TROPICAL AGRICULTURAL RESEARCH AND HIGHER EDUCATION CENTER
EDUCATION PROGRAM FOR DEVELOPMENT AND CONSERVATION
GRADUATE SCHOOL

Estimation and Use of Modified Prior Probabilities for Digital Classification Improvement of Tropical Forests

A thesis submitted to the Tropical Agricultural Research and Higher Education Center in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

by
Lucio Pedroni

Turrialba, Costa Rica
July 14, 2000

DISSERTATION ABSTRACT

This dissertation addresses the problem of tropical forest classification using remotely sensed data. Traditional remote sensing methods have had problems for discriminating tropical secondary and disturbed forests. As a consequence, important information is lacking, that is required for research in biogeography and for a complete assessment of carbon dioxide flows from land-use and land-cover change.

To improve the discrimination of tropical secondary and logged forests using remotely sensed data, a Bayesian classification approach was investigated. The prior probabilities were modified as a function of the pixel's geographical context, which is a non-parametric strategy to incorporate information obtained from ancillary data into the maximum likelihood classification. The method has been proposed before, but found little application, because it presented practical problems for obtaining prior probability estimates.

The dissertation describes and tests a data analysis procedure that generates prior probability estimates from class frequencies modeled with ancillary data and a Mahalanobis Distance threshold of previously classified pixels. The method produces a pixel sample size that is large enough to estimate class prior probabilities in numerous geographic strata, which is particularly desirable for the study of large and complex landscapes, in which stratified random sampling for obtaining class frequency estimates is economically prohibitive.

An experiment is presented, in which the procedure made it possible to estimate 537 sets of prior probabilities for an entire Landsat TM scene of central Costa Rica. After modifying the prior probabilities, the overall classification consistency of the training sites improved from 74.6% (traditional equal priors maximum likelihood classification) to 91.9%, while the overall classification accuracy of sites controlled in the field by independent studies improved from 68.7% to 89.0%. The classification accuracy was most improved for the spectrally similar forest categories.

The usefulness of spectral enhancement using the Normalized Difference Vegetation Index (NDVI) and Tasseled Cap features were also investigated. The results of spectral analysis and of 18 classification experiments using different band and index combination are presented. Weak evidence was found to support the hypothesis that spectral enhancement might help the discrimination of tropical secondary, logged, and undisturbed forest categories.

Keywords: Remote Sensing, Geographical Information System, Landsat, NDVI, Tasseled Cap, Prior Probability, Bayes's Classification, Tropical Forest, Secondary Succession, Costa Rica

This thesis has been accepted in its present form by CATIE's Graduate School, and has been approved by the Advisory Committee of the student as partial fulfillment of the requirements for the degree of:

DOCTOR OF PHILOSOPHY

SIGNATURES:

Markku Kanninen, PhD. (CATIE)
Principal Advisor

Jeffrey Jones, PhD. (CATIE)
Member of the Advisory Commitee

Milton Joseph Trlica, PhD. (Colorado State University)
Member of the Advisory Committee

Denis Joseph Dean, PhD. (Colorado State University)
Member of the Advisory Committee

Gilberto Páez, PhD. (CATIE)
Director and Dean of the Graduate School

DEDICATION

This dissertation is dedicated to my mother, Renate Pedroni Koch, who never failed to support my education; and to my wife, Marisol Morera Jiménez, who with great patience joined with me in all my struggles and efforts.

AKNOWLEDGEMENTS

This research was supported by two partial grants awarded by the Swiss Organization for Development and Cooperation (INTERCOOPERATION) and the Tropical Agricultural Research and Higher Education Center (CATIE).

Several persons and institutions gave important contributions to this study. Among them, a special thank is due (in alphabetical order) to:

Mr. Martin Artavia (CATIE), for providing field-work assistance and essential botanical knowledge;

Mr. José Joaquín Campos (CATIE), for facilitating four-wheel transport;

Mr. Denis Joseph Dean (Colorado State University), for introducing me in ARC/INFO GRID and AML, and for critical review of my papers;

The Foundation for the Development of the Central Volcanic Mountain Range (FUNDECOR) for providing Landsat TM data and independent ground data;

Mr. Manuel Guariguata (CATIE), for critical review of one of my manuscripts;

Mrs. Eileen Helmer (Oregon State University), for providing literature and independent ground-truth data;

Mr. Jeffrey Jones (CATIE), for providing computing facilities and work place at CATIE and reviewing my manuscripts;

Mr. Armond Joyce (NASA), for reviewing one of my articles and encouraging me to publish;

Mr. Markku Kanninen (CATIE), for providing literature and reviewing my manuscripts;

Mrs. Paula Mello (CATIE), for editorial reviewing of my manuscripts;

Mrs. Florencia Montagnini (CATIE), for reviewing one of my articles;

The National Geographic Institute (IGN) of Costa Rica for providing digital cartographic data;

Mr. Jonny Pérez (CATIE), for support in SAS analysis;

Mr. Alexander Salas (CATIE) for assistance in the edition of the digital elevation data;

Mr. Mathias Tobler (Swiss Federal Institute of Technology), for VisualBasic programming;

Mr. Milton Joseph Trlica (Colorado State University), for guidance in preparing the research proposal, teaching of scientific style writing, and manuscript review;

Mr. Sergio Velásquez (CATIE), for all kind of assistance in CATIE's GIS laboratory, supply of literature, and manuscript review;

Mr. Geoffrey Venegas (CATIE) for precious assistance during fieldwork.

BIOGRAPHY

Lucio Pedroni was born in Basel (Switzerland) on February 22th, 1962. He spent his childhood in the Italian speaking region of Switzerland (Ticino), where he concluded his elementary and high school. In November 1986, he obtained a Diploma in Forest Engineering at the Swiss Federal Institute of Technology in Zürich.

His professional career includes prolonged stays in Europe (Switzerland), Africa (Rwanda and Madagascar), Latin America (Costa Rica), and North America (United States). As a student in forest engineering he held positions in the Swiss Forest Service, and as a graduate professional he worked in three international bilateral projects of the Swiss Government and the Swiss Organization for Development and Cooperation (INTERCOOPERATION), leading two of them.

His current professional interests include remote sensing and GIS applications to tropical forests and sustainable development, global change research, international trade and valorization of ecosystem services, and forest management certification.

This thesis consists of a summary part and two substudies which are referred to by roman numerals in the summary part.

Part I: Dissertation summary

Part II: Pedroni, L., Improved classification of Landsat TM data using modified prior probabilities in large and complex landscapes.
(Submitted for publication to International Journal of Remote Sensing)

Part III: Pedroni L., Secondary and logged tropical forests as detected with spectrally enhanced Landsat TM data an modified prior probabilites.
(Submitted for publication to Remote Sensing of Environment)

Part IV: Appendix

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ANOVA	Analysis of Variance
AVHRR	Advanced Very High Resolution Radiometer
BHD	Breast Height Diameter
CATIE	Tropical Agricultural Research and Higher Education Center
ECHO	Extraction and Classification of Homogeneous Objects
FUNDECOR	Foundation for the Development of the Central Volcanic Mountain Range
GHG	Green House Gas
GIS	Geographical Information System
GPS	Global Positioning System
DN	Digital Number
IGN	National Geographic Institute of Costa Rica
INTERCOOPERATION,	Swiss Organization for the Development and Cooperation
IR	Infra Red
ISODATA	Iterative Self Organizing Data Analysis Technique
MANOVA	Multivariate Analysis of Variance
MSS	Multi-Spectral Scanner
m.a.s.l	Meters Above Sea Level
MD	Mahalanobis Distance
NDVI	Normalized Difference Vegetation Index
NGO	Non Governmental Organization
PS	Primary Succession
PTWFTB	Premontane Tropical Wet Forest Transition to Basal
SS	Secondary Succession
TM	Thematic Mapper
TWF	Tropical Wet Forest
UCR	University of Costa Rica

Part I

Dissertation Summary

CONTENTS

1. INTRODUCTION

1.1 Role of remote sensing in global change research and policy

Remotely sensed data are fundamental sources of information for our understanding of global change. Past remote sensing research led to increased attention about the problem of tropical deforestation (Woodwell *et al.*, 1987; Myers, 1988; Sader and Joyce, 1988; Westman *et al.*, 1989; Gilruth *et al.*, 1990; Green and Sussmann, 1990; Hall *et al.*, 1991; Campbell and Bowder, 1992; Grainger, 1993; Downton, 1995; Sohn *et al.*, 1999). Today, our knowledge about land cover dynamics, surface resources, biological diversity, and the global climate system has increased considerably in part because of the multiple types of data that have been made available through remote sensing. Once processed and analyzed, remotely sensed data are particularly useful for computer modeling of patterns and processes in the environment, at local (Helmer, 1999; McCracken *et al.*, 1999), regional (Sader and Joyce, 1988; Iverson *et al.*, 1993), and global scale (Brown *et al.*, 1993). As an example, data on land cover dynamics were indispensable for the estimation of pools and fluxes of carbon from terrestrial ecosystems (Table 1). They allowed to conclude that about 23.6% (1.7±0.8 Gt C/yr) of human induced global carbon dioxide (CO_2) emissions during the decade 1980-89, and 20.3% (1.6±0.8 Gt C/yr) during the period 1989-98 were originated from land-use and land-cover change (IPCC, 2000).

Table 1. Pools, sources, and sinks of carbon in terrestrial ecosystems

	1980			Changes 1850 - 1980		
	Area (Mha)	Vegetation (Gt C)	Soil (Gt C)	Area (Mha)	Vegetation (Gt C)	Soil (Gt C)
Tropical Forests	2167 1755[a,b]	288 212[a,b]	203 216[a,b]	-508	-59	-42
Temperate Forests	1492 1038[a,b]	127 59[a,b]	155 100[a,b]	-91	-26	-17
Boreal Forests	1167 1372[a,b]	96 88[a,b]	237 471[a,b]	-4	-6	-3
Non-Forest Ecosystems	8900 10950[b]	73 107[b]	845 1224[b]	+603	+1	+31
Total Terrestrial Ecosystems	1327 15115[b]	583 466[b]	1440 2011[b]	0	-90	-31

Sources: upper values: Houghton, 1996; lower values: a) Dixon *et al.*, 1994 and b) IPCC, 2000.

However, the uncertainty in these numbers is still very important (IPCC, 2000). One reason is that data on global land cover are still incomplete, coarse, and inconsistent on land cover definitions used. Important questions, such as the area covered by disturbed and secondary forests in tropical regions, remain unanswered. As a consequence, CO_2 flows in these ecosystems are poorly known. Research that investigated CO_2 flows in tropical forests concluded that secondary and disturbed forests might be significant sinks of atmospheric carbon and therefore important for consideration in policy decisions about climate change mitigation activities (Brown and Lugo, 1992; Lugo and Brown, 1992; Fearnside, 1996; Fearnside and Guimaraes, 1996). However, the role of tropical forests in the global carbon balance is still uncertain. A clear picture about the areas covered by different types of forest has not yet been made available and is one of the major sources of uncertainties related to the debate on biological carbon sinks.

Despite these problems, remotely sensed data are the most complete and consistent sources of information about surface patterns and processes in our environment. This explains why in recent years, remote sensing has also acquired an increased importance for the political discussion about global change. As part of their commitments under the United Nation Framework Convention on Climate Change (UNFCCC), nations have to report on domestic emissions of Green Houses Gases (GHG) from land use and land cover change. To do this, they have to rely on land cover data obtained from remote sensing research. In the future, remote sensing will certainly play an important role for assessing compliance with Kyoto Protocol commitments and for refining the global carbon balance.

By providing essential data about the surface and the atmosphere of the Earth, remote sensing gives an important contribution to global change research and policy. The quality of the data output by remote sensing has thus important consequences, which justifies research for improving the quality of these data.

1.2 Problem statement

In the past three decades, emphasis of remote sensing research in tropical regions has been the assessment of deforestation and the improvement of the methods of its measurement (Downton, 1995). Despite the importance of measuring deforestation, the simplified picture "forest/non-forest" provided by most deforestation studies does not inform about areas that have recovered through secondary successions, were altered through timber logging, or were fragmented because of farm frontier expansion. These processes have important implications for carbon cycling, biodiversity conservation, hydrological cycle, and the sustainability of tropical agricultural systems. It is therefore important that remote sensing can be used to identify areas in which these processes occur. If secondary forests are mistaken for primary forests, carbon pools of tropical forests are likely to be overestimated, while carbon sinks are likely to be underestimated (Lugo and Brown, 1992; Fearnside and Guimaraes, 1996). Estimates of biodiversity and species extinction rates are also likely to vary depending on the data used about area, spatial distribution, and fragmentation of these different forest categories (MacArthur and Wilson, 1967; Westman *et al.*, 1989).

Unfortunately, the use of remotely sensed data for mapping secondary, logged, and primary (old-growth) forests is hindered by the weak spectral definition of these different vegetation categories. All information derived from remotely sensed data analysis is inferred from measurements of the electromagnetic energy that has been reflected or emitted from the objects present in the target area. A problem of object discrimination occurs, when different objects have similar reflection and emission patterns. This is the case of most vegetation categories.

The few studies that reported success in the discrimination of tropical secondary forests using remotely sensed data made use of high-density and high-quality ground data (Mausel *et al.*, 1993; Li *et al.*, 1994; Foody *et al.*, 1996). For the study of large tropical regions, methods that rely on large amounts of ground data can rarely be applied, because such data are not available or too costly to acquire. Classifier operations, combined with high-density ground data sampling, were also successfully used to improve the discrimination of tropical forest categories (Li *et al.*, 1994; Brondizio *et al.*, 1996; Foody *et al.*, 1996). However classifier operations (e.g.

Extraction and Classification of Homogeneous Objects, ECHO) are not yet available in standard remotely sensed data analysis packages and need farther validation.

Most study areas selected for remote sensing research are also too small to be representative for the standard remote sensing task: if a project is to utilize Landsat TM data, it is because low-cost information is required for a large area. Sensor systems of higher spatial and spectral resolution – that produce more accurate levels of object discrimination than Landsat TM – do not provide yet this low-cost large-area view. Rugged topography, ecological diversity, and advanced landscape fragmentation, as found in many parts of Central America, complicate the analysis of remotely sensed data (Pedroni and Velásquez, 1998). Conducting research in small areas is an advantage for keeping these sources of "spectral noise" under control, but the task is much more difficult to achieve in large study regions. Methods and conclusions of remote sensing research conducted in simple environments or in small study regions might thus be of limited application in large and complex tropical regions.

Multi-spectral data classification procedures, capable of discriminating tropical vegetation categories of different composition, structure, and function have thus still to be improved or developed from scratch. A need of practicable techniques for effective discrimination of tropical forest categories exists especially when the problem is addressed over large and complex tropical regions.

1.3 Study approach

In the last decade, a considerable amount of digital cartographic data has been made available by the increased use of Geographical Information Systems (GIS). This makes data analysis strategies that integrate remotely sensed data with other types of spatial data potentially usable. However, analysis techniques that merge ancillary spatial data sets and remotely sensed data have found limited application because of technical and practical problems that have been only partially solved. A major problem has been obtaining information about the spatial relationship between land cover categories and the ancillary, non-spectral variables previous to the classification. This is a key type of information for making the ancillary data set usable for the classification. Traditionally, such information has been obtained from sampling random points in the field or in remotely sensed data of finer spatial resolution (e.g. aerial photographs). Statistically, this was a correct approach. However, random sampling might be extremely costly, especially in large and often inaccessible tropical regions. As a consequence, this method was rarely applied.

The present work offers a solution to this problem, and demonstrates that the inclusion of information obtained from ancillary data into the classification process of Landsat TM data might result in improvements of the classification accuracy, that traditional methods could not achieve.

The methods tested in this research were developed keeping in mind the following principles:

- Technical feasibility: Only methods considered technically feasible to apply over large areas were investigated. This included requirements about density and design of the ground data.
- Representative data set: Remote sensing research is often conducted in small areas and using a single set of remotely sensed data. Conclusions might thus be questioned, especially if the study area was small and lacked of the complexity present in other regions. The study region chosen for this study was large compared to that selected

by other studies (see Part II), and presented a high degree of spectral complexity caused by its topography, land cover fragmentation, and ecological diversity.

- Costs of implementation: The methods tested in this research are believed not to be more expensive than those of traditional methods, provided an adequate set of ancillary data is available in digital format.

It was thought that the definition and compliance of these principles was necessary to prevent the investigation of data analysis techniques that would not be usable for potential users in tropical countries.

1.4 Aims and hypothesis of the study

The general aim of this work was to improve our ability to discriminate primary, logged, and secondary forest types in large tropical areas using remotely sensed data

Since the forest categories of interest were known to be spectrally similar, integration of ancillary non-spectral data in the classification process was chosen as the experimental approach. The underlying hypothesis was that individual land cover categories have distinctive associations with ecological and geographical variables such as elevation, accessibility, soil type, etc. Integration of knowledge about this type of spatial relationships in the classification of the satellite data was expected to improve the discrimination of primary, logged, and secondary forest types. To test this hypothesis, three ancillary variables were chosen: elevation above sea level, accessibility, and distance from the Pacific Coast.

Elevation data were hypothesized to help the discrimination of land cover categories that occur at different elevations (mangroves, tropical crops, coffee beans, subalpine paramo, etc.). Elevation data were also expected to improve the discrimination of secondary forests, since different categories of secondary forest occur in the mountains and in the lowlands of the study region (see Part III and Part IV).

Accessibility data were hypothesized to help the discrimination of human induced land cover categories, such as crops, pastures, secondary forests, and logged forests. Several studies have demonstrated that a spatial relationship exists between access conditions, land management, and the likelihood of deforestation (see below).

Data on distance from the Pacific coast were hypothesized to help the discrimination of land cover categories that were known to occur at specific distance ranges from the Costa Rican Pacific coast (mangroves, oil palm fields, banana fields, inundated palm forests, etc.). It was also observed, that patterns of brown vegetation caused by to the dry season were different in the Pacific-Atlantic direction of the study region. Since the satellite image was acquired at the end of the dry season (in March, 1996), this variable was also hypothesized to help the discrimination of land cover categories that had variable spectral patterns across the East-West direction because of variations in moisture content and green leaf biomass.

The specific aims of this work were:

(1) To test the Bayesian classification approach as a way to incorporate ancillary data into the maximum likelihood classification and to improve the accuracy of the classification. According to the literature (see below) the use of modified prior probabilities (Bayesian approach) can produce significant improvements of the classification accuracy, especially of spectrally similar land cover categories. However, the use of this method has been limited because the estimation of prior probabilities has been difficult, especially in large areas. Applications in tropical regions were not found in the literature.

(2) To use a new computer analysis technique, rather than random sampling, to model the spatial variability of the class prior probabilities. With the development of this technique it was hoped to make the Bayesian approach accessible for use in large areas.

(3) To test spectral data enhancement techniques (NDVI and Tasseled Cap) for improving the discrimination of tropical forest categories, and compare the classification results obtained with these techniques with those obtained with the Bayesian classification. The use of vegetation indexes is often claimed to enhance the spectral separability of vegetation categories. However, results reported in the literature are inconsistent. The hypothesis was thus considered worth to test for the discrimination of the forest categories of interest. The Bayesian classification was hypothesized to produce better classification results, because it exploits the ancillary information in addition to the spectral information.

