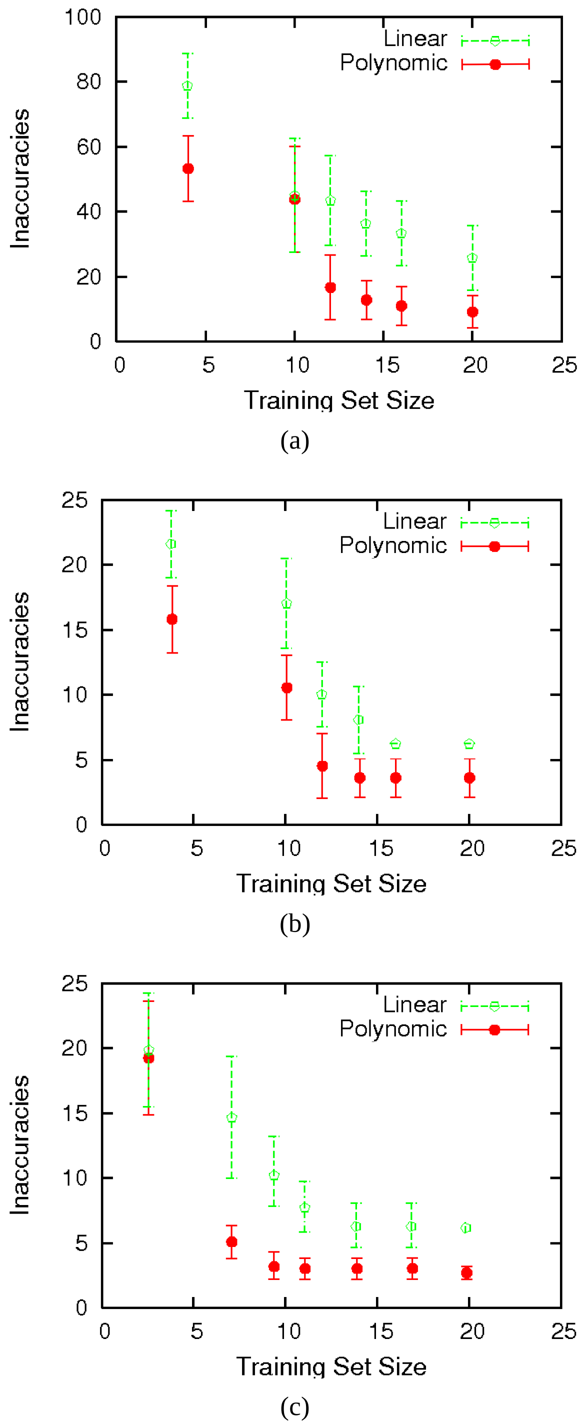


Fig. 4.7. The role of training set cardinality on the classification performance per farm class: 1 (a), 2(b), and 3(c).

For all classes the nonlinear version of the KA algorithm performed much better than the linear one. Average inaccuracies committed under the linear approach were in several orders of magnitude higher compared with polynomic kernel. Moreover, after 10 instances were included in the training process, inaccuracies were reduced by about 10, 72 and 57% for class one, two and three respectively. Then the mean errors become stabilised from 16 samples onwards with 9, 12 and 16% of inaccuracies for the three informational groups. This may indicate that a safe sampling rate to integrate the training set could be around 20 instances, as at this point inaccuracies reached their minimum in all classes, and become more stable with lower standard deviation.

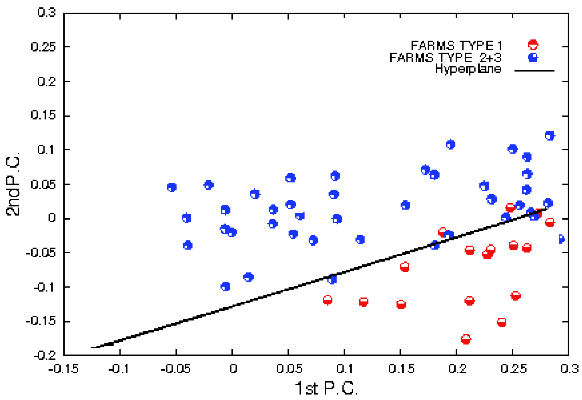
It is worth noting that the standard deviation observed after the maximum rate of sampling was not the same for all groups. This implies that the stability of the algorithm was highly dependant on the type of farm it was trying to separate. Table 4.3 shows some details about the performance of the algorithm on each farm class.

Table 4.3. Error matrix for the supervised classification of three farm classes from multispectral data, using 20 patterns as training set by the kernel adatron (KA) algorithm.