2. THEORETICAL FRAMEWORK

2.1 Tropical forests and the missing carbon sink

Tropical forests are the largest terrestrial pool of global carbon (Table 1). Their clearing is a major source of atmospheric carbon dioxide (IPCC, 2000), but their regeneration through secondary succession is an important carbon (C) sink (Lugo and Brown, 1992; Fuentes, *et al.* 1995; Fearnside and Guimaraes, 1996). Whether tropical forests function as a planetary significant CO_2 sink, and possibly explain a significant part of the "missing carbon sink" in the global carbon balance, is uncertain (Fung, 1996; Houghton, 1996). Houghton (1996) defines these concepts as follows:

> *"Sinks can be defined as the net uptake or net accumulation of carbon. The world's oceans and growing forests are sinks of carbon. The missing sink refers to the imbalance in terms of the global C equation: emissions of C from combustion of fossil fuels (5.4 ± 0.5 Pg C yr^{-1} during the 1980s) and from changes in land use (1.6 ± 0.5 Pg C yr^{-1}) are not balanced by the accumulation of C in the atmosphere and oceans (3.4 ± 0.2 Pg C yr^{-1} and 2.0 ± 0.8 Pg C yr^{-1}, respectively). The imbalance (or missing C) is approximately of the same magnitude as, but opposite sign to, the net release of C from changes in land use. The imbalance is thought to be explained by the accumulation of C on land in undisturbed or incorrectly understood disturbed ecosystems"*[1].

The most important sources of uncertainties leading to the missing carbon sink are:

- The extent of area covered by forests and the rates of deforestation and forest recovery (Flint and Richards, 1994; Downton, 1995; Houghton, 1996).

---

[1] Newest figures presented in the IPCC report (2000) for the reference period 1850-1998 are, that global CO_2 emissions have been 270±30 Gt C from fossil fuel burning and cement production, and 136±55 Gt C from land use and land cover change. The increase of CO_2 in the atmosphere was 176±10 Gt C. Atmospheric concentrations increased from about 285 to 366 ppm. Thus, about 43% of the total emissions over the reference period have been retained in the atmosphere. The remainder is estimated to have been taken up in approximately equal amounts by the oceans and terrestrial ecosystems.

- Carbon flow rates after disturbance, particularly those related with biomass accumulation in secondary forests (Houghton and Hackler, 1994; Apps and Price, 1996; Fearnside, 1996).
- Biomass and carbon contents of both undisturbed and disturbed tropical forests (Brown and Lugo, 1992; Brown and Iverson, 1992; Brown, 1996).
- The ambiguity of definitions of biomes (IPCC, 2000), that is one classification scheme often used in remote sensing research.

The area covered by undisturbed, disturbed and secondary forests is thus an important question that remote sensing has to answer. However, as will be discussed below, answering this question in tropical regions is particularly challenging.

2.2 Categorization of secondary forests

Land cover maps are discrete representations of continuous and complex realities. Transitions between vegetation types are frequent in nature, while sharp boundaries are relatively rare. Where they exist, they are often man-made. This contrasts with the requirement of mutually exclusive and exhaustive definitions for map legend categories in remote sensing (Congalton, 1991). The variety of definitions existing for forest and secondary successions makes the definition of such categories yet more difficult, and is also a major cause of uncertainties about land cover and global carbon balance (IPCC, 2000). A review of the ecological literature, however, let conclude that some agreement about general patterns of forest successions is emerging. Succession stages, as defined in models of the succession process, might therefore be used for a classification scheme.

In Costa Rica, three models of secondary successions have been described in the literature. They have been used to define the classification scheme of secondary forests used in this research, which is presented and discussed with more detail in Part IV of the present work.

Finegan (1992 and 1996) describes secondary successions in the humid neotropical lowlands in terms of three phases. A different plant community dominates each phase of the succession, but the site is colonized by all species about at the beginning of the succession process. For most species, seed dispersal is assured by vertebrates. The more long-lived and shade tolerant species survive the whole succession process and arrive to dominance in the third phase of the succession. Short lived, fast growing, and shade intolerant species dominate the early stages of the succession, but disappear in later phases because of their inability to regenerate under shadow.

In the dry forest zone, Janzen (1988) and Sabogal (1992) describe a different succession process. Here, seed dispersal and site colonization depend on wind rather than vertebrates. Wind dispersed seeds fall close to their parent trees (Harper, 1977; Howe and Smallwood, 1982), and this makes tree colonization of abandoned pastures that are far from forest patches or remnant trees difficult and slower than in the humid tropics. The first colonizing tree population of abandoned dry zone pastures is composed of long-lived wind-dispersed species. A short-lived second phase of the succession, as described in the humid tropics, is thus absent in the tropical dry zone.

Kappelle (1995) and Kappelle *et al.* (1996) describe secondary successions in the high mountains of the Talamanca Range. Here, as in the dry zone, the succession process appears to develop in two phases, with species commonly found in old-growth forests already dominating in the second phase of mountain successions. This is particular feature of mountain successions, that is not observed for successions in the humid and in the dry tropical regions.

Secondary forests are thus a broad category encompassing very different types of vegetation. Most research that was successful in mapping secondary forests using remotely sensed data defined each phase of the succession as a separate land cover category (Mausel *et al.*, 1993; Foody and Curran, 1994; Foody *et al.*, 1996; Moran *et al.*, 1996). This approach appeared to be the best available classification scheme of secondary forests for the present work. However, its use for the classification of remotely sensed data is not lacking of conceptual and practical problems. Finegan (1996) points out that his succession model is valid for sites that have not been seriously degraded and for which seed sources are nearby. This observation suggests that the succession process might differ depending upon local conditions and the history of land use. In other tropical regions, variations in terms of species composition, growth rates, and time-frame, have been observed even for succession processes occurring at closely located sites (Foody *et al.*, 1990; Uhl, 1987). Biomass and carbon accumulation rates have been shown to vary among succession processes, depending on initial site conditions and previous land use, by a factor of 10 (Uhl *et al.*, 1988).

While providing a conceptual framework for the classification of secondary forests, the ecological literature remind us the secondary forest categories based on phase-models are neither mutually exclusive, nor exhaustive in terms of biophysical parameters such as biomass.

An approach for classifying remotely sensed data that does not imply obtaining overlapping categories in terms of biomass and other biophysical properties in the classified output data set is far to have been developed. Unfortunately, the exact relationship between biophysical properties of the vegetation and remotely sensed radiation has rarely been investigated (Foody and Curran, 1994). A correlation between data on biophysical properties of the vegetation and data on remotely sensed radiation is often hypothesized in remote sensing research. However, since few experiments have been realized to test this hypothesis, conclusions should not be precipitated about the relationship between land cover categories defined in a classification scheme and biomass and other biophysical properties of these categories.

A critical discussion of the relationship between biophysical parameters and spectral response of the vegetation is presented in Part IV of the present work.

2.3 Use of Landsat data in tropical regions

Among available remote sensing systems, Landsat Multi-Spectral Scanner (MSS) and Thematic Mapper (TM) have been the most widely used. These sensor systems provide thirty years of relatively low-cost data at a spatial and spectral resolution that is sufficient to discriminate most surface features over large areas (Jensen, 1996). The TM scanner system has been especially designed for the study of vegetation and Earth resources. It provides data for almost the entire world on a 16 days cycle. However, the use of Landsat TM data for mapping tropical vegetation has three inherent limitations:

- <u>Spectral similarity.</u> All green vegetation has about the same pattern of spectral response. Chlorophyll absorbs blue and red visible radiation and reflects green wavelengths. Leaf mesophyll reflects near infrared (IR), while increased water content and shadow absorb all wavelengths, but especially those in the IR portion of the spectrum. If chlorophyll content, mesophyll structure, water content, and shadow vary among vegetation categories, as a consequence of differences in species composition, structure, and moisture, then these vegetation categories should exhibit different patterns of spectral response. This is the underlying hypothesis of vegetation studies using remote sensing. However, only few studies managed to register remotely sensed data and biophysical vegetation data of the same date to test for correlation between the two data sets. Observed correlation were generally weak (Sader *et al.*, 1989; Foody and Curran, 1994).

- Sensor system resolution. Landsat TM data have a spatial resolution of 28.5 x 28.5 m. Thus, only one multi-spectral datum is available to describe all Earth features encompassed in a ground surface of that size (812.25 m^2). This is particularly problematic at the boundary of land cover categories. Highly fragmented landscapes are more difficult to classify correctly because the proportion of spectrally "mixed pixels" (boundary pixels) is high. Spatial resolution was found to be a major obstacle for the discrimination of secondary forests when using MSS data (80 x 80 m) and Advanced Very High Resolution Radiometer (AVHRR) data (1.1 x 1.1 km) (Woodwell *et al.*, 1986 and 1987; Sader *et al.*, 1990).

  The spectral resolution of Landsat TM data is of seven relatively broad spectral bands (three in the visible and four in the IR portion of the spectrum). Much information contained in the electromagnetic energy released by the Earth features is thus absent in the data of the sensor system. If spectral differences between vegetation categories exist only at specific wavelengths, Landsat data might not contain such information. For this reason, there is considerable hope that by increasing the number of spectral bands (hyper-spectral sensor systems) it will become possible to improve the discrimination among vegetation categories and other Earth resources. However, currently available hyper-spectral sensor system do not provide the synoptic view of Landsat.

- Clouds. Most tropical environments are almost permanently covered by clouds. The probability of obtaining cloud-free images for tropical regions from space-borne optical sensor systems is thus very small (Foody and Curran, 1994). Time series of cloud-free imagery would be advantageous for the detection of secondary forests less than thirty years old, which corresponds approximately to the time-frame in which secondary successions accumulate biomass (Brown and Lugo, 1990). However, because a sequence of cloud-free Landsat sensor imagery is seldom available for tropical regions, remote sensing research has to develop single-date data classification methods that are accurate enough to be used for the production of informative land cover maps.

2.4 Digital classification techniques

Several reasons have already been mentioned that made it difficult to discriminate among tropical secondary, logged, and undisturbed forests using Landsat TM data analysis (forest definition; spectral similarity; spatial, spectral, and temporal resolution of the remotely sensed data). The techniques more frequently used to classify Landsat sensor data represent another type of problem. Most of these techniques take the classification decision for each pixel individually, independently of its surrounding pixels and its geographical context. However, much information is present in the data of pixels surrounding a point, and in geographic context in which a given multi-spectral datum is classified. This information can potentially be analyzed and included in the digital classification process to improve the chance of a correct classification. Visual interpretation of aerial photographs and color composites of digital images does always take such information into account. For that reason, Tuomisto *et al.* (1995) could achieve a better visual interpretation of Landsat TM color composites than a reliable pixel-by-pixel digital classification of the data. Visual interpretation does not only observe the 'color' of a pixel (its spectral pattern), but also its spatial context, its relationship with neighboring pixels, the shape of the object to which it belongs, etc. All these features are not considered in digital pixel-by-pixel classification techniques, which are thus unable to extract the whole information content inherent in the multi-spectral data set.

Previous research that has attempted to improve the extraction of information from the multi-spectral data set has been based on two hypothesis. The first hypothesis is that subtle spectral differences between the land cover categories of interest are present in the data, but not adequately enhanced to allow the maximum likelihood decision rule to be sensitive to them. Research based on this assumption developed spectral data enhancement techniques, the best known among them being the computation of so called 'vegetation indexes', such as the Normalized Difference Vegetation Index (NDVI) and the Tasseled Cap transformation (Sader *et al.* 1989; Foody and Curran, 1994; Crist and Cicone, 1994; Sader, 1995; Helmer, 1999).

The second hypothesis is that variations in the spectral response of neighboring pixels reflect texture variations of the Earth features of interest, for example differences in canopy roughness and vegetation structure between secondary and old-growth forests. Information about image texture can be extracted digitally and included in the classification process. Proposed techniques have been spatial filtering of the data to reduce data variance and enhance spectral-textural differences among the land cover categories of interest (Hill and Foody, 1994), or addition of texture measures extracted from an n x n window to the multi-spectral data set (Franklin and Peddle, 1989; Jensen, 1996).

The technique of adding data layers to the original image data about information extracted from the multi-spectral data set itself has most commonly been used in the case of vegetation indexes. These are obtained through mathematical combinations of the original spectral data and are intended to enhance some particular feature present in the data that might be correlated with the land cover categories of interest. For example, NDVI has widely been used because it is supposed to help the detection of green leaf biomass, since it enhances the difference between red, that is absorbed by chlorophyll, and near-IR, that is reflected by the leaf mesophyll (Jensen, 1996). The NDVI is simply a ratio of the difference between near-IR and red band values over the sum of the same band values:

$$NDVI_{TM} = \frac{TM4 - TM3}{TM3 - TM4}, \qquad (1)$$

where TM3 is the third band of Landsat TM (red, 0.63-0.69 μm) and TM4 is the fourth band of Landsat TM (near infrared, 0.76-0.90 μm) (Jensen, 1996).

Other frequently used indexes are the "Brightness", "Greenness", and "Wetness" features derived from the tasseled cap transformation (Crist and Cicone, 1984; Jensen, 1996; Helmer, 1999). These indexes are obtained from regression equations applied to the original spectral data, and are supposed to relate with green leaf biomass, vegetation structure, and moisture (Table 2).

Table 2. Transformation coefficients for the creation of the Tasseled Cap Indexes using Landsat TM data

Feature	TM1	TM2	TM3	TM4	TM5	TM7
Brightness	0.33183	0.33121	0.55177	0.42514	0.48087	0.25242
Greenness	-0.24717	-0.16263	-0.40639	0.85468	0.05493	-0.11749
Third (Wetness)	0.13929	0.22490	0.40359	0.25178	-0.70133	-0.45732
Fourth	0.84610	-0.70310	-0.46400	-0.00320	-0.04920	-0.01190

Source: Crist and Cicone, 1984.

Research that used these techniques has not always achieved an improvement of the classification accuracy of tropical forest categories (Sader *et al.*, 1989; Foody and Curran, 1994; Puig, 1996; Helmer, 1999). Either a strong correlation between vegetation indexes and specific ecological features of the vegetation measured in the field could not be found (Foody and Curran, 1994), or, in the case of the addition of textural measures to the multi-spectral data set, the normality assumption of the data distribution was violated (Hutchinson, 1992; Jensen, 1996).

Only classifier operations, such as the use of textural classifiers, that are algorithms that do not take a per-pixel but a per-object (group of pixels) classification decision, were reported to consistently improve the classification accuracy of secondary forest types (Li *et al.*, 1994; Brondizio *et al.*, 1996; Foody *et al.*, 1996). Such research was however related to an effort for ground data collection that is seldom feasible for the study of large tropical regions. Object classifiers are also not yet included in standard image processing software.

2.5 Use of ancillary data in remote sensing

The use of contextual information, going beyond that extracted from the multi-spectral data themselves, has rarely been explored to solve the classification problem of tropical secondary, logged, and undisturbed forests. However, spatial data might include useful information to mathematically describe the geographical and ecological context of a particular image pixel that has to be classified. The potential usefulness of such data to improve the classification accuracy of remotely sensed data is straightforward.

Most vegetation types occur only within a range of ecological conditions, such as elevation above sea level, latitude, soil type, and others. Human choices do also largely determine the spatial distribution of land cover categories: crops are grown close to roads and markets, forests are logged where timber extraction is financially lucrative, and secondary forests are left to grow on abandoned farms that must therefore have been made accessible for humans.

In Costa Rica, an association between deforestation and climate, slope, soil fertility and infrastructure for human access to forests has been demonstrated by previous studies (Sader and Joyce, 1988; Veldkamp *et al.*, 1992). Similar correlations between deforestation and landscape variables were also found in Brazil (Stone *et al.*, 1991; Moran *et al.*, 1994), Guatemala (Sader, 1995), Guinea (Gilruth *et al.*, 1995), Honduras (Ludeke *et al.*, 1990), Madagascar (Green and Sussman, 1990), Mexico (Dirzo and Garcia, 1992), Philippines (Kummer and Tuner, 1994), and in other parts of the world. Since secondary successions develop on lands that were cleared for farming or grazing, and then abandoned, a correlation should exist between secondary forests and certain landscape features such as infrastructure for human access.

If the patterns of spatial distribution of the cover types of interest are correlated with variables that can be georeferenced using a GIS, then different strategies are available to use the information content of such ancillary variables to improve the accuracy of the classification (Hutchinson, 1982).

The simplest way of including ancillary variables in the discriminant analysis is to add an ancillary data layer to the image data set, before the classification. This method does not always improve the classification accuracy (Hutchinson, 1992; Gallo, 1999). Environmental variables have often skewed or multimodal distributions (Hutchinson, 1992; Flack, 1995; Jensen, 1996). Simply adding them to the multispectral data set thus violates the normal distribution assumption required for the maximum likelihood discriminant analysis. To avoid this problem, the ancillary variables can be used in pre-classification image stratification (Franklin and Wilson, 1992) or in post-classification sorting (Hutchinson, 1992; Cibula and Nyquist, 1987). These are simple and effective ways to improve the classification accuracy, but they are limited by the artificially sharp boundaries they create in the classified output data set due to their "*deterministic, inflexible nature*" (Maselli *et al.*, 1995).

2.6 Bayes's classification with modified prior probabilities

A technique that allows inclusion of ancillary data into the classification process without requiring assumptions about the distribution of the ancillary data or forcing the classification to become deterministic, is the classification with modified prior probabilities, also known as Bayesian approach (Swain and Davis, 1978; Strahler, 1980; Hutchinson, 1982; Mather, 1985; Maselli *et al.*, 1995). To understand this technique, it is necessary to briefly review the mathematics of the maximum likelihood classification.

The maximum likelihood decision rule is based on a normalized (Gaussian) estimate of the probability density function of each spectral class. The probability density function for a pixel x_k can be expressed as (Foody *et al.* 1992):

$$p(x_k \mid i) = \frac{e^{[-1/2(x_k - u_i)' V_i^{-1}(x_k - u_i)]}}{[(2\pi)^{n/2} \mid V_i \mid^{1/2}]} \qquad (2)$$

where $p(x_k \mid i)$ is the probability density function for a pixel x_k to be member of class i, n is the number of channels present in the image, x_k is the data vector for the pixel in all channels, u_i is the mean vector for class i over all pixels, and V_i is the variance-covariance matrix for class i. The maximum likelihood decision rule assigns the pixel x_k to the class for which equation (2) results in the greatest probability value. In practice, classification algorithms use a logarithmic form of the maximum likelihood decision rule in which all constants are eliminated. Following the mathematical manipulations shown in Strahler (1980), the maximum likelihood decision rule can be expressed by the following discriminant function $F_{1,k}(x_k)$:

$$F_{1,k}(x_k) = \ln \mid V_i \mid + (x_k - u_i)' V_i^{-1}(x_k - u_i) \qquad (3)$$

The pixel x_k is assigned to the class for which the discriminant function results in the lowest value.