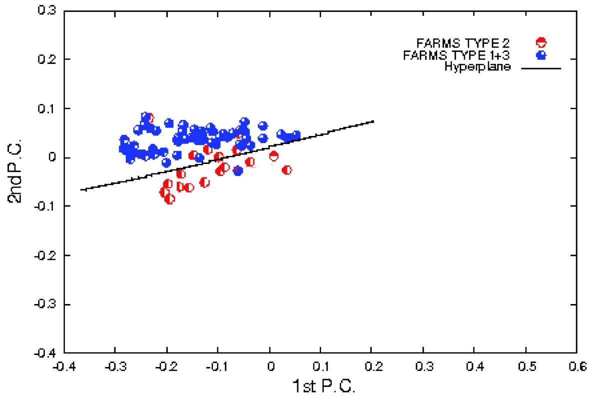
	KA	*Predicted*				
		Class 1	*Class* 2	*Class* 3	Σ	Accuracy (%)
Actual	*Class* 1	127	7	5	139	92.36
	Class 2	2	31	1	34	91.17
	Class 3	3	2	48	53	90.56
	Σ	132	40	54	226	*Overall Accuracy* %
	Accuracy (%)	96.21	77.50	88.88		91.15

As can be seen, once the training set size have been fixed to 20 cases, the general accuracy improves but the algorithm wrongly allocated cases from one class into another; with most of the mistakes occurring between class 1 and 2. One possibility is that this performance arises from the fact that farms belonging to group 2 represent the intermediate stage within the two remaining discrete levels of farm intensification: groups 1 and 3. On the other hand, farm type 1 also included those households with the lowest level of human transformation resembling the typical spectral complexity of natural landscapes; while in farms type 3 most of the inner land cover is the result of management practices that consequently restrict their spectral response to a more narrow electromagnetic space compared with typologies 1 and 2.

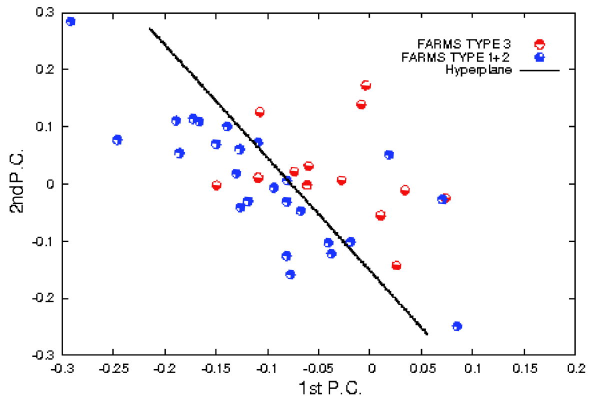
Fig. 4.8 illustrates the decision boundaries for the three informational classes. It is interesting to note that the complexity of the kernel function used to map the data and generate the separating hyperplanes was higher for classes 1 and 2 (order 6) than for farm class 3 (order 3). This may corroborate the above referred complexity on the separation of those farm typologies that resembles natural landscapes as groups 1 and 2. Given the tendency to generate more complex decision boundaries, higher polynomial degree may fit well with those patterns that are difficult to learn; while low order polynomials may be suitable for those situation where less complex classes segmentations are required (Schölkopf and Smola, 2002).



(a)



(b)



(c)

Fig. 4.8. Optimal decision boundaries of a model kernel adatron with polynomic kernel order 3(a);2(b); and 2(c). Using 20 patterns as training set for farm's class 1(a), 2(b) and 3(c).

Finally, this study has demostrated that the use of learning machines might lead to robust solutions that overcome techniques traditionally used in farming system classification such as linear principal component for feature extraction (Kobrich et al., 2003) and linear discriminant analysis for supervised classification (Escobar and Berdegué, 1990). Table 4.4 shows the performance of a supervised classification using discriminant analysis after linear feature extraction with principal component analysis. As can be seen, accuracy between predicted and actual classes were highly discrepant, with an important degree of class understimation for all groups: 36.5, 45, and 52.9 % for class 1, 2, and 3 respectively.

Table 4.4. Confussion matrix for the segmentation of three farm classes trained on 275 cases using linear discriminant analysis (LDA)

	LDA	*Predicted*				
		Class 1	*Class* 2	*Class* 3	Σ	Accuracy (%)
Actual	*Class* 1	106	54	7	139	63.47
	Class 2	10	22	8	34	55
	Class 3	15	21	32	53	47.1
	Σ	131	97	47	226	
	Accuracy (%)	80.9	22.6	68		*Overall Accuracy* % 58.18

This result seems to replicate the findings of Sanchez-Hernandez et al. (2007) who found important levels of omission errors using multispectral data under linear supervised classification. This implies that such approaches are not suitable to deal effectively with classifying complex objects such as collections of land cover labelled as farms, apart from the fact that the minimum amount of training data required might sometimes be very large.

Nevertheless, additional research may be required to test whether this situation applies for all cases and to determine whether other feature extraction methodologies might affect its performance.

4.6 Conclusion

In this research the supervised classification of farming systems using the kernel adatron algorithm has been addressed. Results indicated that solutions to farm class segmentation from multispectral data exist and are robust. Kernel principal component analysis has been demonstrated to be an effective way to reduce multispectral dimensionality and feature extraction with minimum lost of information. It is worth mentioning that the training process was also very cost-effective from a sampling viewpoint, given the reduced training set required to reach good performance. However, in further research, it would be interesting to evaluate the impact of training set size on the performance of the classifier under different sampling strategies and in different geographical location given the relatively low stability showed by the solution explored in this research.