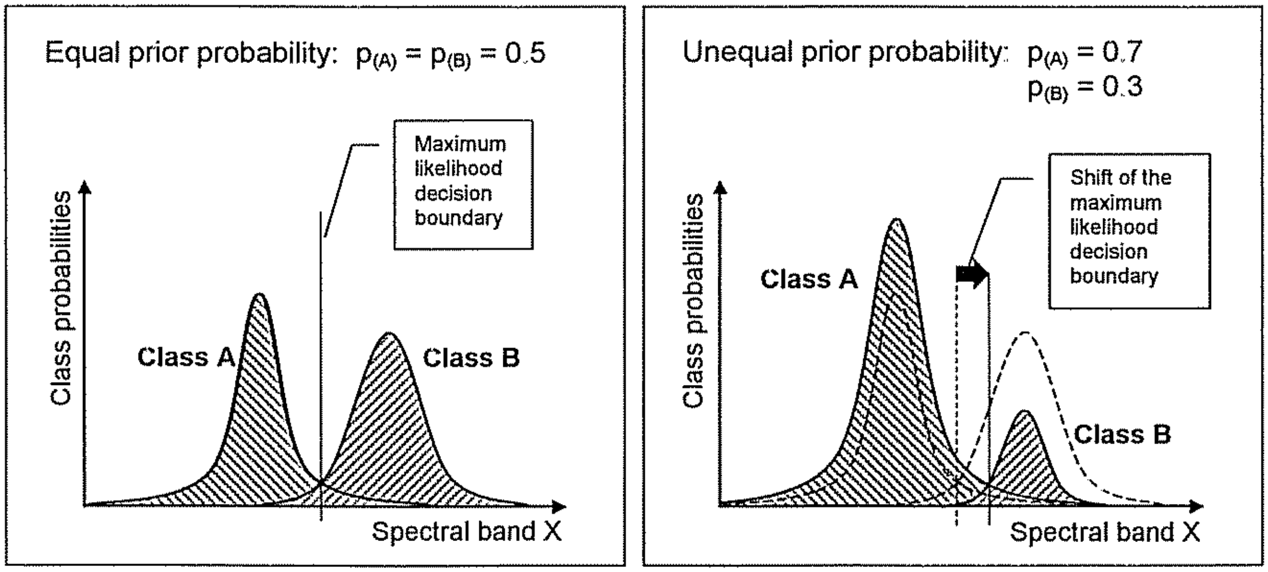
The Bayes's decision rule is identical to the maximum likelihood decision rule, except that it does not assume that each class has the same probability to occur. In almost every remote sensing application, some classes are encountered more often than others. The maximum likelihood decision rule can be modified to take into account the prior probabilities of each class, which are simply the expected area proportions of the classes in a particular scene or stratum. Their incorporation into the maximum likelihood decision rule occurs via manipulation of the Law of Conditional Probabilities (Strahler 1980). Mathematically, this occurs by adding the term $-2\ln P_i$ to equation (3):

$$F_{2,k}(x_k) = \ln|V_i| + (x_k - u_i)'V_i^{-1}(x_k-u_i) - 2\ln P_i \qquad (4)$$

where P_i is the prior probability of class i and $F_{2,k}(x_k)$ is the discriminant function that takes into account the class prior probabilities. The sum of the P_i of all classes must be 1.0 for each pixel, because each pixel must be assigned to a class.

Graphically, it can be observed, that changing the prior probabilities results in a shift of the maximum likelihood decision boundary (Figure 1).

Figure 1. Shift of the maximum likelihood decision boundary caused by modified prior probabilities



As a consequence of this shift, a greater range of spectral data values will be assigned to the more frequent class (A) and less to the rarer one (B). In the extreme case that the prior probability of class B is zero, all pixels will be assigned to class A.

Availability of ancillary data for the definition of the pixel's geographical and ecological context allows a different set of prior probabilities to be assigned to each pixel. For example, to a pixel closely located to the Costa Rican Pacific coast - that might actually be mangrove forest - a small prior probability might be assigned to the class "subalpine paramo" and a high prior probability to the class "mangrove forest". The opposite would then be done for a pixel located in a mountain region that is actually a subalpine paramo. In that way, the probability of having both pixels classified correctly will be higher than with the traditional approach, that would assume equal class prior probabilities for both pixels.

In the praxis, assigning a different set of prior probabilities to each individual pixel will not make sense because it would be impossible to estimate a different set of prior probabilities for each of the numerous pixels present in an image. Instead, the ancillary variables relevant to the spatial distribution of the cover types of interest can be used to stratify the study region in portions of approximately invariant ecological and geographical conditions. The same set of prior probabilities can then be assigned to each pixel within each of these strata. For determining which set of prior probabilities should be assigned to the pixels within a stratum, it is necessary to know the relative surface occupied by the individual land cover categories within each stratum. Observed land cover frequencies (for example from stratified random sampling in the field) can thus be used as an estimate of the class prior probabilities.

Using the methods described above to classify remotely sensed data, only the land cover categories that are likely to occur in a given geographical and ecological context have a chance to classify the image pixels. Spectrally similar categories that do occur in the scene, but not at the same geographical and ecological context, have therefore little chance to become classified. This mechanism increases the classification accuracy of spectrally similar land cover categories, unless they occur with the same frequency in similar conditions.

It has been shown that this classification method can produce great improvements of the classification accuracy of spectrally similar land cover categories, while not affecting the classification decision for spectral classes that are clearly different from others, and for which the spectral information is sufficient to allow a reliable classification (Mather, 1985; Maselli *et al.*, 1995).

The major drawback of this method has been the need of random sampling, or stratified random sampling, to obtain class frequency estimates. In large tropical environments, where major portions of the scene can have prohibitive access conditions and up-to-date aerial photographs may not be available, random sampling techniques might be simply impossible to achieve. For that reason, the method has not found widespread application. An approach to solve this problem is shown in this dissertation.

3. MATERIALS AND METHODS

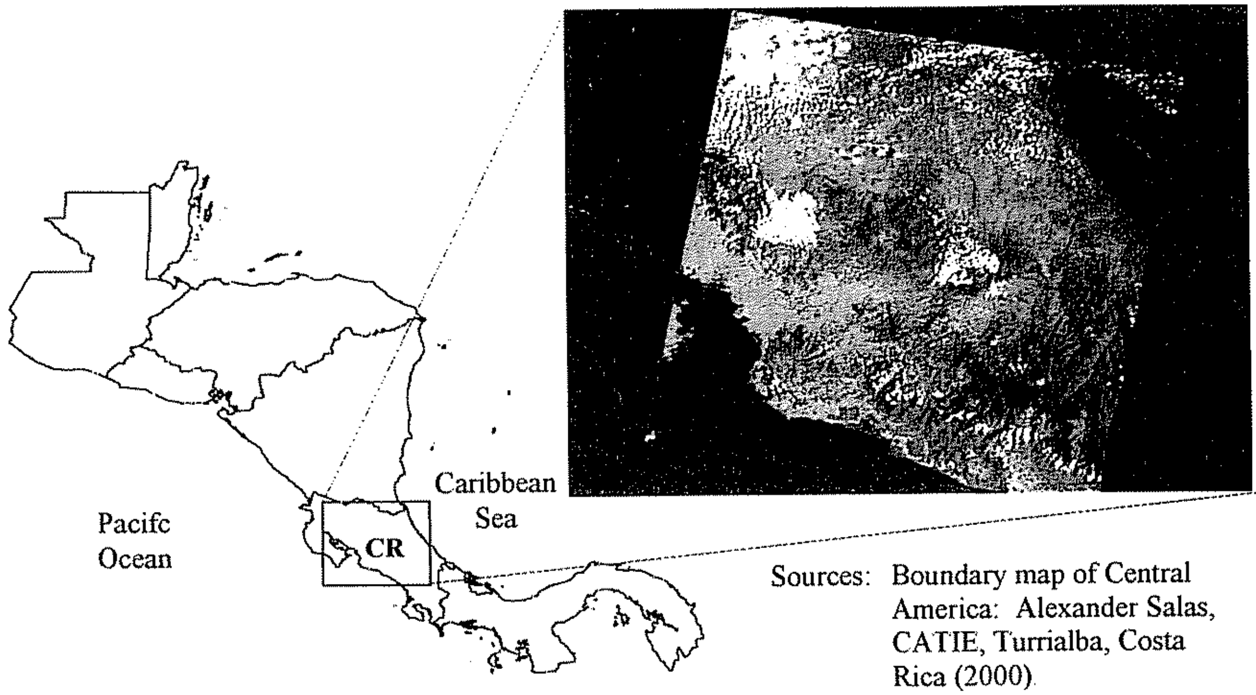
3.1. Study region

The tropical study region selected for this research was the central portion of Costa Rica corresponding to path 15 and row 53 of Landsat TM (Figure 2).

The study region includes approximately 30,950 km^2 and is therefore large compared to that used in most classification experiments that have been reported in the literature (see Part II). About 5433 km^2 (17.5%) are covered by water, and the rest is land with elevations ranging from 0 to 3825 m.a.s.l., and with considerable variations in slope and aspect. Because of the presence of the Central Volcanic and Talamanca mountain ranges and the influence of trade winds, climatic conditions are extremely variable, with yearly average precipitation ranging from 1400 mm yr^{-1} to more than 7000 mm yr^{-1} (IMN, 1987).

Because of the broad range of climatic conditions, a wide array of natural ecosystems, represented by 12 life-zones and 11 transition-zones (*sensu* Holdridge *et al.*, 1971), exists in the study region. As a consequence of this ecological diversity, the study region was spectrally complexity and particularly challenging to classify. The small to medium sized land tenure, the fragmented patterns of forest cover, the mixed forms of land-use, the rugged topography, and the presence of clouds and haze added considerable spectral "noise" to the image data.

Figure 2. Study region



Sources: Boundary map of Central America: Alexander Salas, CATIE, Turrialba, Costa Rica (2000).

Satellite image: Landsat TM from March, 5th, 1996 (Path15, Row 53)

Note: CR = Costa Rica

3.2 Data collection and preparation

Three sets of data were used in this research: satellite data, ancillary spatial data (models generated with a GIS of landscape variables considered relevant to the spatial distribution of the spectral classes), and ground-truth data. The three data sets were used for classification (satellite data), estimation of georeferenced sets of class prior probabilities (satellite data + ancillary data), spectral signature extraction (satellite data + ground-truth data), and classification accuracy assessment (ground-truth data).

Satellite data:

- A Landsat TM image from March, 5th, 1996 (path 15 / row 53) was obtained from the Foundation for the Development of the Central Volcanic Mountain Range, FUNDECOR. Before classification, the image was georeferenced to the coordinate system Lambert Conformal Conic of north Costa Rica.
  To reduce the quantity of data to be processed, the thermal band (TM band 6) was eliminated from the satellite image because it was considered of too low spatial resolution and spectral contrast to be helpful for the discrimination of forest categories.
  No atmospheric correction were performed (because of lack of data about atmospheric conditions in March, 5th, 1996), but the Normalized Difference Vegetation Index (NDVI) and the three first features of the Tasseled Cap transformation were added to the reflective bands, because the literature review suggested that they helped to improve the discrimination of

vegetation categories. The Tasseled Cap features were computed using the regression equations shown in Table 2.

Ancillary spatial data:

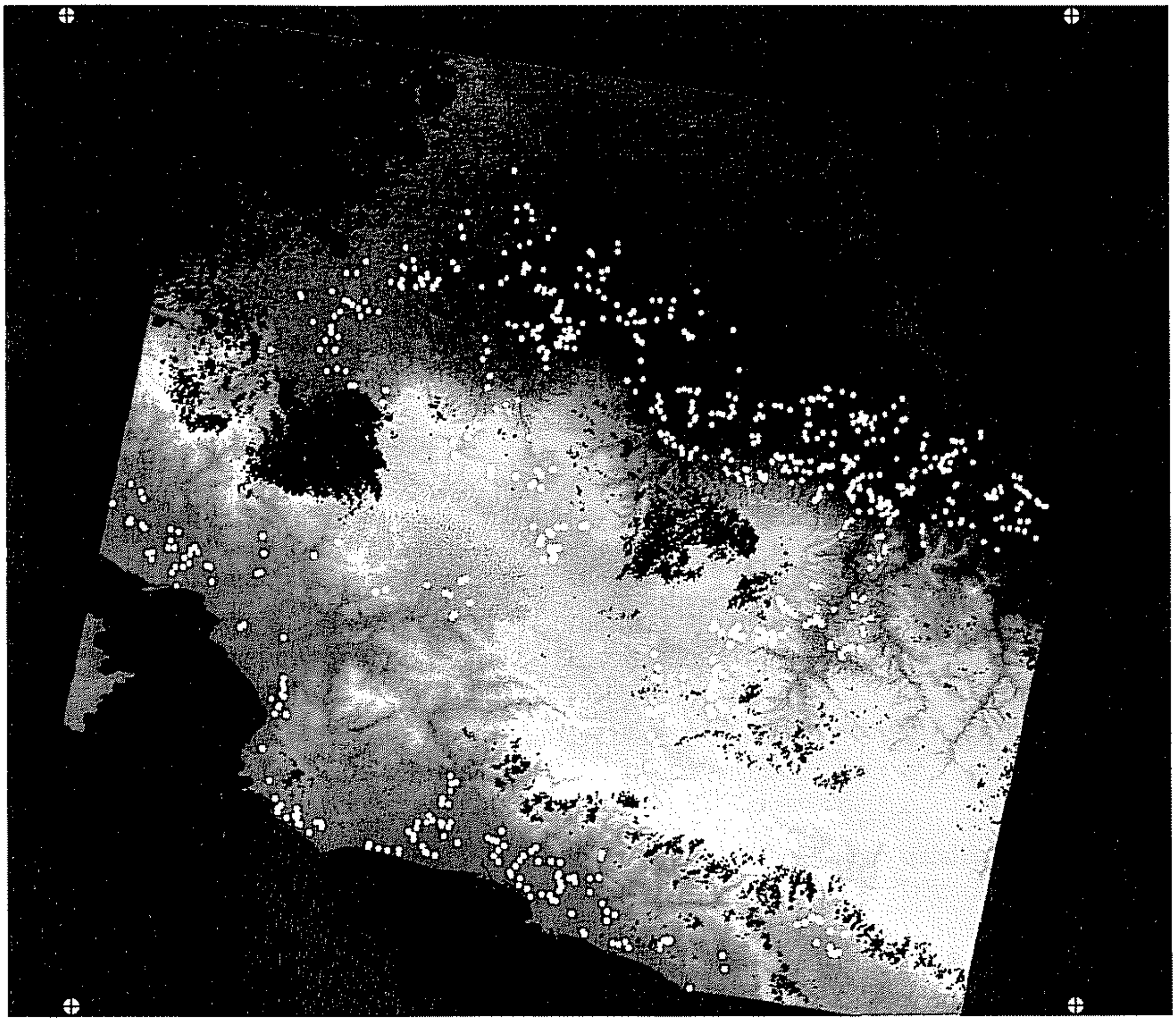
- Digital Elevation Model (DEM). A DEM was generated from interpolation of contour lines data using ARC/INFO 7.2.1 TOPOGRID module. The digital contour lines data were obtained from the Costa Rican National Geographic Institute (IGN), that generated them from scanned 1 : 50,000 scale maps. Since the IGN could not provide digital contour line data for the entire study region, additional 1 : 50,000 and 1 : 200,000 scale contour lines data were obtained from the Department of Geography at the University of Costa Rica (UCR), and from CATIE's GIS laboratory, respectively, to complete a digital contour line data base for the entire study region. The DEM, as all other ancillary data used in this research, was generated at the same spatial resolution of the satellite image.
- Digital accessibility model. This model was generated from slope data derived from the DEM and digital road data provided at a 1 : 50,000 map scale by the IGN. Because a large proportion of the study region presented rugged terrain, it was thought that the use of an Euclidian distance function from roads would not adequately represent the access conditions existing in the study region. In reality, the time and energy required to access a field is a function of several variables. In this research, slope and distance from the closest road were considered to be the most important variables to model conditions of access to a spot. Firstly, using walking time data obtained from interviews to local farmers and foresters, a regression equation was constructed to estimate the time a person would need to cross a pixel (28.5 m) as a function of its slope. Then, the regression equation was applied to each image pixel using slope data derived from the DEM. Finally, the least cumulative walking time was calculated for each pixel, using the closest road as starting point of the calculation. This walking time model was considered to adequately represent the access conditions of each image pixel.
- Digital model of distance categories from the Pacific Coast and export banana fields. Using a boundary coverage of Costa Rica (digitized from 1 : 200,000 scale maps), and boundary polygons of banana fields (screen digitized over the satellite image), a distance model was created representing distance categories increasing from the boundary of the banana fields toward the Pacific Coast. As discussed previously, this model was created because it was hypothesized that under the particular conditions of the study region it would have been helpful to improve the discrimination of land cover categories such as mangrove forests, oil palm fields, banana fields, inundated palm forest, and others.
- Digital life zone map *sensu* Holdridge *et al.* (1971). This map was digitized from 1 : 200,000 scale maps obtained from the Tropical Science Center (CCT) of Costa Rica and was used for the selection of spectral observations from ecologically similar conditions (Part III).

Ground-truth data:

- 826 field site descriptions were obtained from field surveys carried out during the period from March, 1998, to March, 1999. The center coordinates of each field site were determined using a Global Positioning System (GPS) unit (Garmin XL 12). This data set was used for signature extraction and classification consistency evaluation (Figure 3).

Figure 3. Location of ground-truth data of this study

84°57'16" W / 10°58'38"N | 83°13'03"W / 10°58'38"N



84°57'16"W / 9°14'00"N | 83°13'03" W / 9°14'00" N

Notes:

- White spots represent the location of 826 ground-truth sites.
- Black areas represent regions covered by clouds, shadow, oceans, and no data (image background).
- Colors represent different geographical and ecological conditions according to the ancillary variable (R,G,B, display of accessibility, DEM, and distance categories).

The location of most field described sites was determined previous to field work, using the following criteria:

(1) Representative of all spectral patterns present in the image. Spectral classes not present in the training data set have a great probability of not being classified correctly by the maximum likelihood algorithm. Through visual interpretation of the satellite image it was made sure that all spectral patterns were present in the training data obtained from field described sites.

(2) Representative of the study region: As much as possible, ground truth data were gathered well distributed in all three coordinates of the study region (Latitude, Longitude, and elevation). Study regions not surveyed in the field might harbor land cover categories not present in the map legend. A good spatial distribution of the ground-truth data reduces the probability of forgetting a land cover category in the classification scheme. Classification schemes must be exhaustive (Congalton, 1991).

(3) Useful for the development of spectral signatures: Field sites used to develop training signatures must be spectrally homogeneous and large enough to make it possible to gather at least 10 spectral measurements (pixels) for each band of the satellite image (Jensen, 1996). For the classification of a 7 band Landsat TM image, this requirement would represent 70 pixels, or an area of about 5.68 hectares, boundary pixels excluded. In highly rugged and fragmented landscapes, sites that meet these conditions are rare and must therefore be sought in the computer screen previous to field work.

(4) Not under clouds or shadow: About 13.2% of the study region was either under clouds or shadowed. It was necessary to avoid collecting ground-truth data in areas that were not visible in the satellite image.

(5) Easy to access: Limited resources were available for field survey. To reduce the costs of fieldwork the collection of ground-truth data at location that were difficult to access was reduced to a necessary minimum.

- 252 field site descriptions made in 1996 by other researchers were obtained from Oregon State University (Ph.D. research data of Helmer, 1999) and FUNDECOR. This data set was used for classification accuracy assessment only.