Chapter 5

General discussion

The process of identifying levels of intensification on farms by recognizing patterns in multispectral data, could be divided into a sequence of four operations: unsupervised generation of informational classes into which sensed farm will be assigned (clustering), sensed farms isolation from their surroundings and other farms in the image (segmentation), a feature extractor measures the multispectral properties of the farms that are useful for the classifier (feature extraction), and a classifier uses these features to assign the sensed farm to a category (classification).

According to these terms, both the second and the third chapter of this thesis clearly belong to the stage of clustering. Nevertheless, unlike most clustering methods, which require a priori definition of the number of clusters to use, and therefore, address the problem of unsupervised classification as a sort of subjective decision making (Duda et al., 2001). This paper presents the use of methodologies that can infer the appropriate number of clusters from the topology of the input data, with the advantages inherent to the simplification of the clustering process; and the greater range of applicability, given the generality of starting conditions by not including the researcher's subjective decisions about the number of clusters (Kostov and McErlean, 2006). Additionally, a kind of robustness to noise is derived from the fact that the self-organizing method presented does not interpolate, but approximates the distribution of input data considered, and moreover, show little sensitivity to parametric assumptions that govern most clustering methods.

Within the field of learning and other associated areas such as unsupervised classification, different representations are used to solve the problem of cluster validation. Basically in this investigation the representations used were mainly of two types:

- Cluster cohesion and separation: it is an expression defined in terms of continuous correspondence by the sum of proximities between the objects that make up the cluster (farms) and its centroid. The quality parameters obtained in this research are quite close to similar studies made by Ultsch and Mörchen (2005), where the hierarchical linear model (Chapter 2) produces clusters of low explained variability, and less farms classified correctly when the number of sets was above or below the threshold of three groups of farms, and therefore it is expected that the representations in these extremes were not good enough according to the cohesion criterion indicated in previous studies by Tan et al. (2006) and Brun et al. (2007).

- Techniques based on proximity matrix: this is a discrete description defined as the correlation between actual and ideal similarity matrix. Based on these results, it is clear that the clusters obtained show little similarity between objects located in different groups, and high similarity between objects located within the same group. While seeing some points outside the block diagonal structure, this follows a pattern which shows a clear grouping. This diagonal structure usually shows a diffuse distribution when the similarity tends to be zero within the block (Ultsch, 2003; Tan et al., 2006), which is not the case for this research. In this sense, these results are consistent with those found by Ultsch and Mörchen (2005) and Jain (2010), who compared several methods of clustering against SOM and found that these self-organizing maps can be highly non-linear and topologically correct, ensuring in this way, a proper representation of the neighborhood points that are close in the input space by adjacent points in the two-dimensional space inhabited by the mesh of neurons and weight vectors.

Although in this research, the self-organizing maps exhibits a superior ability to approximate the mapping function on a classification problem for farms when compared with discriminant analysis. Some issues remain outstanding, for example, relatively small training set resulting in suboptimal network architectures, presence of local minima and the stochastic nature of the solutions achieved (Jain et al., 1999). Local minimum and its stochastic nature has shown a significant influence on the use of neural approaches in unsupervised learning, and there is reason to believe that these limitations are present in training sets with similar dimensions to this investigation (Zhang, 2000).

In this study, this effect was minimized by selecting only those farms that integrate the same informational category in both classification systems. Evidence for the importance of combining methods of classification comes from the studies of Dubes (1987); Jain and Dubes (1988b); Kettering (2006); and Duda et al. (2001). Who have conducted studies about evaluation external, internal and relative validity of the clusters

With regard to supervised classification, this paper has described the application of machine learning methods to the classification of farms based on multispectral data. The underlying idea is simple: all information is in the input data, the repeated presentation allows to find existing topological relations. The quality of representation in general improved with increasing the size of the training set, however, even for moderate sampling levels, good models can be built.

The methodologies used are versatile, can be applied to objects whose spectral configuration is more or less arbitrary as farms, and produce results that are either for samples distributed in time, or for joint training which spectral information was collected recently. In this respect the Adatron kernel approach is more powerful than the linear approach because of the ability of the kernel functions to extract significant features. In contrast, linear approaches such as discriminant analysis can not be used to induce multispectral features on farms due to its high variability over time.

Chapter 6

General conclusion

This thesis has delineated the procedure followed to classify farms by intensification levels from landsat imagery. In addressing this problem a kernel adatron algorithm was implemented, which as a method belonging the learning methodology is capable of learning from examples and classify unseen samples. This methodology is particularly useful in cases like the one treated in this research, in which there is no known method for computing the separation of farms by intensification levels from their spectral response.

To address the central aim the kernel adatron algorithm needed to be complemented with an unsupervised classifier that was able to provide the intensification labels for the selected training sample of farms. This task was accomplished with a tradittional agglomerative hierarchical and also by using the self-organizing map which is a relatively recent technique belonging the learning machine approach. Two main conclusions were drawn from the unsupervised classification; firstly, it was confirmed that the capacity of proportion of land in cultivation and under irrigation, stocking rate, machinery, equipment and labour index all represent farm intensification levels. These variables could then be used as replacement of yield (commonly referred as to the best output intensification indicator) in those cases like farms of Urdaneta municipality where data is not available; and secondly, the advantages offered by the Kohonen self-organizing map to identify without any data preprocessing, the number of groups.

The supervised classification revealed the most important conclusion of this research, since it was shown that through the integration of unsupervised and supervised classification techniques it is feasible to segregate farms by intensification levels based on their spectral signature in a landsat image, by using the generalisation capacity of the kernel machine to compute the correct output from the input data, without precisely specifying the method by which this task must be done, but using input/output examples as a training data from which the kernel can extract the decision function required in classifying.

The possibilities offered by the learning methodology to deal with agricultural intensification classification from remoted sensed satellite are huge, given the ability of this methodology to cope with tasks that can not be solved by traditional programming approach. Taking advantages of the remoted sensed satellites in providing synoptic and repetitive imageries over large areas, reduces the need for expensive ground survey which has been one of the main barriers impeding the effectively monitoring of agricultural intensification process and thus by implementing it would give a better understanding of the environmental and socio economic impacts associated with intensification processes and the drivers forces involved.

One of the limitations of this research is related to the level to which the algorithm can be applied, since it was proved that it can work properly if the area to be classified is delimited (farm boundaries from which a sample of pixels is taken). This requirement could be considered as an impedement to conduct national or regional studies, in which the objective could be to identify intensification trends of geographical areas for instance, and then map them. This is just a limitation strongly related to this thesis, however, they are obviously as varied as the objectives to be pursued for different disciplines, so that a closer integration between remote sensing specialist and computer sciences with agricultural economist, agronomist, ecologist and sociologist would be necessary in order to fully exploit the advantages of this technique.

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.

Aldenderfer, M. and Blashfield, R. (1984). *Cluster analysis*. Number 44 in Quantitative applications in the social sciences. Sage University Paper.

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

Andersen, E., Verhoog, A., Elbersen, B., Godeschalk, F., and Koole, B. (2006). A multidimensional farming system typology. Technical Report 12, SEAMLESS. System for Environmental and Agricultural Modelling: Linking European Science and Society.

Angelsen, A. (1999). Agricultural expansion and deforestation: Modelling the impact of population, market forces and property rights. *Journal of Development Economics*, 58:185–218.

Angelsen, A. and Kaimowitz, D. (1999). Rethinking the causes of deforestation: Lessons from economic models. In *The World Bank Research Observer*, volume 14, pages 73–98. The World Bank.

Arias, L. (1993). *La tecnología en la agricultura venezolana: evolución y perspectivas*. Sistema Alimentario Venezolano. Estudios Especiales. Fundación Polar, Caracas, Venezuela.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, (68):337–404.

Aune, J. and Bationo, A. (2008). Agricultural intensification in the sahel-the ladder approach. *Agricultural systems*, 98:119–125.

Baltenweck, I., Staal, S., Ibrahim, M., Manyong, V., Wis, T., nad F. Holmann, M. J., Patil, B., and DeWolff, T. (2003). Broad dimensions of crop-livestock intensification and interaction across three continents. In *Working paper*. International Livestock Research Institute (ILRI), international Institute Centre for Tropical Agriculture (CIAT), International Institute of Tropical Agriculture (IITA), University of Peradeniya and BAIF, Nairobi.

Baltenweck, I., Staal, S., Owango, M., Muriuki, H., Lukuyu, B., Gichungu, G., Kenyanjui, M., Njubi, D., Tanner, J., and Thorpe, W. (1998). Intensification of dairying in the greater Nairobi milk-shed: Spatial and household analysis. In *Collaborative Research Project Report*. Kenya Agricultural Research Institute (KARI), Ministry of Agriculture, Livestock Development and Marketing (MoA), International Livestock Research Institute (ILRI), Nairobi.

BCV (2006). Sistema de cuentas nacionales. Banco Central de Venezuela.

Benedict, M., Tolley, H., Elliot, F., and Taeuber, C. (1944). Need for a new classification of farms. *Journal of Farm Economics*, 26(4):694–708.

Bernhardt, K., Allen, J., and Helmers, G. (1996). Using cluster analysis to classify farms for conventional alternative systems research. *Review of Agricultural Economic*, 18(4):599–611.

Bilsborrow, R. and Carr, D. (2000). *Tradeoffs or synergies? Agricultural intensification, economic development and the environment*, chapter Population, agricultural land use and the environment in developing countries. CABI.

Bishop, C. (2006). *Pattern recognition and machine learning*. Information Science and Statistics. Springer, Singapore.

Blaut, J. (1977). Two views of diffusion. *Annals of the Association of American Geographers*, (67):343–349.

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.

Boserup, E. (1965). *The conditions of agricultural growth: the economics of agrarian change under population pressure*. G. Allen & Unwin, London.