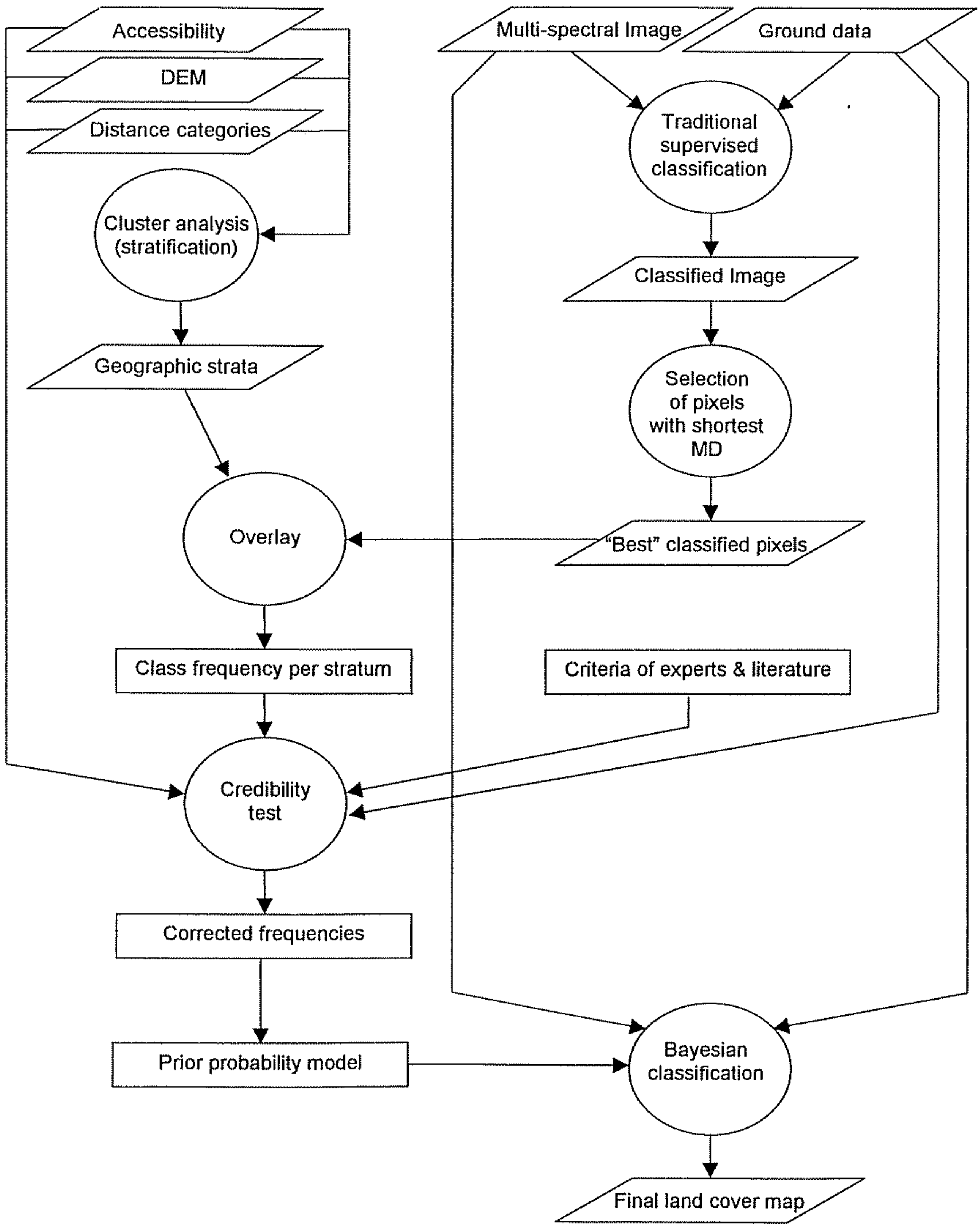
3.3 Sequence of data analysis

The basic approach of this research has been to merge two sources of information for improving the quality of the classification. The first source of information is traditional in remote sensing: the spectral information. This information was obtained from well known supervised procedures, and will thus not be discussed farther (more details are given in Parts II and III).

The second source of information was obtained from the ancillary data, and was basically a model of how the frequencies of the different land cover categories change as a function of the ancillary variables. This model was then incorporated into the Bayesian maximum likelihood classification in form of modified prior probabilities.

The different steps required to accomplish this data analysis procedure are shown in a flow diagram (Figure 4) and are described with more detail in the next sections.

Figure 4. Flow-diagram of data analysis

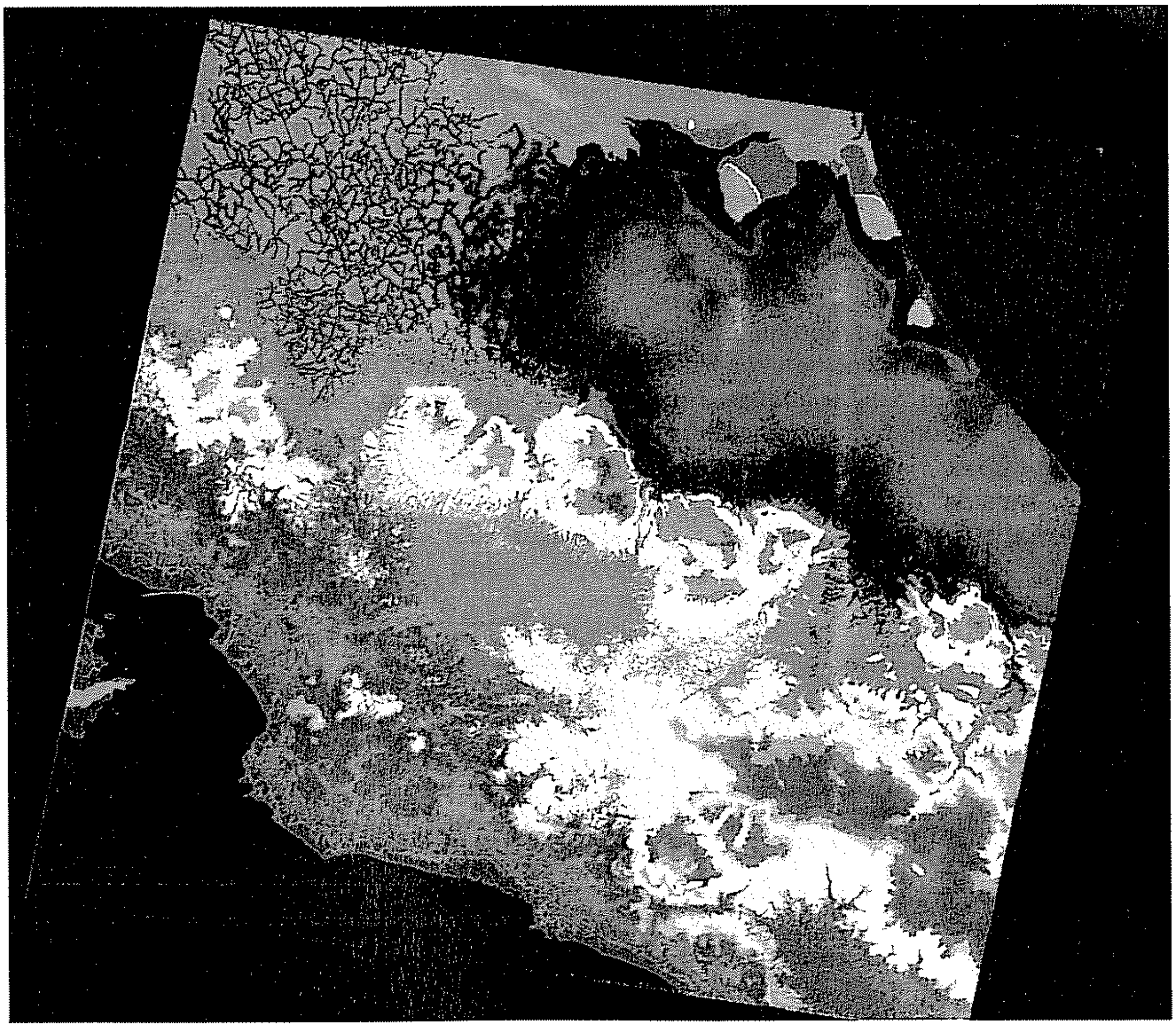


3.4 Stratification of the study region

Theoretically, a different set of prior probabilities can be used for the bayesian classification of each pixel. In practice, estimating a separate set of prior probabilities for each individual pixel is impossible. Instead, the study region can be stratified in portions of about invariant ecological and geographical conditions. All pixels contained in the same stratum can then be classified using the same set of prior probabilities. This approach was chosen for the present study.

Figure 5. Stratification of the study region

84°57'16" W / 10°58'38"N | 83°13'03"W / 10°58'38"N



84°57'16"W / 9°14'00"N | 83°13'03" W / 9°14'00" N

Note: The colors of this image represent 537 geographical strata. For each stratum, a different set of prior probabilities was estimated and used for the Bayesian classification.

To stratify the study region the three ancillary variables representing elevation, accessibility and distance categories were used However, 256^3 possible strata (supposing 8 bit data) would have resulted from the linear combination of the data of these variables (the colors of Figure 3 are a graphical representation of these combinations). To reduce the number of strata to some number that was considered easier to handle, ERDAS Imagine 8.3.1 ISODATA clustering algorithm was used. The convergence threshold was set to 99% and reached after a night of iterations. As result of the clustering process, 537 strata were delimited (Figure 5).

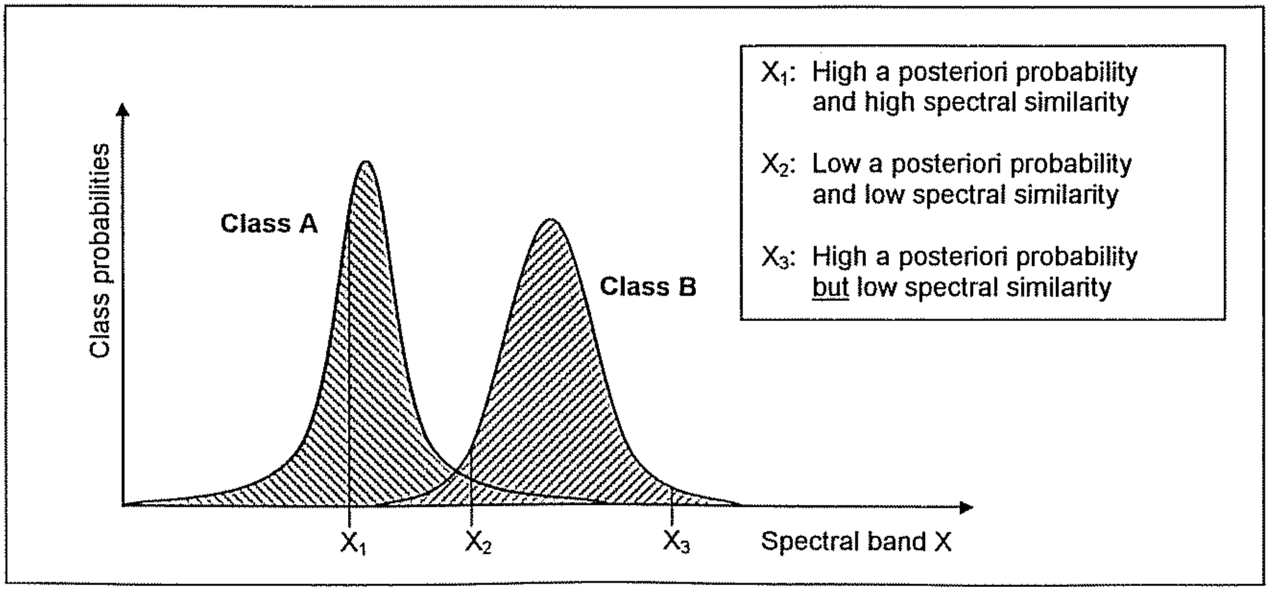
3.5 Estimation of prior probabilities

Because stratified random sampling to determine the class frequencies within each of the 537 strata would have been impossible to achieve through fieldwork or interpretation of aerial photographs, an alternative method was developed. The method used a selection of pixels previously classified using the traditional equal prior maximum likelihood decision rule to sample the spectral class frequencies in each image stratum. The sample pixels were selected for their shortest Mahalanobis Distance to the mean values of the spectral signatures used for training. It has been shown, that the Mahalanobis Distances is a good a measure of spectral similarity between a pixel and the mean of a spectral class (Foody *et al.*, 1992). Mathematically, the MD is expressed by the following formula:

$$MD = (x_k - u_i)' V_i^{-1} (x_k - u_i), \qquad (5)$$

where MD is the Mahalanobis Distance between the pixel x_k and the mean of class i, u_i is the mean vector for class i over all pixels, and V_i is the variance-covariance matrix of class i (Foody *et al.* 1992). It was thus assumed that pixels with greatest spectral similarity to the mean of their assigned spectral class had a greater chance to be classified correctly. As measure of spectral similarity the MD was preferred over the *a posteriori* probability, since some pixels might have high *a posteriori* probability values even if they are spectrally dissimilar to the class to which their were allocated (Figure 6). This is for example the case, when some spectral classes are not adequately represented in the training data set (Foody *et al.*, 1992).

Figure 6. *A posteriori* probability as an inadequate measure of spectral similarity



Even if the threshold value of the Mahalanobis Distance was set for a high confidence level (95%), 909,311 pixels passed the selection criterion. Nevertheless, not all of them were assumed to have been classified correctly. A small proportion of incorrectly classified pixels was considered likely to exist even among pixels that were spectrally identical to the mean of the spectral class to which they were allocated. Their use as class frequency samples would therefore have resulted in biased frequency estimates. To minimize this risk, all selected pixels whose incorrect class allocation was recognizable were eliminated before using them to estimate the prior probability.

To identify possibly misclassified pixels among the MD threshold selection, the range was defined for ecological and geographical conditions that was considered allowable for each particular spectral class according to interviews with local agronomic experts and personal experience. Pixels found beyond this allowed range were eliminated. This process required ARC/INFO arc macro language (aml) data processing of the MD pixel selection and of the three ancillary variables.

A low pixel frequency (0.1 pixels) was then added to the number of pixels left over for all classes that could not be excluded from a particular stratum, according to the ecological and geographical criteria. As suggested by Maselli *et al.* (1995), this trick is required to avoid the exclusion of a class in a particular geographic stratum just because it is not present in the frequency sample. Finally, the pixel frequencies of the training data were added to the refined MD selection of pixels, which yielded a body of 764,636 pixels for the estimation of the class frequencies. From these pixels, 14.38% were obtained from the training data, and 85.62% were selected with the MD threshold criterion and surpassed the ecological and geographical credibility test.

The sample size of 764,636 pixels would have been impossible to gather through stratified random sampling in the field, but was large enough to estimate the frequency of 33 land cover categories in the 537 image strata.

To generate the prior probability set to be applied in each stratum, the pixels frequency of each class was divided by the total number of pixels sampled in the stratum.

3.6 Classification experiments

To test the effectiveness of different spectral band and vegetation index combinations for improving the discrimination, 18 classification experiments were performed using a supervised maximum likelihood approach. The band and index combination that resulted in the best classification output for the forest categories was then classified again using the spatially variant prior probability sets that were estimated using the methods described above. The computing routine of this classification required ERDAS macro language (eml) and spatial modeler (sml) programming, as well as a Visual Basic program written in house to control the flow of the program execution.

3.7 Classification accuracy assessment

A statistically sound evaluation of the classification accuracy could not be achieved in this research, because ground truth data meeting all requirements for such an assessment were not available (Congalton, 1991).

Nevertheless, enough evidence to compare the performance of the different classification experiments was obtained from:

- The classification self-consistency of the training data set (979 sites, from which 826 were visited in the field, and 153 identified through visual interpretation of the satellite image), and
- The classification accuracy of the control data set (252 sites visited and described in the field by other researchers).

4. RESULTS

4.1 Prior probability model

The final prior probability model was a set of 537 tables containing prior probability values for each class. Each table was related with a particular stratum, and contained prior probability values for the 927 spectral classes used to classify the scene. Graphically, the model is presented in Figure 7, and discussed in detail in Part II. In these figures the prior probabilities are represented as functions of the three ancillary variables that were used to stratify the study region. As was expected, the likelihood of occurrence of the different categories of land cover showed great variations along the gradients. For most categories the range of variations was from 0% to 100%. This pattern contrast with the implicit assumption made with the traditional equal prior classification (Figure 8), in which each class, at whatever location, has exactly the same prior probability to occur (3.03% in a classification with 33 land cover categories).

The tail regions of the probability distributions do not appear truncated, which indicates that the elimination of pixels from the 95% confidence threshold did not generate artifacts in the model.

The more horizontal the distribution of probabilities in the graphics of Figure 7, the less the variable contributed to stratification. Time of access appears thus to have given the smallest contribution to stratification. However, it was the most important variable for improving the discrimination of forest categories, especially old-growth forests from logged and secondary forests. This is particularly evident, when one observes the likelihood distribution of these categories in the time of access variable. A variable can thus be important for improving discrimination, but less important for stratification (see Part II).

Elevation above sea level did also contribute to improve the discrimination of old-growth forests, because greater abundance of undisturbed ecosystems is found at higher elevations, probably because protected areas larger in mountain and foothill regions.

The distribution of secondary forests and logged forests is also slightly different to that of old-growth forests for the variable distance from the Pacific Coast. The reason is probably the same as for the variable elevation above sea level, because regions of higher elevation are concentrated at specific ranges of distance from the Pacific Coast.