Boyazoglu, J. (1998). Livestock farming as a factor of environmental, social and economic stability with special reference to research. *Livestock Production Science*, 57(1-14).

Brookfield, H. (1972). Intensification and disintensification in Pacific agriculture. *Pacific Viewpoint*, (13):30–48.

Brookfield, H. (2001). Intensification, and alternative approaches to agricultural change. *Asia Pacific Viewpoint*, 42(2/3):181–192.

Brown, P. and Podolefsky, A. (1976). Population density, agricultural intensity, land tenure, and group size in the New Guinea. *Ethnology*, (15):211–238.

Brun, M., Sima, C., Hua, J., Lowey, J., and Carroll, B. (2007). Model-based evaluation of clustering validation measures. *Pattern Recognition*, 40:807–824.

Burney, J., Davis, S., and Lobell, D. (2010). Greenhouse gas mitigation by agricultural intensification. *PNAS*, 107(26):12052–12057.

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.

Caraveli, H. (2000). A comparative analyis on intensification and extensification in Mediterranean agriculture: dilemmas for lfas policy. *Journal of Rural Studies*, (16):231–242.

Carswell, G. (1997). Agricultural intensification and rural sustainable livelihoods. A think piece. In *IDS working paper 64*. IDS, Brighton.

Carswell, G. (2000). Agricultural intensification in Ethiopia and Mali. In *IDS working paper 68*. IDS, Brighton.

Carswell, G. (2002). Farmers and fallowing: agricultural change in Kigezi district, Uganda. *The Geographical Journal*, 168(2):130–140.

Chamberlain, D., Fuller, R., Bunce, R., Duckworth, J., and Shrubb, M. (2000). Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. *The Journal of Applied Ecology*, 37(5):771–788.

Chang, W. (1983). On using principal components before separating a mixture of two multivariate normal distributions. *Applied Statistics*, 32(3):267–275.

Chiuderi, A., Fini, S., and Cappellini, V. (1994). An application of data fusion to landcover classification of remote sensed imagery: a neural network approach. In *Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 756–762. IEEE.

CIRAD (1989). Programa estadístico del servicio informatico. CIRAD, Montpellier, France.

Conelly, W. (1992). Agricultural intensification in a Philippine frontier comumunity: impact on labor efficiency and farm diversity. *Human Ecology*, (20):203–223.

Conforti, D. and Guido, R. (2005). Kernel based support vector machine classifiers for early detection of myocardial infarction. *Optimization Methods and Software*, (2 & 3):401–413.

Conley, D., Paerl, H., and Howart, R. (2009). Controlling eutrophication: nitrogen and phosphorus. *Science*, 323(1014-1015).

Corden, W. and Neary, P. (1982). Booming sector nad de-industrialisation in a small open economy. *The Economic Journal*, 368:825–848.

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

Dasgupta, P. (1995). Population, poverty and the local environment. *Scientific American*, 272(2):40–45.

DaSilva, A., Escobar, M., Colmenares, O., and Martínez, C. (2003). Aplicación de métodos multivariados en la clasificación de unidades de producción con vacunos doble propósito en el norte del estado Carabobo, Venezuela. *Revista Científica FCV-LUZ*, 13(6):471–479.

Decaens, T. and Jiméenez, J. (2002). Earthworm communities under and agricultural intensification gradient in Colombia. *Plan and Soil*, (240):133–143.

Demont, M., Jouve, P., Stessens, J., and Tollens, E. (2007). Boserup versus Malthus revisited: Evolution of farming systems in northern Côte d'ivoire. *Agricultural Systems*, (93):215–228.

DeWilde, J. (1967). Experiences with agricultural development in tropical Africa. International Bank for Reconstruction and Development, Baltimore.

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

Doan, P. (1995). Population density, urban centrality, and agricultural intensification in Jordan. *Population Research and Policy Review*, (14):29–44.

Dubes, R. (1987). How many clusters are best? *Pattern Recognition*, 20(6):645–663.

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.

EEA (2009). European Environment Agency. Core set of indicators N 020. Nutrients in freshwater. Technical report, Copenhagen.

Ehrlich, P. (1968). *The population bomb*. Ballantine Books, New York.

Escobar, G. and Berdegué, J. (1990). *Tipificación de sistemas de produccion agrícola*, chapter Conceptos y metodología para la tipificación de sistemas de finca. La experiencia de RIMISP, pages 13–44. Red Nacional de Metodología de Investigación de Sistemas de Producción RIMISP.

Evenson, R. and Gollin, D. (2003). Assessing the impact of the green revolution, 1960 t0 2000. *Science, New Series*, 300(5620):758–762.

Everitt, B. (1974). *Cluster Analysis*. Heinemann Educational Books Ltd.

Everitt, B. (1993). *Cluster analysis*. Halsted, New York, third edition.

Fernández, L., Florentino, Y., and Rey, J. (1998). Aplicación de un sistema informático integrado para la evaluación de la degradación ambiental en el trópico. INIA CENIAP Memorias. Maracay.