Only stratified random sampling in the field could have been used to make a quantitative assessment of the quality of the prior probability model. However, common sense and previous knowledge about the spatial distribution of land cover in central Costa Rica can be used to allow the conclusion that the prior probability model reflected, with an high degree of certainty, the true land cover patterns in the study region. The methods used to estimate the prior probabilities appear thus appropriate to produce useful information for the Bayesian classification. The elevated number of sample points produced by the proposed technique (Figure 9) might offset, up to a certain degree, the bias inherent in a design that was not random and that might have resulted in the selection of dome sample points that were not assigned to their true class

Figure 7. Prior probability model (equal-area representation)

Figure 7.a. Land cover probability as a function of walking time from the closest road

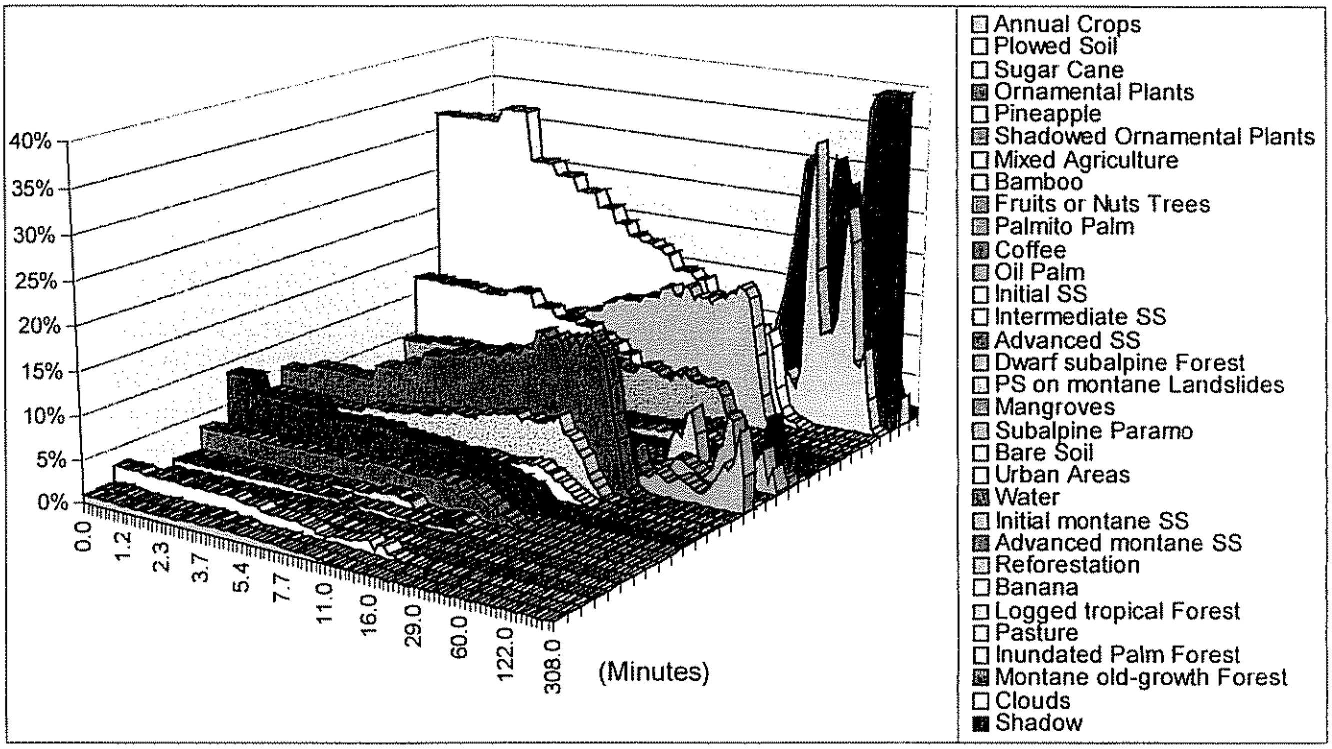


Figure 7.b. Land cover probability as a function of elevation above sea level

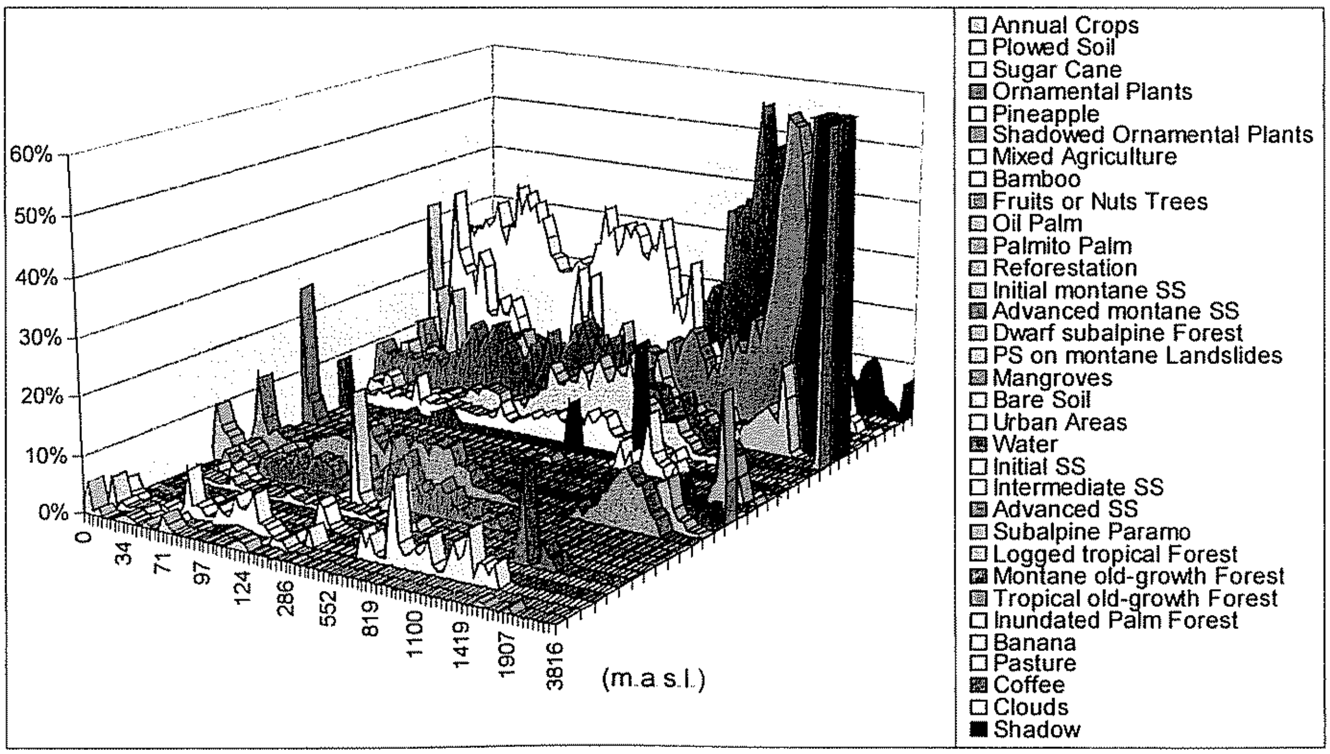


Figure 7.c. Land cover probability as a function of distance from the Pacific Coast

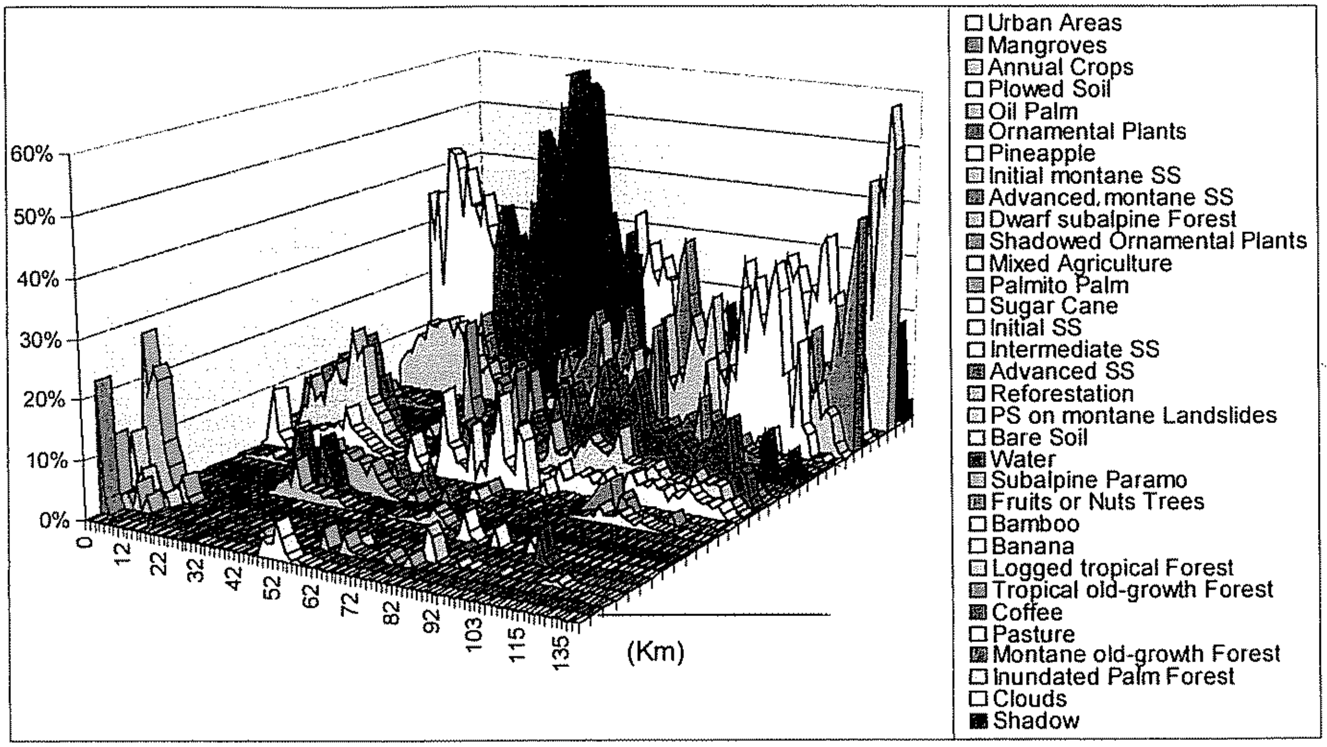


Figure 8. Prior probability assumed in equal prior classifications

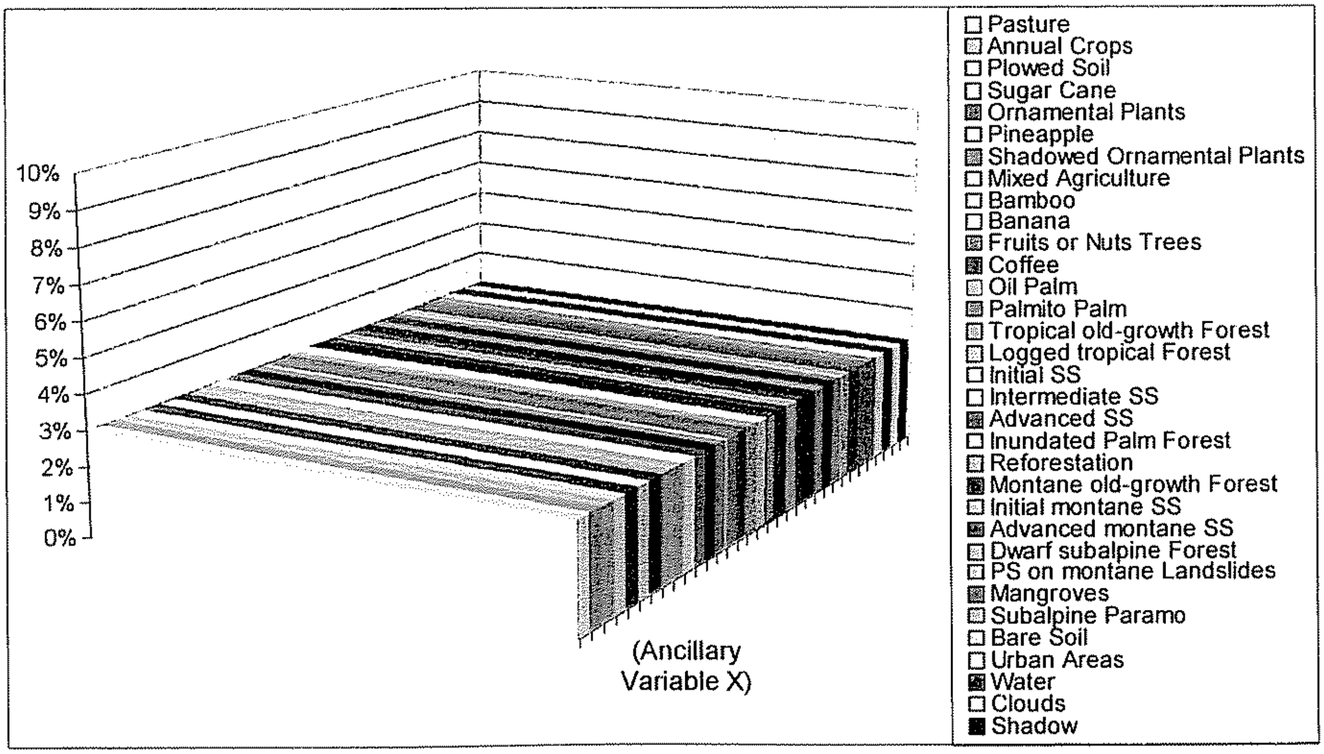
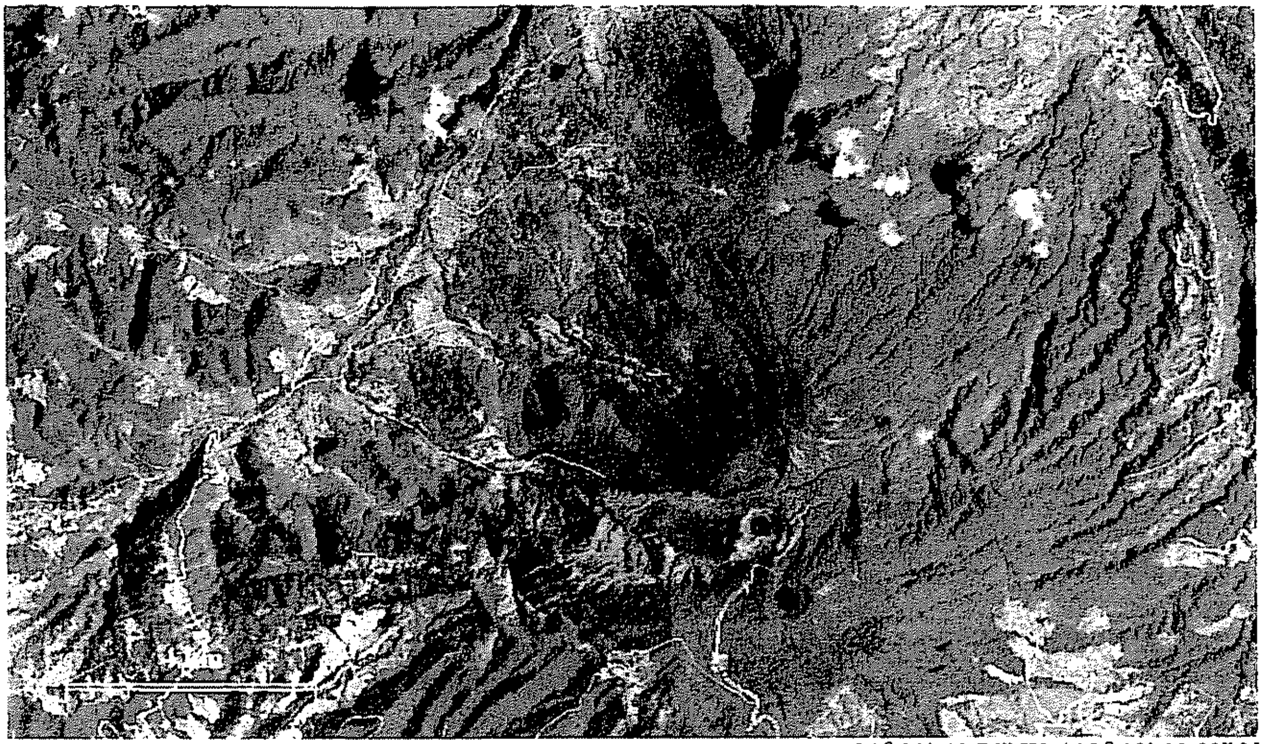
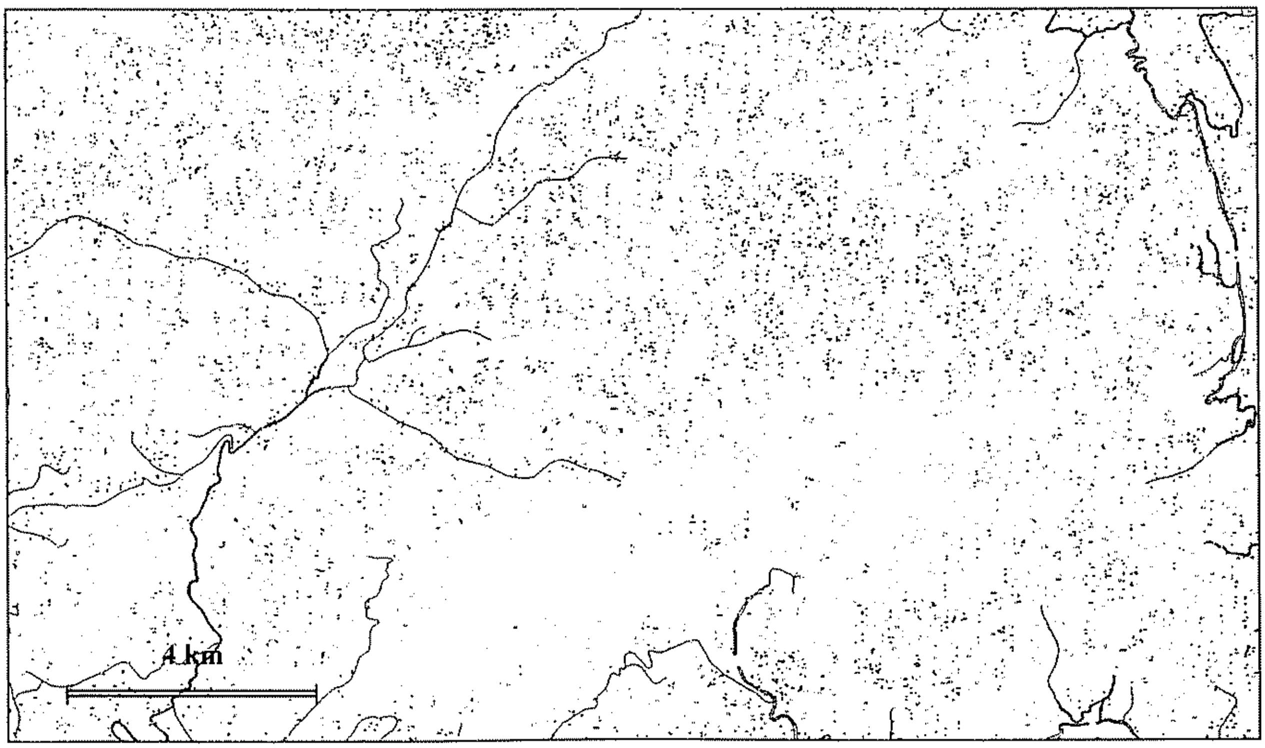


Figure 9. Sample point density (black spots) used to estimate the class prior probabilities (Image subset corresponds to the region of the Poás Volcano, aprox. 240 km^2)

84° 20' 24.58" W / 10° 16' 30.65" N



84° 09' 49.76" W / 10° 10' 12.93" N



4.2 Classification results

The usefulness of the Bayesian classification for improving the discrimination of land cover categories is demonstrated by the classification results (Table 3).

The overall consistency of the classification in the 979 training sites using the Bayesian classification was, with 91.9% (Kappa 0.914), 17.3% higher than using the traditional equal prior maximum likelihood classification. The overall classification accuracy in the 252 control sites was 89.0% and 68.7% respectively. The land cover categories that improved the best were the spectrally similar forest categories.

The classification results presented in Table 3 refer to the band and index combination that resulted in the best discrimination. The results of 18 classification tests using different band and index combinations are presented in Part III. Variations in classification accuracy of these different tests were only a few percentage points. From this it was concluded that use of NDVI and Tasseled Cap indexes was not as effective as the estimation and use of prior probabilities for improving the discrimination of land cover, especially the spectrally similar categories.

The superior quality of the classification using modified prior probabilities could also be appreciated visually. Part II shows two subsets of the study region, in which "salt and pepper" effects, and the classification of land cover categories that could not exist, according to the geographical context, are significantly reduced in comparison to the results obtained with the traditional classification method.

4.3 Spectral enhancement using NDVI and Tasseled Cap

As discussed in depth in Part III, the use of vegetation indexes did not result in important changes of the classification accuracy. To test the usefulness of vegetation indexes for improving the discrimination of different forest categories, spectral observations were selected from field visited sites and then *sensu* Holdridge *et al.*, 1971). At the 95% confidence level, the spectral mean values were statistically different among forest categories within the same ecological life-zone (ANOVA and Tukey test), but were also different within the same forest category between the two ecological life-zones (t-test and Wilcoxon rank sum test). Forest categories were thus spectrally different between and within categories.

However, in all spectral bands and indexes, the mean spectral values of individual forest categories were found to be within the standard deviation range of the mean values of the other forest categories. While the mean values of the forest categories were different, spectral overlap was important, thus suggesting little spectral dissimilarity, especially between advanced secondary successions, logged forests, and old-growth forests.

The use of NDVI and Tasseled Cap indexes was of limited usefulness for increasing the spectral separation between the forest categories. NDVI values were only useful for separating forests categories from other land cover categories such as pastures. Tasseled Cap values showed some difference among categories, but they had also greater variances than NDVI.

The spectral patterns of the forest categories of interest are analyzed in depth in part III.

Table 3. Percentage of overall classification consistency and accuracy
(Band combination: TM1, TM2, TM3, TM4, TM5, TM7, NDVI, Brightness, Wetness)

	Classification Consistency in the Training Sites			Classification Accuracy of Independently Controlled Sites		
Land Cover Category	*n*	*equal priors*	*mod. priors*	*n*	*equal priors*	*mod. priors*
Pasture	27970	90.4	96.9	1763	86.3	95.0
Annual Crops	2592	88.8	97.1	0	---	---
Plowed Soil	2997	93.9	99.0	0	---	---
Sugar Cane	7624	94.2	98.8	482	67.6	82.8
Ornamental Plants	715	70.1	96.6	0		
Pineapple	1033	99.8	100.0	297	98.3	98.3
Shadowed Ornamental Plants	905	98.3	100.0	236	94.9	95.8
Mixed Agriculture	543	25.1	55.9	0	---	---
Bamboo	3322	65.9	93.5	0	---	---
Banana	22841	81.9	97.9	355	84.5	96.3
Fruits and Nuts Trees	4383	61.9	85.8	246	50.8	69.9
Coffee	4890	77.6	95.4	528	79.9	97.2
Oil Palm	3868	69.0	94.5	170	52.4	88.2
Palmito Palm	2680	79.6	90.1	0	---	---
Tropical Old-growth Forest	7819	45.4	82.0	1706	34.6	93.8
Logged Tropical Forest	8347	27.5	71.6	285	26.7	55.8
Initial SS[(1)]	1694	33.0	66.0	0	---	---
Intermediate SS[(1)]	3673	29.8	66.0	0	---	---
Advanced SS[(1)]	7514	31.1	68.6	296	57.8	77.4
Inundated Palm Forest	7201	64.2	96.0	352	85.2	97.7
Reforestation	7108	55.8	87.0	412	40.5	55.6
Montane Old-growth Forest	8109	83.7	99.1	2115	71.6	95.4
Initial Montane SS[(2)]	589	55.5	84.7	494	15.8	58.9
Advanced Montane SS[(2)]	669	50.8	63.8	618	30.3	73.9
Dwarf Subalpine Forest	267	67.7	67.5	444	11.9	48.2
PS[(3)]on Montane Landslides	82	89.2	98.8	0	---	---
Mangroves	2513	93.0	99.2	490	95.1	95.9
Subalpine Paramo	1920	99.1	99.8	1782	86.9	93.0
Bare Soil	1821	97.4	99.9	160	70.0	93.8
Urban Areas	742	96.6	100.0	160	48.8	84.4
Water	10426	99.9	100.0	754	99.9	100.0
Clouds	3589	100.0	100.0	963	100.0	100.0
Shadow	4020	99.7	99.9	98	89.8	98.9
Overall Accuracy		74.6	91.9		68.7	89.0
Kappa		0.73	0.91		0.66	0.88

(1) Secondary Succession of tropical lowland forests after Finegan's model (1996)
(2) Secondary Succession of montane forests after Kappelle's model (1995)
(3) Primary succession

5. DISCUSSION

5.1 Advantages and drawbacks of the methods used

The methods used in this research achieved a much higher level of discrimination of tropical forest categories than conventional methods. The estimation of georeferenced prior probabilities of land cover categories was achieved using a computer modeling approach (instead of a random sampling approach) for a large and complex tropical region and then used in a Bayesian classification of spectral data. The research thus contributed to solve two critical problems: the estimation of prior probabilities in large and complex areas, and the discrimination of tropical forest categories.

However, compared to other classification methods of Landsat TM data, the methods used in this research required a considerable amount of additional data, skills, and computer processing time. While obtaining additional data is only a problem when GIS data of good quality are not available, computer processing time and skills must be acquired.

Bayesian classifiers might also be very sensitive to variations in the prior probability estimates. Classification results will thus depend on the quality of the prior probability model. This would make the availability of a validation data set to test the quality of the prior probability model of advantage. However, obtaining a validation data set has the same practical problems than the estimation of prior probabilities using random sampling techniques. As a consequence, it will rarely be possible to quantitatively assess the quality of the prior probability model, and only the classification output will provide information about its quality and usefulness. Part II includes an in depth discussion of the classification with modified prior probabilities.

5.2 Usefulness of spectral enhancement for improving the discrimination

The results obtained from spectral patterns analysis and from 18 classification tests suggest that spectral data enhancement techniques are of little advantage for the study of regions as complex as that chosen for this research. The use of such techniques implicitly assumes that spectral patterns observed at the patch level are repeated over the entire scene. Instead, the spectral response observed at the patch level might vary, within the same land cover category, in space, and not only in time. This has also been observed in other remote sensing research of secondary successions (Mausel *et al.*, 1993; Foody *et al.*, 1996). In this research, spectral variations within forest categories were expected, because during fieldwork important differences in species composition, biomass, and site moisture were observed for forest patches belonging to the same category. Such variations are also reported in the ecological literature, since species composition, biomass accumulation rates, and time-frame of secondary successions have been shown to vary depending on initial site conditions and the history of land use (Uhl, 1987; Uhl *et al.*, 1988; Foody *et al.*, 1996). A more in depth discussion of these aspects can be found in Parts III and IV of this work.

5.3 Potential use of the classified data

Despite the improvements in classification accuracy obtained with the methods used in this research, the levels of discrimination achieved are insufficient to provide information about location and area of secondary and logged forest at the farm-level. However, unlike other classification techniques, the Bayesian classification using spatially variant sets of prior probabilities minimizes the number of pixels allocated to land cover categories that are unlikely to exist in the geographical context described by the ancillary variables. Within major strata, such as ecological life-zones, the area estimates obtained for the different forest categories are therefore likely to approximate the true values, particularly if corrected for commission and omission errors. Thus, for regional research applications, such as carbon budgeting over large areas, the data classified with the methods used in this research could be useful. Nevertheless, for such applications, attention should be paid to the relationship between the classification scheme and its biophysical meaning.

6. CONCLUSION

Landsat TM data have been shown to be useful for the discrimination of secondary and disturbed forest categories in tropical regions of low relief and reduced forest fragmentation. Important requirements for using them effectively for this purpose were high-density ground data of good quality, classifier operations, and – sometimes – ancillary data. It is at least doubtful that Landsat TM data might be used for the same purpose and with similar levels of success in more complex tropical regions. Rugged topography, ecological diversity, advanced forest fragmentation, and continuous human intervention in the forest, are sources of "spectral noise" that reduce the possibility of making a good discrimination using spectral data alone. In such situations, ancillary data describing the forest site conditions at the pixel level might be used to improve the quality of the discrimination.

The Bayesian classification used in this research appeared to be an effective strategy to include the information provided by ancillary data in the discriminant analysis. This classification technique was more effective for improving the accuracy of the discrimination of tropical forest categories than the use of spectral enhancement techniques in the pre-processing phase. NDVI and Tasseled Cap indexes were useful for the discrimination of broad land cover categories, but little evidence was found to support the hypothesis that they can contribute to more subtle discrimination objectives. The classification test using the Bayesian approach was the only one, out of 18 classification experiments, that produced a significant improvement of the discrimination of tropical forest categories.

If remotely sensed data are to be used to address carbon and biodiversity issues, ecological and geographical modeling of the pixel's context appears to be a necessary complement of spectral data analysis and ground data collection. This will certainly increases analysis costs. However, the computer modeling approach used in this research proved to be a viable alternative to generate the information about prior probability required for the Bayesian classification. Its use in large and inaccessible tropical regions should be less expensive, and probably equally accurate, than the traditional random sampling approach.

Classification schemes based on qualitative descriptions harbor the danger of not being mutually exclusive and exhaustive in terms of biophysical properties that are now important to assess for global change research. On the other hand, the use of quantitative classification schemes does not ensure that the classified categories are quantitatively different in terms of the biophysical properties used to define the categories. Further multidisciplinary research thus required for improving our understanding of the biophysical properties that explain the spectral response of tropical forests.

As shown in this work, the most commonly used classification techniques appears unable to exploit the whole information content of existing data. More research is required to develop spatial-spectral algorithms and classification techniques that are capable of identifying Earth features by their spectral response, shape, size, geographical context, and spatial relationship with other neighboring features.

Finally, an important question that should be investigated, is how classification errors of commission and omission are propagated when output data from remote sensing research are put in other research applications or scaled up and down.

REFERENCES CITED

Apps, M. and D. T. Price (eds.), 1996. Forest Management and the Global Carbon Cycle. NATO ASI Series, Series I: Global Environmental Change, Vol. 40, Springer Verlag, Berlin, 452 p.

Brondizio, E., E. Moran, P. Mausel, and Y. Wu, 1996. Land cover in the Amazon estuary: linking of Thematic Mapper with botanical and historical data. Photogrammetric Engineering and Remote Sensing, 62:921-929.

Brown, S., 1996. Tropical forests and the global carbon cycle: estimating state and change in biomass density. In: Apps, M. and D. T. Price (eds.), 1996. Forest Management and the Global Carbon Cycle. NATO ASI Series, Series I: Global Environmental Change, Vol. 40, Springer Verlag, Berlin.

Brown, S., and A. E. Lugo, 1990. Tropical secondary forests. Journal of Tropical Ecology, 6:1-31.

Brown, S., and L. R. Iverson, 1992. Biomass estimates for tropical forests. World Resource Review, 4:366-384.

Brown, J. F., T. R. Loveland, J. W. Merchant, B. C. Reed, and D. O. Ohlen, 1993. Using multisource data in global land-cover characterization: concepts, requirements and methods. Photogrammetric Engineering and Remote Sensing, 59:977-987.

Campbell, J. B., and J. O. Bowder, 1992. SPOT survey of agricultural land uses in the Brazilian Amazon. International Arch. Photogrammetry Remote Sensing, Commission VII, 29:159-165.

Cibula, W. G., and M. O. Nyquist, 1987. Use of topographic and climatological models in a geographical database to improve Landsat MSS classification for Olympic National Park. Photogrammetric Engineering and Remote Sensing, 54:587-592.

Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 37:34-46.

Crist, E. P., and R. C. Cicone, 1994. Application of the tasseled cap concept to simulated Thematic Mapper data. Photogrammetric Engineering and Remote Sensing, 50:343-352.

Dirzo, R., and M. C. Garcia, 1992. Rates of deforestation in Los Tuxtlas, a neotropical area in southeast Mexico. Conservation Biology, 6:84-90.

Downton, M. W., 1995. Measuring tropical deforestation: development of the methods. Environmental Conservation, 22:229-240.

Fearnside, P. M., 1996. Amazonian deforestation and global warming: carbon stocks in vegetation replacing Brazil's Amazon forest. Forest Ecology and Management, 80:21-34.

Fearnside, P. M., and W. M. Guimaraes, 1996. Carbon uptake by secondary forests in Brazilian Amazonia. Forest ecology and Management, 80:35-46.

Finegan, B., 1992. El potencial de manejo de los bosques húmedos secundarios neotropicales de las tierras bajas. Colleción Silvicultura y Manejo de Bosques Naturales, No. 5. Tropical Agricultural Research and Higher Education Center (CATIE), Turrialba, Costa Rica.

Finegan, B., 1996. Pattern and process in neotropical secondary rain forests: the first 100 years of sucession. Tree, 11:119-124.

Flack, J., November 1995, *"Interpretation of Remotelñy Sensed Data Using Guided Techniques"* http://www.per.dem.csiro.au/staff/FlackJulien/thesis/node2.html (2-February-2000).

Flint E. R., and J. F. Richards, 1994. Trends in carbon content of vegetation in south and southeast Asia associated with changes in land-use. pp 185-190. In: Dale H. V. (ed.), Effects of Land-Use Change on Atmospheric CO2 Concentrations: South and Southeast Asia as a Case Study. Ecological Studies: Analysis and Synthesis, vol. 101, Springer Verlag, New York.

Foody, G. M., N. A. Campbell, N. M. Trodd, and T. F. Wood, 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. Photogrammetric Engineering and Remote Sensing, 58: 1335-1341.

Foody, G. M. and P. J. Curran, 1994. Estimation of tropical forest extent and regenerative state using remotely sensed data. Journal of Biogeography, 21:223-244.

Foody, G. M., G. Palubinskas, R. M. Lucas, P. J. Curran, M. and Honzak, 1996. Identifying terrestrial carbon sinks: classification of successional stages in regenerating tropical forest from Landsat TM data. Remote Sensing of Environment, 55:205-216.

Franklin, S. E., and D. R. Peddle, 1989. Spectral texture for improved class discrimination in complex terrain. International Journal of Remote Sensing, 10:1437-1443.

Franklin, S. E., and B. W. Wilson, 1992. A three-stage classifier fore remote sensing of mountain environments. Photogrammetric Engineering and Remote Sensing, 58:449-454.

Fuentes, A. M., C. L. Guevara, and N. C. Zamora, 1995. Fijación de dióxido de carbono: una alternativa para contrarrestar el efecto invernadero. Seminario de la Universidad Nacional, Facultad de Ciencias de la Tierra y el Mar, Escuela de Ciencias Ambientales, Heredia, Costa Rica.

Fung, I., 1996. The global carbon cycle and the atmospheric record: "the problem definition". pp. 25-34. In: Apps, M. and D. T. Price (eds.), 1996. Forest Management and the Global Carbon Cycle. NATO ASI Series, Series I: Global Environmental Change, Vol. 40, Springer Verlag, Berlin.

Gallo, M., 1999. Identificación de Tipos de Bosques Primarios en la Zona Norte de Costa Rica. MSc Thesis (Turrialba, Costa Rica: Tropical Agricultural Research and Higher Education Center).

Gilruth, P. T., C. F. Hutchinson, and B. Barry, 1990. Assessing deforestation in the Guinea highlands of West Africa using remote-sensing. Photogrammetric Engineering and Remote Sensing, 56:1375-1382.

Gilruth, P. T., S. E. Marsh, and R. Itami, 1995. A dynamic spatial model of shifting cultivation in the highlands of Guinea, West Africa. Ecological Modelling, 70:179-197.

Grainger, A., 1993. Rates of deforestation in the humid tropics: estimates and measurements. The Geographical Journal, 159:33-44.

Green, G. M., and R. Sussman, 1990. Deforestation history of the eastern rainforests of Madagascar from satellite images. Science, 248:212-215.

Hall, F. G., D. B. Botkin, D. E. Strebel, K. D. Woods, and S. J. Goetz, 1991. Large-scale patterns of forest succession as determined by remote-sensing. Ecology, 72:628-640.
Harper, J. L., 1977. Population Biology of Plants. Academic Press, New York.
Helmer, E. H., 1999. The Landscape Ecology of Secondary Tropical Forest in Montane Costa Rica. PhD Dissertation, Oregon State University, Oregon.
Hill, R. A., G. M. and Foody, 1994. Separability of tropical rain-forests types in the Tambopata-Candamo reserved zone, Peru. International Journal of Remote Sensing, 15: 2687-2693.
Holdridge, L. R., W. C. Grenke, W. H. Hatheway, T. Lang, and J. A. J. Tosi, 1971. Forest Environments in Tropical Life Zones: a Pilot Study. Pergamont Press, Oxford.
Houghton, R. A., and J. L. Hackler, 1994. The net flux of carbon from deforestation and degradation in south and southeast Asia. pp 301-328. In: Dale V. H. (ed.) Effects of Land-Use Change on Atmospheric CO_2 Concentrations: South and Souteast Asia as a Case Study. Springer-Verñag, New York.
Howe, H. F., and J. Smallwood, 1993. Ecology of seed dispersal. Annual Review of Ecological Systems, 13:201-228.
Hutchinson, C. F., 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. Photogrammetric Engineering and Remote Sensing, 48:123-130.
International Panel on Cliamte Change, 2000. IPCC Special Report: Land Uses, Land Use Change and Forestry. Summary for Policymakers. IPCC Secretariat, c/o WMO, 7bis, Avenue de la Paix, C.P.No 2300, 1211 Geneva 2, Switzerland.
Iverson, L. R., S. Brown, A. Prasad, H. Mitasova, A. J. R. Gillespie, and A. E. Lugo, 1993. Use of GIS for estimating potential and actual forest biomass for continental South and Southeast Asia. Ecological Studies: Analysis and Syntesis, 101:67-116.
Janzen, D. H., 1988. Management of habitat fragments in a tropical dry forest: growth. Annals of the Missouri Botanical Garden, 75:105-116.
Jensen, J. R., 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. 2nd edition, Printice Hall, New Jersey.
Kappelle, M., 1995. Ecology of Mature and Recovering Talamancan Montane *Quercus* Forests, Costa Rica. Ph.D. Dissertation, University of Amsterdam/School of Environmental Sciences, National University, Heredia (Costa Rica).
Kappelle, M., T. Geuze, M. E. Leal, and A. M. Cleef, 1996. Successional age and forest structure in a Costa Rican upper montane *Quercus* forest. Journal of Tropical Ecology, 12:681-698.
Kummer, D. M., and B. L. Turner, 1994. The human causes of deforestation in southeast Asia. BioScience, 44:323-328.
Li, Y, E. F. Moran, E. S. Brondizio, P. Mausel, and Y. Wu, 1994. Discrimination between advanced secondary succession and mature moist forest near Altamira, Brazil using Landsat TM data. Proceedings of the American Society for Photogrammetry and Remote Sensing. Annual meeting of ASPRS in Reno, NV., 350-364.
Ludeke A. K., R. C. Maggio, and L. M. Reid, 1990. An analysis of anthropogenic deforestation using logistic regression and GIS. Journal of Environmental Management, 31:247-257.
Lugo, A. E., and S. Brown, 1992. Tropical forests as sinks of atmospheric carbon. Forest Ecology and Management, 54:239-255.

MacArthur, R. H., and E. O. Wilson, 1967. Theory of Island Biogeography. Princeton University Press, Princeton.

Maselli, F., C. Conese, T. De Filippis, and R. Romani, 1995. Integration of ancillary data into a maximum-likelihood classifier with nonparametric priors. Journal of Photogrammetry and Remote Sensing, 50:2-11.

Mather, P. M., 1985. A computationally-efficient maximum-likelihood classifier employing prior probabilities for remotely-sensed data. International Journal of Remote Sensing, 6:369-376.

Mausel, P., Y. Wu, Y. Li, E. F. Moran, and E. S. Brondizio, E. S., 1993. Spectral identification of successional stages following deforestation in the Amazon. Geocarto International, 4:61-71.

McCracken, S., E. S. Brondizio, D. Nelson, E. F. Moran, A. D. Siqueira, and C. Rodriguez-Pedraza, 1999. Remote sensing and GIS at farm property level: demography and deforestation in the Brazilian Amazon. Photogrammetric Engineering and Remote Sensing, 65:1311-1320.

Moran E. F., E. Brondizio, P. Mausel, and Y. Wu, 1994. Integrating Amazonian vegetation, land-use, and satellite data. BioScience, 44:329-338.