Field, A. (2005). *Discovering statistics using SPSS*. SAGE publications, second edition.

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

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.

Forbes, F. and Raftery, A. (1999). Bayesian morphology: fast unsupervised bayesian image analysis. *Journal of the American Statistical Association*, 94(446):555–568.

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

Galloway, J., Townsend, A., and Erisman, J. (2008). Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science*, 320:889–892.

García, C. and Moreno, J. (2004). *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). Springer-Verlag, Berlin, Germany.

Gleriani, J., da Silva, J., and Neves, J. (2004). Comparative performance of neural network and maximum likelihood for supervised classification of agricultural crops: single date and temporal analysis. In *International Joint Conference Proceedings*, volume 4, pages 2959–2964. International Joint Conference, IEEE.

Gómez, H., Tewolde, A., and Nahed, J. (2002). Análisis de los sistemas ganaderos de doble proposito en el centro de Chiapas México. *Archivos Latinoamericanos de Producción Animal*, 10(3):175–183.

González, A. (2008). *Spatial pattern recognition for crop-livestock systems using multispectral data*. PhD thesis, GeoSciences School, University of Edinburgh, Edinburgh, Scotland.

Grigg, D. (1982). *The dynamics of agricultural change: the historical experience*. Hutchinson, London.

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

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002a). Cluster validity methods: part I. *SIGMOD Record (ACM Special Interest Group on Management of Data*, 31(2):40–45.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002b). Clustering validity checking methods. *SIGMOD Record (ACM Special Interest Group on Management of Data*, 31(3):19–27.

Hardiman, R., Lacey, R., and Yi, Y. M. (1990). Use of cluster analysis for identification and classification of farming systems in Quingyang county, Central North China. *Agricultural Systems*, 33(115-125).

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.

Hayami, Y. and Ruttan, V. (1971). *Agricultural development: an international perspective*. The John Hopkins University Press, Baltimore.

Hazell, P. and Ramasamy, C. (1991). *The green revolution reconsidered: The impact of high yielding rice varieties in South India*. John Hopkins, Baltimore.

Hensall, J. and King, L. (1966). Some structural characteristics of peasant agriculture in Barbados. *Economic Geography*, (42):74–84.

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

Huang, C., Davis, L., and Townshend, J. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749.

IFAD (2007). Rural poverty in the Bolivarian Republic of Venezuela. International Fund for Agricultural Development. Available on the internet.

INE (2007). Síntesis estadística estadal 2007. Instituto Nacional de Estadística.

INE (2009). Estadísticas de comercio exterior. Instituto Nacional de Estadísticas.

INNOVA (2009). Base de Datos Agroalimentaria de Venezuela. INNOVA.

Iraizoz, B., Gorton, M., and Davidova, S. (2007). Segmenting farms for analysing agricultural trajectories: A case study of the Navarra region in Spain. *Agricultural Systems*, (93):143–169.

Jain, A. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, (31):651–666.

Jain, A. and Dubes, R. (1988a). *Algorithms for clustering data*. Prentice Hall.

Jain, A. and Dubes, R. (1988b). *Algorithmsor clustering Data*. Prentice-Hall advance reference series. Prentice-Hall, Inc, Saddle River, NJ.

Jain, A., Murty, M., and Flynn, P. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323.

Jeon, Y., Choi, J., and Kim, J. (2004). *Lecture Notes in Computer Science*, chapter A study on supervised classification of remote sensing satellite image by bayesian algorithm using average fuzzy intracluster distance, pages 597–606.

Jewitt, S. and Baker, K. (2007). The green revolution re-assessed: Insider perspectives on agrarian change in bulandshahr district, western uttar pradeh, india. *Geoforum*, (38):73–89.

Jolliffe, I. (1972). Discarding variables in a principal component analysis. I: Artificial data. *Applied Statistics*, 21(2):160–173.

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

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

Kaufman, L. and Rousseeuw, P. (1990). *Finding groups in data: an introduction to cluster analysis*. Wiley Series in Probability Statistics. John Wiley and Sons, New York.

Kerr, J. and Cihlar, J. (2003). Land use and cover with intensity of agriculture for Canada from satellite and census data. *Global Ecology and Biogeography*, (12):161–172.

Kettering, J. (2006). The practice of cluster analysis. *Journal of Classification*, (23):3–30.

Kobrich, C., Rehmanb, 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):141–157.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.

Kohonen, T. (1995). *Sel-Organizing maps*. Springer.

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 Systems*, 2(1):33–45.

Lambin, E., Rounsevell, M., and Geist, H. (2000). Are agricultural land-use models able to predict changes in land use intensity? *Agriculture, Ecosystems and Environment*, (82):321–331.

Land, W. and Bryden, M. (2003). Kernel adatron implementation for breast cancer data. In *International workshop on soft computing in industrial applications*. IEEE.

Le-Saux, B. and Amato, G. (2004). Image recognition for digital libraries. In *International Multimedia Conference. Proceedings of the 6th ACM SIGMM International workshop on Multimedia information retrieval*, pages 91–98.