Myers, N. 1988. Tropical deforestation and remote-sensing. Forest Ecology and Management, 23:215-225.

Pedroni, L. and S. Velásquez, 1998. Medición y diagnóstico de cobertura boscosa: ¿Qué tan útiles son las imágenes de satélite en nuestro medio? In: E. M. Flores and E. E. Chaves (Editors), Conservación del Bosque in Costa Rica. Academia Nacional de Ciencias, San José (Costa Rica), pp. 85-99.

Puig, C. J., 1996. Identificación de Fases de Desarrollo de Bosques Secundarios Húmedos Tropicales de Tierras Bajas en Costa Rica, mediante Análisis Digital de Imágenes de Sensores Remotos. MSc Thesis, Tropical Agricultural Research and Higher Education Center (CATIE), Turrialba (Costa Rica).

Sabogal, C., 1992. Regeneration of tropical dry forests in Central America, with examples from Nicaragua. Journal of Vegetation Sciences, 3:407-416.

Sader, S. A., 1995. Spatial characteristics of forest clearing and vegetation regrowth as detected by Landsat Thematic Mapper imagery. Photogrammetric Engineering and Remote Sensing, 61:1145-1151.

Sader, S. A., and A. T. Joyce, 1988. Deforestation rates and trends in Costa Rica: 1940-1983. Biotropica, 20:11-19.

Sader, S. A., R. B. Waide, W. L. Lawrence, and A. T. Joyce, 1989. Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data. Remote Sensing of Environement, 28:143-156.

Sader, S. A., T. Stone, and A. T. Joyce, 1990. Remote sensing of tropical forests: an overview of research and applications using non-photographic sensors. Photogrammetric Engineering and Remote Sensing, 56:1343-1351.

Sohn, Y., E. Moran, and F. Gurri, 1999. Deforestation in North Central Yucatan. Photogrammetric Engineering and Remote Sensing, 65(8):947-958.

Stone T. A., I. F. Brown, and G. M. Woodwell, 1991. Estimation, by remote sensing, of deforestation in central Rodonia, Brazil. Forest Ecology and Management, 38:291-304.

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

Swain, P. H., and S. M. Davis, 1978. Remote Sensing: The Quantitative Approach, McGraw-Hill, New York.

Tuomisto, H., A. Linna, and R. Kalliola, 1995. Use of digitally processed satellite images in studies of tropical rain vegetation. International Journal of Remote Sensing, 15:1595-1610.

Uhl, C., 1987. Factors controlling succession following slash-and-burn agriculture in Amazonia. Journal of Ecology, 75:377-407

Uhl, C., R. Bushbacher, and E. A. S. Serrao, 1988. Abandoned pastures in eastern Amazonia. I. Patterns of plant succession. Journal of Ecology, 76:938-958.

Veldkamp E., A. M. Weitz, I. G. Staritsky, and E. J. Huising, 1992. Deforestation trends in the Atlantic zone of Costa Rica: a case study. Land Degradation and Rehabilitation, 3:71-84.

Westman, W. E., L. L. Strong, and B. A. Wilcox, 1989. Tropical deforestation and species endangerment: the role of remote-sensing. Landscape Ecology, 3:97-109.

Woodwell, G. M., R. A. Houghton, T. Stone, and A. Park, 1986. Changes in the areas of forests in Rodonia, Amazon Basin, measured by satellite imagery. In: J. Trabalka and D. Reichle (Editors), The Changing Carbon Cycle: A Global Analysis. Springer-Verlag, New York, NY, pp. 242-257.

Woodwell, G. M., R. A. Houghton, T. Stone, R. Nelson, and W. Kovalick, 1987. Deforestation in the tropics: new measurements in the Amazon Basin using Landsat and NOAA Advanced Very High Resolution Radiometer imagery. Journal of Geophysical Research, 92:2157-2163.

Part II

Improved Classification of Landsat TM Data Using Modified Prior Probabilities in Large and Complex Landscapes

(Submitted for publication to International Journal of Remote Sensing)

Improved Classification of Landsat TM data using modified prior probabilities in large and complex landscapes

LUCIO PEDRONI
Tropical Agricultural Research and Higher Education Center (CATIE), 7170 Turrialba, Costa Rica
Tel.: (++506) 556-15-30, Fax.: (++506) 556-79-54, E-mail: lpedroni@catie.ac.cr

ABSTRACT

The use of modified prior probabilities to exploit ancillary data and increase classification accuracy has been proposed before. However, this method has not been widely applied because it has heavy computing requirements and because obtaining prior probability estimates has presented practical problems. This article presents a procedure that generates large sets of prior probability estimates from class frequencies modeled with ancillary data and a Mahalanobis Distance selection of previously classified pixels. The method produces a pixel sample size that is large enough to estimate class frequencies in numerous strata, which is particularly desirable for the study of large and complex landscapes.

A case study is presented, in which the procedure made it possible to estimate 537 sets of prior probabilities for an entire Landsat TM scene of central Costa Rica. After modifying the class prior probabilities, the overall classification consistency of the training sites improved from 74.6% to 91.9%, while the overall classification accuracy of sites controlled in the field by independent studies improved from 68.7% to 89.0%. The classification accuracy was most improved for spectrally similar classes. The method improves classification accuracy in large and complex landscapes with spectrally mixed land-cover categories.

1. Introduction

One important advantage of Landsat TM is its 'synoptic view', that makes it possible to study large portions of the Earth's surface at a relatively low cost. Sensor systems of higher spatial and spectral resolution do not provide yet this large area view at a low cost. For this reason, improvement of classification methods of Landsat sensor data should focus on problems faced by large-area projects. Instead, this type of research has often been limited to small image subsets, sometimes comparable in size to that of only a few aerial photographs. As a consequence, innovative classification methods that produce more accurate results, but require more software, hardware, or ground data, have rarely been applied to large-areas. The large area view is now particularly important to understand and address global change (Detwiler and Hall 1988, Houghton 1991, Dixon *et al.* 1994, UNEP 1995). As an example, several countries have committed themselves to realize national inventories of green house gases and need therefore data on land-use, land-use change, and forestry (LULUCF) at the national level. Improved classification of land cover at national and supra-national scale, and better understanding of the driving forces underlying land cover and land use change (Harrison 1991, Skole *et al.* 1994, Kummer and Turner 1994), will both become essential to monitor leakage of carbon dioxide emissions of

LULUCF activities under the Kyoto Protocol. Land cover data of large areas are also essential to assess biological diversity at the ecosystem level (Sader *et al.* 1991, Lewis 1994, Caicco *et al.* 1995, Tuomisto *et al.* 1995, Nagendra and Gadgil 1999). The rates of tropical deforestation and forest regeneration, however, are still uncertain (Kleinn *et al.* 2001), as are the methods by which to accurately measure them (Myers 1988, Skole and Tucker 1993, Downton 1995). The remote sensing community is thus asked not only to refine traditional land-cover measurement, but also to differentiate ecological, functional, structural, and compositional features on the Earth's surface. This implies the need to discriminate vegetation categories of different biomass content and taxonomic composition, such as secondary, logged, and undisturbed old-growth forests. Some of these land-cover categories have overlapping spectral signatures and are therefore difficult to discriminate using remotely sensed data alone. Studies that address these problems are relatively recent and few (Sader *et al.* 1989, Mausel *et al.* 1993, Foody and Curran 1996, Moran *et al.* 1996, Helmer 1999), and generally utilize an amount and density of ground data that is not available for large-area projects.

Among data analysis methods, the supervised maximum likelihood classification procedure of just the spectral data is still the most widely used classification strategy in digital processing of Landsat sensor data (Booth and Oldfield 1989, Foody *et al.* 1992, Maselli *et al.* 1995). An attractive feature of the maximum likelihood classifier is that it can be modified (Jensen 1996), which has led to the investigation of multiple variations of the procedure. With the steadily growing availability of digital geographic data and the increased speed of today's computers, some of these modified procedures are now more applicable. This offers new opportunities for solving classification problems in large-area projects.

This article presents a classification procedure, that makes it possible to integrate spatial data in the classification process through the strategy of modifying the class prior probabilities. This strategy was proposed by Swain and Davis (1978), Strahler (1980), and Hutchinson (1982), and has been successfully implemented by Mather (1985) and Maselli *et al.* (1995) in small areas. However, the procedure that will be discussed here extends the possibilities of the approach proposed by previous investigations. A strategy is shown, that produces estimates of large sets of prior probabilities, without necessarily requiring an expensive stratified random sampling of the training data. This does not imply, that such a sampling procedure would be the most desirable technique in all situations where it can be applied. Nor does it mean that no ground data are required to develop the training statistics and to evaluate the classification accuracy. However, the proposed method addresses a number of problems, from which the most relevant is probably that of an economically feasible estimation of class prior probabilities in large-area applications.

2. Previous work

Among the most frequent experiments designed to improve the maximum likelihood classification performance were computations of band ratios and vegetation indexes for the enhancement of some spectral features present in the remotely sensed data set (Sader *et al.* 1989, Boyd *et al.* 1996, Helmer 1999). Important efforts were also made to incorporate image texture measurements into the analysis (Jensen and Toll 1982, Gong and Howarth 1990, Joen and Landgrebe 1992) or to reduce texture variations before classification (Hill and Foody 1994). Results of these investigations were mixed (Sader *et al.* 1989, Foody and Curran 1994).

Since the spatial distribution of most land-cover categories has a strong relationship with environmental variables, some researchers have investigated the possibility of including data about these variables into the classification process (Strahler 1980, Hutchinson 1982, Mather 1985, Cibula and Nyquist 1987, Skidmore and Turner 1988, Maselli *et al.* 1995). These strategies were generally more successful than the manipulation of spectral data in the pre-classification phase. However, some of them violate the normal distribution assumption of the data (Hutchinson 1982, Flack 1995, Jensen 1996), and other result in artificially sharp boundaries in the classified output data set (Maselli *et al.* 1995).

The more successful procedure for incorporating ancillary data into the maximum likelihood classification has been the use of modified prior probabilities. This method, well described by Strahler (1980) and Mather (1985), is independent from the data distribution of ancillary variables, does not create sharp boundaries in the output data set, and usually results in greatly improved classification accuracy (Strahler 1980, Mather 1985, Skidmore and Turner 1988, Maselli *et al.* 1995).

The method consists of producing estimates of the expected class frequencies in the classified image, and using them to modify the prior probability of each class in the maximum likelihood decision rule. When a higher prior probability is assigned to a particular class, the decision boundary of the discriminant function shifts away from the mean of that spectral class, thus allowing a greater number of pixels to be classified to that class. The result is that classes that are more likely to appear in the scene, or in a subset of it, tend to classify more pixels, which in turn increases the chance of producing an accurate classification. The different class prior probabilities, also called 'weights' or 'priors', become decisive for assigning a pixel to a class, only when the spectral information is insufficient to make the discrimination from other classes, that is to say, at pixel brightness values where signatures are largely overlapping (Maselli *et al.* 1995). The priors do not affect the classification decision at pixel brightness values where the spectral signatures are widely separate (Mather 1985).

If a separate set of prior probabilities can be estimated for all image strata that have been defined through ancillary data analysis, the information content of these data is transposed to the maximum likelihood decision rule probabilistically. As pointed out by Maselli *et al.* (1995), this particular feature of the method is 'useful to avoid introducing artifacts from nominal scale variables'.

Strahler (1980) and Maselli *et al.* (1995) concede that one of the procedure's major drawbacks is that a separate set of prior probabilities must be estimated for each stratum. The class frequencies of the training sites can be used to estimate the priors, provided they were taken randomly (Maselli *et al.* 1992, Conese *et al.* 1993). However, stratified random sampling involving field surveys and aerial photograph interpretation does not necessarily guarantee that all classes will be sampled in all the strata in which they actually occur. To avoid artificially excluding a class, just because it is not present in the frequency sample set, Strahler (1980), Hutchinson (1982), and Maselli *et al.* (1995) suggest assigning a 'non-completely excludable ancillary prior' (Maselli *et al.* 1995), a low probability threshold, to all classes. Strahler (1980) also suggests modeling the probabilities from a much smaller set, assuming there is no high-level interaction and using statistical techniques to calculate expected values for a multidimensional contingency table for which only certain marginal totals are known. To limit the number of strata to be sampled, Maselli *et al.* (1995) suggest taking into consideration only the ancillary variables that are more explicative of the class frequencies. For this purpose, they used Mutual Information Analysis, a non-parametric technique described by Davis and Dozier (1990), and Conese and Maselli (1993), that can be helpful to identify the data layers that are most important to the stratification process.

The work of Strahler (1980), Mather (1985), and Maselli *et al.* (1995) demonstrated that the method of modifying the class prior probabilities has the potential to improve classification accuracy. However, due to increased computing and sampling requirements, their methods have not found the widespread application they deserve.

If a classification procedure requires software that is not commercially available, as well as an intensive stratified random sampling of the landscape, it will not even be considered for use in large area applications because of the heavy investment it implies. This fact has often been overlooked by researchers developing new classification methods. Because large sets of image data are not necessary when developing new classification methods, most studies have been conducted over small areas (Strahler 1980: 220 km^2, Mather 1985: 45 03 km^2, and Maselli *et al.* 1995: 40.96 km^2). The advantage being, that small data files have to be processed, high-density ground data can be sampled, and research cost remains low. For projects involving the classification of one or several Landsat scenes, each encompassing more than 30 000 km^2, only a small fraction of the ground data density typically presented in scientific studies can be gathered. Random sampling, and separate training and control sites are often sacrificed because of time and budget restrictions.

Before discussing the theoretical background of a cost-effective procedure to estimate the class prior probabilities in large and complex areas, it is necessary to briefly review the maximum likelihood classification with modified prior probabilities.

3. Maximum likelihood classification with modified prior probabilities

The maximum likelihood decision rule is based on a normalized (Gaussian) estimate of the probability density function of each class. The probability density function for a pixel x_k can be expressed as (Foody *et al.* 1992):

$$p(x_k \mid i) = \frac{e^{[-1/2(x_k - u_i)' V_i^{-1}(x_k - u_i)]}}{[(2\pi)^{n/2} \mid V_i \mid^{1/2}]} \tag{1}$$

where $p(x_k \mid i)$ is the probability density function for a pixel x_k to be member of class i, n is the number of channels present in the image, x_k is the data vector for the pixel in all channels, u_i is the mean vector for class i over all pixels, and V_i is the variance-covariance matrix for class i. The maximum likelihood decision rule simply assigns the pixel x_k to the class for which the right hand expression of equation (1) results in the greatest probability value. In practice, classification algorithms use a logarithmic form of the maximum likelihood decision rule in which all constants are eliminated. Following the mathematical manipulations shown in Strahler (1980), the maximum likelihood decision rule can be expressed by the following discriminant function $F_{1,k}(x_k)$:

$$F_{1,k}(x_k) = \ln \mid V_i \mid + (x_k - u_i)' V_i^{-1}(x_k - u_i) \tag{2}$$

The pixel x_k is assigned to the class for which the discriminant function results in the lowest value.

The maximum likelihood decision rule can be modified to take into account the prior probabilities of each class, which are simply the expected area proportions of the classes in a particular scene or stratum. Their incorporation into the maximum likelihood decision rule occurs via manipulation of the Law of Conditional Probabilities (Strahler 1980). Mathematically, this occurs by adding the term $-2\ln P_i$ to equation (2):

$$F_{2,k}(x_k) = \ln|V_i| + (x_k - u_i)'V_i^{-1}(x_k-u_i) - 2\ln P_i \quad (3)$$

where P_i is the prior probability of class i and $F_{2,k}(x_k)$ is the discriminant function that takes into account the class prior probabilities. The sum of the P_i of all classes must be 1.0 for each pixel, because each pixel must be assigned to a class.

4. Stratified sampling with the Mahalanobis Distance

Most land-cover categories have typical distributions along ecological and socioeconomic gradients such as elevation, precipitation, soil types, and access conditions, among others (Veldkamp *et al.* 1991, Franklin and Wilson 1992) The assumption that the class prior probabilities are constant along these gradients does not hold, but is always made by the maximum likelihood algorithm, even if the priors are modified without a scene stratification. Therefore, before sampling the landscape to estimate the class prior probabilities, it is advisable to stratify the scene. The goal is to identify landscape units in which the range of ecological and human induced conditions is small enough to maintain approximately constant class frequency – or prior probability – along the gradients still present within each of them.

Once these units of homogeneous context have been identified, a sample procedure is required to estimate the class prior probabilities within each stratum. Because the area of a particular land-cover category can be estimated through random sampling, the frequency of randomly selected training sites can be used as an estimate of class prior probability (Maselli *et al.* 1992). However, when numerous strata are present in the study region, access is difficult, or the area to be surveyed is very large, the classical procedure of random sampling in the field can not be applied. The alternative method suggested here is to sample the class frequencies using a subset of previously classified pixels. The problem is then reduced to the selection of a subset of pixels with minimum spatial bias and maximum likelihood of correct class allocation.

As discussed by Foody *et al.* (1992), two measures can potentially be used to evaluate the quality of the classification at the pixel level: the *a posteriori* probability and the Mahalanobis Distance (MD). The *a posteriori* probability of a pixel x_k to belong to class i, $L(i|x_k)$, can be determined as follows (Foody *et al.* 1992):

$$L(i|x_k) = \frac{P_i(x_k|i)}{\sum_{r=1}^{t} P_r\, p(x_k|r)} \quad (4)$$

where i is the class number, t is the total number of classes, and P_i is the *a priori* probability of membership in class i.