Lele, U. and Stone, W. (1989). *Population pressure, the environmente and agricultural intensification. Variations on the Boserup hypothesis*. The World Bank, Washington.

Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, pages 4–22.

Lipton, M. (1989). Responses to rural population growth: Malthus and the moderns. Population and development review. In *Rural development and populations: Institutions and policy*, volume 15, pages 215–242.

Lozano, Z., Lobo, D., Ildelfonso, L., and Pla, L. (2002). Susceptibilidad a la degradación física de los anfisoles de los llanos centrales y occidentales de Venezuela. INIA. Docuemto de trabajo. Maracay.

MAC (1997). Estructura de la base de datos del sexto censo agrícola. Ministerio de Agricultura y Cria.

MAC (1998). Boletin informativo de resultados preliminares. Ministerio de Agricultura y Cria. Dirección General Sectorial de Planificación y Políticas.

MacLeod, C. and Moller, H. (2006). Intensification and diversification of New Zealand agriculture since 1960: An evaluation of current indicators of land use change. *Agriculture, Ecosystems and Environment*, 115:201–218.

Makhura, M., Goode, F., and Coetzee, G. (1998). A cluster analysis of commercialisation of farmers in developing rural areas of South Africa. *Development Southern Africa*, 15(3).

Malthus, T. (1783). *An essay on the principle of population*. J. Johnson.

MARNR (1983). Sistemas ambientales Venezolanos. In *Region Natural 24. Llanos ondulados centrales*, number 24 in II. Ministerio del Ambiente y de los Recursos Naturales Renovables, Caracas.

Matson, P. A., Parton, W. J., Power, A. G., and Swift, M. J. (1997). Agricultural intensification and ecosystem properties. *Science*, 277:504–509.

May, R., Maier, H., and Dandy, G. (2010). Data splitting for artificial neural networks using som-based sampling. *Neural Networks*, 23:283–294.

McAlpine, J. and Freyne, D. (2001). Land use change and intensification in Papua New Guinea. *Asia Pacific Viewpoint*, (42):209–218.

McIntire, J., Bouzart, D., and Pingali, P. (1992). *Crop-livestock interactions in Sub-Saharan Africa*. Washington.

Meadows, D. (1972). *The limits to growth*. Earth Island Ltd, London.

Meertens, H., Fresco, L., and Stoop, W. (1996). Farming systems dynamics: impact of increasing population density and the availability of land resources on changes in agriculturalsystems: the case of Sukumaland, Tanzania. *Agriculture, Ecosystems and Environment*, 56(3):203–215.

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

Milán, M., Bartolomé, J., Quintanilla, R., Garcá-Cachń, 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):197–199.

Miller, D., Kaminsky, E., and Rana, S. (1995). Neural network classification of remote-sensing data. *Computers & Geosciences*, 21(3):377–386.

Montilla, J., Marin, D., and Briceno, M. (2004). *Agricultura: base del progreso*. Imprenta Nacional.

Moseley, W. (2000). Paradoxical constraints to agricultural intensification in Malawi: the interplay between labor, land and policy. In *Discussion Paper Series*. The University of Georgia.

Munton, R. and Norris, J. (1969). The analysis of farm organisation: An approach to the classification of agricultural land in Britain. *Geografiska Annaler. Series B, Human Geography*, 51(2):95–103.

Netting, R. (1993). *Smallholders, householders: farm families and the ecology of intensive, sustainable agriculture*. Stanford University Press, Stanford.

Niazi, T. (2004). Rural poverty and the green revolution: The lessons from Pakistan. *Journal of Peasant Studies*, 31(2):242–260.

Odhiambo, L., Freeland, R., Yoder, R., and Hines, J. (2004). Investigation of a fuzzy-neural network application in classification of soils using ground penetrating radar imagery. *Applied Engineering in Agriculture*, 20(1):109–117.

Olmstead, A. and Rhode, P. (1993). Induced innovation in American agriculture: a reconsideration. *The Journal of Political Economy*, (101):100–118.

Paarlberg, R. (2009). The ethics of modern agriculture. *Society*, 46:4–8.

Padoch, C. (1985). Labor efficiency and intensity of land use in rice production. an example from Kalimantan. *Human Ecology*, (13):271–289.

PDVSA (2007). Informe de gestión anual. Petróleos de Venezuela S.A. y sus Filiales.

PDVSA (2010). Ministerio del Poder Popular para la Energía y Petróleo. Gaceta Oficial de la República Bolivariana de Venezuela. Caracas, 17 de Marzo de 2010.

Picazo, A. and Hernandez, F. (1993). Tipologias agrarias Valencianas. *Revista de Estudios Agro-Sociales*, (164):74–91.

Pichon, F. (1996). The forest conversion process: a discussion of the sustainability of predominant land uses associated with frontier expansion in the Amazon. *Agriculture and Human Values*, 13(1):32–51.

Pingali, P., Bigot, Y., and Binswanger, H. (1987). *Agricultural mechanization and the evolution of farming systems in Sub-Saharan Africa*. Johns Hopkins University Press for the World Bank, Washington.

Pla, I. (1990). La degradación de los suelos y el desarrollo agrícola de Venezuela. *Agronomía Tropical*, 40(1-3):7–27.

Rodríguez, D., Carnero, J., and Navarro, L. (2003). Nuevos enfoques en el manejo de sabanas en los llanos orientales Venezolanos. Centro de Investigaciones Agropecuarias del estado Anzoátegui.

Rodríguez, J. (1997). Proceso de ajuste y seguridad alimentaria: El caso Venezolano. Universidad Central de Venezuela. Facultad de Agronomía Instituto de Economía Agrícola y Ciencias Sociales.

Rosenberg, A. and Turvey, C. (1991). Identifying management profiles of Ontario swine producers through cluster analysis. *Review of Agricultural Economics*, 13(2):201–213.

Ruthenberg, H. (1980). *Farming systems in the tropics*. Claredon Press. Oxford University Press, New York.

Sampath, R. (1992). Farm size and land use intensity in Indian agriculture. *Oxford Economics Papers*, 44(3):494–501.

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.

Saunders, W. and Webster, D. (1987). *Social archaeology: beyond subsistence and dating*, chapter Unilinealism, multilinealism and the evolution of complex societies. Academic Press, New York.

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. (1983). Classification of milk production systems in Tropical South America: A first approximation. *Tropical Animal Production*, (8):99–110.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge University Press.

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

Shively, G. and Pagiola, S. (2004). Agricultural intensification, local labor markets, and deforestation in the Phillippines. *Environment and Development Economics*, (9):241–266.

Shriar, A. (2000). Agricultural intensification and its measurements in frontier regions. *Agroforestry Systems*, (49):301–318.

Smith, P., Martino, D., and Cai, Z. (2008). Greenhouse gas mitigation in agriculture. *Philosphical transactions of the royal society*, (363):789–813.

Smith, R., Jennings, N., and Harris, S. (2005). A quantitative analysis of the abundance and demography of European hares lepus europaeus in relation to habitat type, intensity of agriculture and climate. *Mammal Review*, 35(1):1–24.

Stone, G. (2001). Agricultural change theory. *International Encyclopedia of the Social and Behavioral Sciences*, pages 329–333.

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

Tachibana, T. (2001). Agricultural intensification versus extensification: A case study of deforestation in the Northern-hill region of Vietnam. *Journal of Environmental Economics and Management*, (41):44–69.

Tan, P., Steinbach, M., and Kumar, V. (2006). *Introduction to data mining*. Addison-Wesley, USA.

Tappan, G. and McGahuey, M. (2007). Tracking environmental dynamics and agricultural intensification in southern Mali. *Agricultural Systems*, 94(1):38–51.

Thapa, G. and Rasul, G. (2005). Patterns and determinats of agricultural systems in the Chittagong hill tracts of Bangladesh. *Agricultural Systems*, 84(3):255–277.

Tiffen, M., Mortimore, M., and Gichuki, F. (1994). *More people, less erosion. environmental recovery in Kenya*. John Wiley & Sons, Chichester, England.

Turner, B. and Doolittle, W. (1978). The concept and measure of agricultural intensity. *The Professional Geographer*, XXX(3):297–301.

Turner, B., Hanham, R., and Portararo, A. (1977). Population pressure and agricultural intensity. *Annals of the Association of American Geographers*, 67(384-396).

Turner, D., Koerper, G., Gucinski, H., Peterson, C., and Dixon, R. (1993). Monitoring global change: comparison of forest cover estimates using remote sensing and inventory approaches. *Environmental Monitoring and Assessment*, 26(2-3):295–305.

Ultsch, A. (2003). Maps for the visualization of high dimensional data spaces. In *WSOM'03*, Japan.

Ultsch, A. and Mörchen, F. (2005). ESOM-Maps: tools for clustering, visualization, and classification with emergent SOM. Technical Report D-35032, Data Bionics Research Group, University of Marburg, Marburg, Germany.

United Nations (2007). World population prospects. The 2006 revision. New York.

Urdaneta, F., Materan, M., Pena, M., and Casanova, A. (2004). Tipificación tecnológica del sistema de producción con ganaderia bovina doble proposito (Bos Taurus x Bos Indicus). *Revista Científica FCV-LUZ*, 14(3):254–269.

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

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

von Neumann, J. (1966). *Theory of self-reproducing automata*. Illinoins Press.

Wang, F. (1990). Fuzzy supervised classification of remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 28:194–201.

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

Wise, M., Calvin, K., Thomson, A., Bond-lamberty, L. C. N. B., Sands, R., Smith, S., Janetos, A., and Edmonds, J. (2009). Implications of limiting co2 concentrations for land use and energy. *Science*, 324:1183–1186.

Zhang, G. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 30(4):451–462.

Zhao, Y. and Karypis, G. (2004). Empirical and theoretical comparisons of selected criterion functions for document clustering. *Machine Learning*, 55(3):311–331.

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.