The term:

$$MD = (x_k - u_i)'V_i^{-1}(x_k - u_i) \tag{5}$$

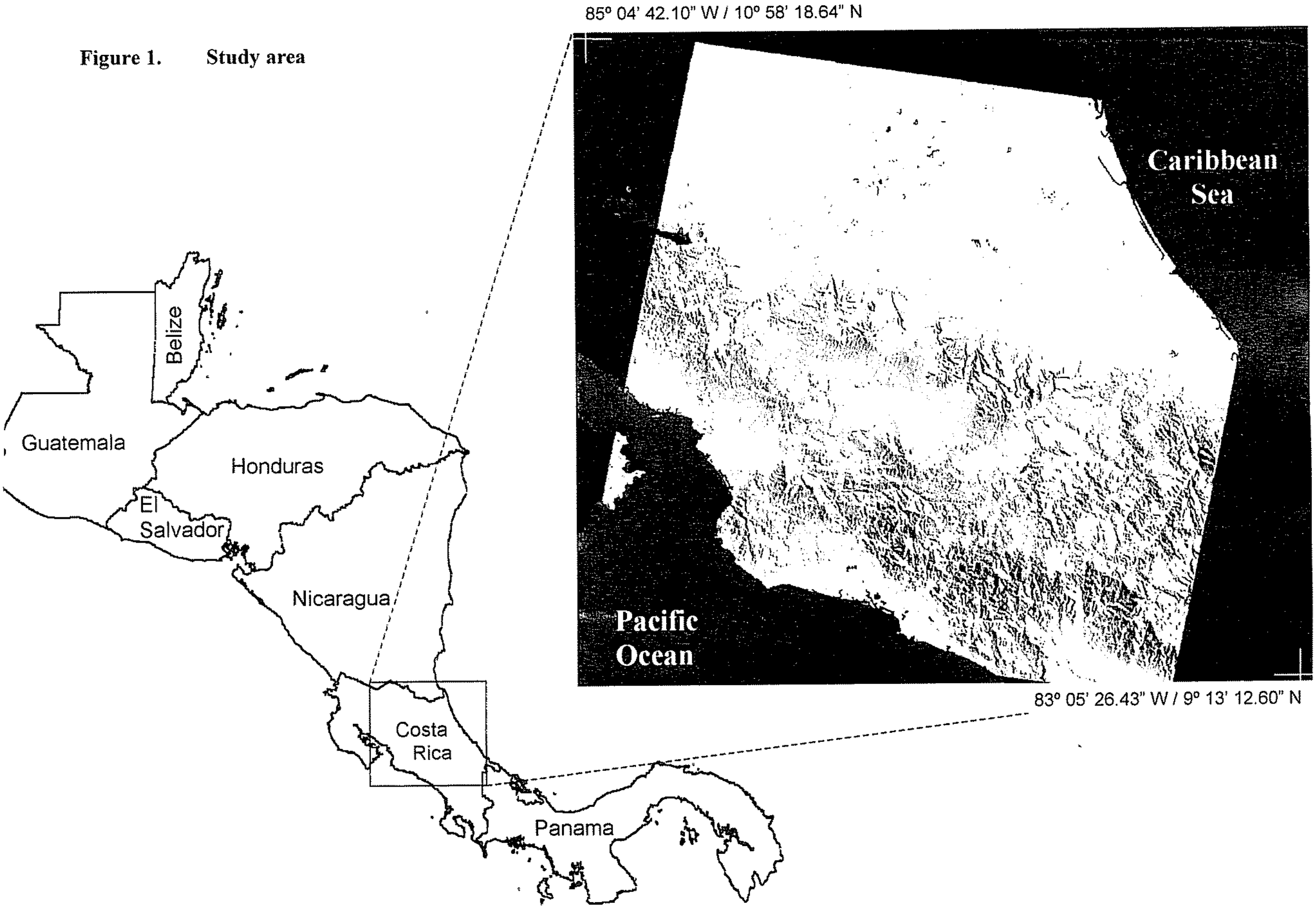
of equation (1) is the Mahalanobis Distance between the pixel x_k and the mean of class i, which is a measure of how spectrally typical the pixel is for that class (Foody *et al.* 1992).

The *a posteriori* probability indicates the relative probability that a pixel belongs to the class to which it was assigned, while the MD indicates whether the spectral signature of the pixel is actually similar to the class spectral signature.

While misclassified pixels are always present in a classified output data set, it can be assumed that their number will decrease at greater *a posteriori* probability and shorter MD values. Provided the training data set is representative of the spectral patterns of the land-cover categories of interest, a threshold of pixels with short MD values or high *a posteriori* probability can therefore provide a better estimation of the spatial distribution and frequency of these categories than the whole set of classified pixels. The MD measure is better suited to this purpose than the *a posteriori* probability measure, because it represents a spectrally related measurement of classification confidence. The risk of counting pixels that are spectrally unrelated to the class to which they were assigned is therefore minimized. Particularly in large-area projects with low training site density, there is a good chance that a specific land-cover category will not be well represented in the signature training data set or that individual pixels will have spectral patterns that are atypical for their category. This makes it possible for pixels with high *a posteriori* probability to belong to a different class from that to which they were assigned.

An advantage of using a MD threshold for the estimation of class prior probabilities is that even for high confidence levels, such as 95%, it is still possible to select a much larger number of pixels than what could possibly be obtained through stratified random sampling in the field. However, unlike the frequency estimates obtained from field and aerial photograph sampling, it must be assumed, that the class allocation is not correct for an unknown proportion of the pixels selected with the MD threshold criterion. Even high spectral typicality does not completely eliminate the possibility of wrong class allocation, especially for closely related spectral classes. This can bias the frequency and therefore the prior estimates obtained from the MD sampling technique. Also, a class may not appear in the sample of a particular stratum, as happens with samples taken from training sites and photography interpretation, thus giving the false impression that the class is not present in that particular stratum. However, because of the large sample size, this situation is less likely to occur with the MD sampling technique, than with the comparatively smaller sample size that can be obtained from training data and photo-interpretation. These two problems, wrong class allocation and absence of a class from a sample, can be minimized with further computer processing of the pixels selected using the MD threshold criterion. As will be shown below, a set of ecological, economical, and geographical criteria can be applied to the selected sample data set in order to eliminate those pixels that are unlikely to have a correct class label. The same set of criteria can then be applied to each stratum to create low prior probability thresholds for all classes that can theoretically occur in it. This follows the suggestion about 'non completely excludable ancillary priors' made in the aforementioned studies (Strahler 1980, Hutchinson 1982, Maselli *et al.* 1995).

Figure 1. Study area



5. Case study

5.1 Study area

The study area encompasses the entire Landsat TM scene (path 15 and row 53) of the central portion of Costa Rica (figure 1). The scene covers approximately 30 950 km^2, of which 5433 km^2 (17.5%) is covered by water and the rest is land with elevations ranging from 0 to 3825 m.a.s.l. and with considerable variation in slope and aspect. The image is from March 1996 (dry season) and was taken on a day with relatively low cloud cover for this part of the world (approximately 13.2%). The ecological complexity of the area includes 12 life-zones and 11 transition-zones (*sensu* Holdridge *et al.* 1971) according to the Ecological Map of Costa Rica (Bolaños and Watson 1993). Due to the influence of trade winds and the presence of the Central Volcanic and Talamanca mountain ranges, climatic conditions are extremely variable, with yearly average precipitation ranging from 1400 mm yr^{-1} to more than 7000 mm yr^{-1} (IMN 1987). One consequence of the climatic variability within the scene is that some vegetation is brown in the Pacific Slope during the dry season, while most of it remains green in the Atlantic Zone and in the mountains. The small to medium- sized land tenure, the fragmented patterns of forest cover, the frequent mixed forms of land-use, the sometimes extremely rugged terrain, and the presence of clouds and haze add considerable spectral complexity to the scene.

5.2 Data processing

The Landsat TM image was georeferenced to the coordinate system Lambert Conformal Conic of North Costa Rica. The thermal band was eliminated from the data set, because of its lower spatial resolution and contrast. The Normalized Difference Vegetation Index (NDVI), and the brightness and greenness indexes of the Tasseled Cap transformation were added to the reflective bands, because it was found that they helped to improve the discrimination of some cover types.

During the period from March, 1998 to March, 1999, 826 sites were visited in the field and their coordinates collected with a non-differential GPS unit (Garmin 12 XL) operating in average mode for approximately 5 minutes. The coordinates of a majority of the field sites visited were previously located on the screen to make sure that all spectral patterns visually present in the scene were appropriately represented in the training data. The structure and floristic composition of the vegetation in each site were briefly described to ensure correct class labeling of the different forest types. The pixels with field descriptions and additional visually interpreted image portions were then used to extract spectral signatures for the cover types of interest. Sites that were thought to have changed before they were surveyed in the field, and spectrally confused signatures were eliminated from the training data set, so that in the end, 979 spectral signatures were maintained for use in classifying the scene.

To stratify the study area, three ancillary variables were considered: elevation, time of access, and a discrete model of distances from the Pacific Coast and export banana fields. The choice of ancillary variables was determined by the availability and accuracy of digital data for the study region.

Elevation data were obtained from a digital elevation model of the study area generated using the ARC/INFO TOPOGRID contour line data interpolation algorithm. The digital contour line data were provided by the Costa Rican National Geographic Institute (IGN), that generated them from scanned 1 : 50 000 scale maps. Because these data were not available for all portions of the image, the contour line data set was complemented with 1 : 200 000 scale contour lines from CATIE's GIS laboratory and additional 1 : 50 000 scale contour line data provided by the Department of Geography at the University of Costa Rica (UCR). Elevation data were thought to be of great ecological significance in the spatial distribution of cover types.

It was assumed that accessibility is the most important factor affecting human- induced spatial distribution of crops, pasture, undisturbed, logged, and secondary forests. However, because difficulty of access is not only a matter of the Euclidian distance to roads, a 'time of access' model was created, using slope data derived from the digital elevation model and data on distance from roads derived from digital road data. To estimate the time required to cross pixels of different slopes, an empirical regression model was estimated using walking time data obtained from interviews with local farmers and foresters. The digital road coverage was obtained from 1 : 50 000 scale maps scanned by the IGN and complemented with screen digitized lines over the satellite image. This was necessary, because the cartographic base was quite outdated for certain portions of the study area. Information about protected areas was included in the 'time of access' model by increasing the time of access by 15 minutes at the borders of protected areas where tracts of undisturbed forests are known to exist. It was therefore assumed that if a piece of land lay within 15-minutes walking distance, people would be willing to walk in to conduct logging, agricultural, or cattle raising activities. However, if no such land is accessible within this range, then people would not hesitate to trespass on legally protected areas to farm or log.

Because of their shape, export banana fields in the Atlantic Zone are visually easy to identify on the satellite image. However, spectrally they tend to blend in with oil palm fields (Pacific Coast), and swampy palm forests (Atlantic Zone). Mangroves, some types of crops and other land-cover categories also occur at specific distance ranges from the Pacific Coast. During the dry season, humidity and precipitation levels increase as one moves from the Pacific Coast toward the central mountain ranges, and across to the Atlantic Coast. As a consequence, the spectral response of land cover categories, such as pasture, forest and some crops, changes along this gradient. Banana fields and swampy palm forests both occur in the Atlantic Zone, at similar distance ranges from the Pacific Coast. To better discriminate between these two cover types, banana fields were digitized on the screen, and a distance model was then created from them. This model was combined with a 'distance from the Pacific Coast' raster, thus creating a discrete continuous model of two distance variables. This distance model was created to better discriminate particular land cover categories present in the study area and is not necessarily a relevant ancillary variable in other parts of the world. Nevertheless, the concept of a distance model is applicable to other features that might be present in other situations (e.g. distance from a city relative to crops that must be trucked in).

Using custom tables, the three layers representing elevation, time of access, and continuous distance categories were recoded to 8 bit values to reduce data dimensionality and then clustered using the ERDAS-Imagine 8.3.1 ISODATA clustering algorithm. As a result of the clustering process, 537 strata were delimited, for which different sets of prior probabilities had to be estimated.

To estimate the class prior probabilities within each of the 537 strata, a gray level MD image was first computed using a maximum likelihood classification with equal priors. The chi-square approximation to the F distribution of the MD was then used to select, with 95% degree of confidence, the pixels for each spectral signature that were closer to their estimated class means. This yielded a threshold selection of 909 311 pixels. The training pixels not present in this MD threshold selection were added, so that finally 14.38% of the pixels used to sample the class frequency were obtained from the training data, and 85.62% from the MD threshold selection. Visually, these pixels appeared well distributed in space and thus acceptable for estimating class frequencies.

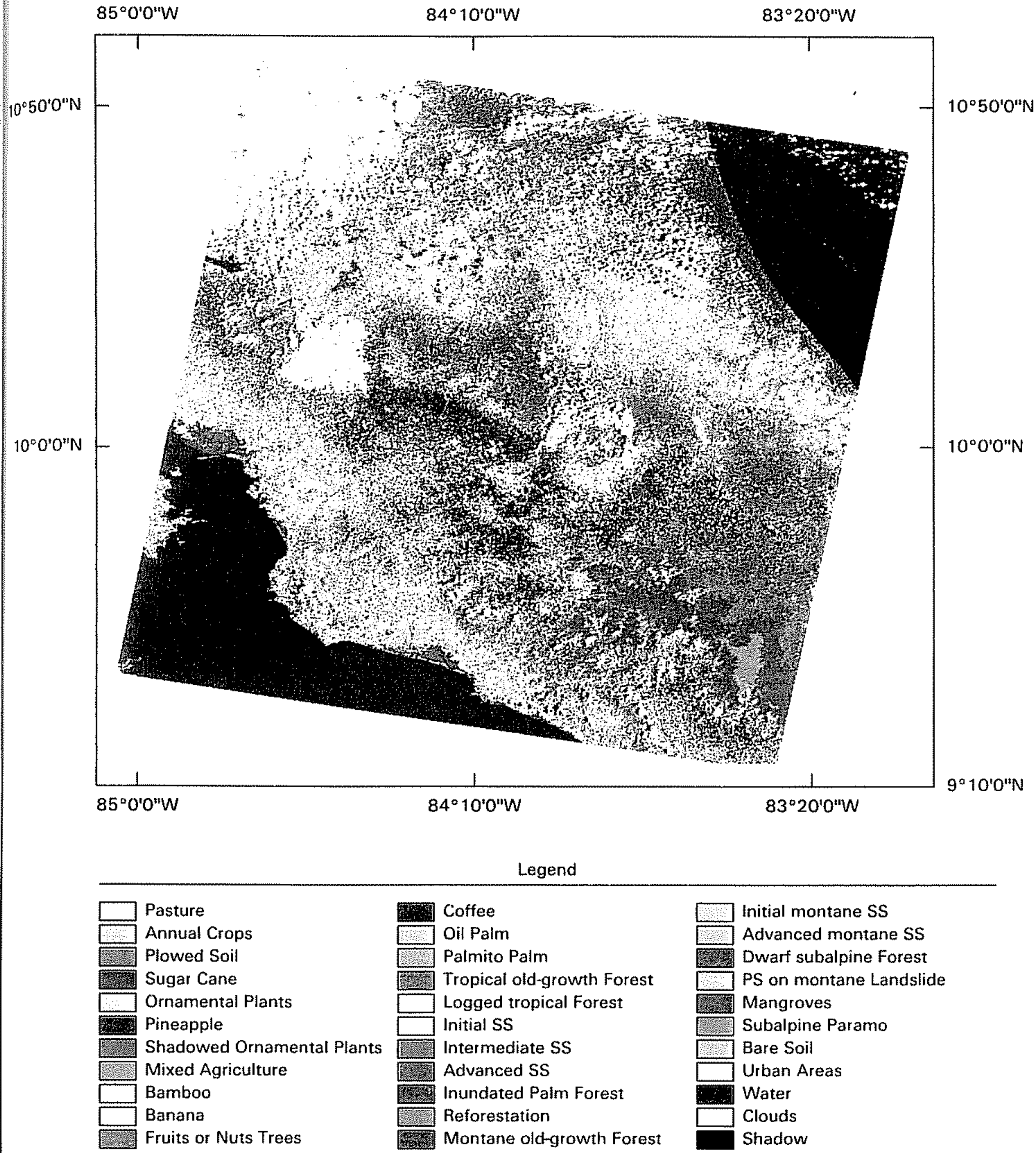
To eliminate misclassified pixels from the MD selection of pixels, lower and upper limits of each ancillary variable were defined for each spectral signature. Pixels classified outside this range were eliminated from the class frequency sample. For example, no signature was allowed that classified outside an elevation range of ± 500 m of the elevation of its corresponding training site (except for water, shadow, and cloud signatures). Signatures for reforestation classified at walking distances greater than 60 minutes were not allowed; nor were crop types beyond 40 minutes, and urban pixels beyond 10 minutes. For most cover types, especially crops, upper and lower elevation ranges were defined, according to the information obtained from interviews with local agronomic experts, from literature, and from personal experience. The same was done for certain natural vegetation types (mountain oak forests above 1600 m, subalpine paramo above 2800 m, mangroves below 10 m, etc.). Some cover types were allowed to be classified only within certain distance ranges from the Pacific Coast: mangroves at less than 5 km, oil palm within 15 km, and export banana fields beyond 85 km. After the application of these selection criteria, 764 636 pixels were left to estimate the class frequencies in the 537 strata.

To avoid exclusion of a class signature in a particular stratum just because it was not sampled, a frequency corresponding to 0.1 pixels was added to the pixel selection of each stratum existing within the range of conditions allowed for each signature. The resulting frequencies were normalized so that their sum was 1.0 for all signatures and used as an estimate of the class prior probabilities. The image was then classified again, using a computing routine that classified each stratum with its corresponding prior probability set. The scene stratification was performed using ERDAS Imagine 8.3.1 spatial modeler language. The sequence of computing routines to extract the image pixels corresponding to a stratum, their classification with the appropriate prior probability set, and union with previously classified strata was controlled by a Visual Basic program written in house.

For the classification accuracy assessment, 252 sites encompassing 15 206 pixels were selected from ground data provided by three independent studies. The selection was necessary because not all reference data could be assigned to the class definitions used in this study. The information provided by the independent studies did not include reference data about some of the categories classified in this study, whose classification accuracy could therefore not be assessed independently. The independent reference data used to evaluate the classification accuracy were provided by Helmer (1999), a local NGO (FUNDECOR), that collected them in 1996, and a student of the Swiss Federal Institute of Technology, who collected them in an undisturbed mountain region at the beginning of the year 2000

To allow a comparison of the classification self-consistency for all classes, a contingency matrix was also computed using the 979 sites (164 466 pixels) of the training data as reference data.

Figure 2. Land-cover of central Costa Rica in March, 1996 (Result of the classification with modified prior probabilities)



5.3 Results

The result of the classification with stratified modification of prior probabilities is shown in figure 2. The classification accuracy of independently selected control sites was clearly superior using the method of modified prior probabilities (table 1). The self-consistency of the classification of the training sites was also higher using this method. After modifying the class prior probabilities the overall classification consistency in the 979 training sites improved from 74.6% to 91.9% (Kappa 0.912), while the overall classification accuracy of sites controlled in the field by independent studies improved from 68.7% to 89.0%. These results are consistent with those reported by other researcher that modified the prior probabilities within strata (Maselli *et al.* 1995). Of the 33 land-cover categories present in the map legend, especially the most spectrally similar classes (different forest types) were classified with increased accuracy. This demonstrates that the method of stratified modification of prior probabilities is helpful when spectral information is insufficient for discrimination.

Figure 3 shows the prior probability model generated with the methods described in section 5.2. The tail regions of the probability distributions do not appear truncated, which indicates that the criteria used to eliminate potentially misclassified pixels from the 95% confidence threshold did not create artifacts. No a single class has equal prior probability within the study region. Instead most classes exist only within a specific range of values of the ancillary variables. This is quite different from the implicit assumption made with the equal priors classification approach. The prior probability model actually represents assumptions that are consistent with previous knowledge about the distribution of classes along ancillary gradients. As expected, the likelihood of undisturbed forest types increases, while that of other classes, especially crops, decreases rapidly where walking distances are greater (figure 3a). At short walking distances all land-cover categories are present, but at longer ones their number goes down progressively, which increases the chance of correct class allocation to one of the remaining classes. The same occurs at higher elevations (figure 3b), where climate conditions become prohibitive for most tropical crops, and forests are less rich in species and, above 1600 m.a.s.l, increasingly dominated by the genus *Quercus*. The model also shows that cover types that appear at comparable elevations and access conditions, such as banana and oil palm fields, sometimes exist at different distances from the Pacific Coast (figure 3c).

The linear scale of the ancillary variables (graphics on the left side of figure 3) helps to check if the model assumptions are consistent with previous knowledge. However, because equal intervals of ancillary data represent very different surfaces of the study region (figure 4), the linear scale is inadequate for understanding the contribution of ancillary variables to the discrimination of categories and stratification of the study region.

In the right side graphics of figure 3, the three ancillary variables are scaled in 100 equal-area intervals, and the prior probabilities are represented as a function of these equal-area intervals. The more horizontal the distribution of a cover type in the equal-area model, the less the ancillary variable was helpful for stratification.

The prior probabilities for low values of time of access and elevation above sea level appear stretched in the equal-area graphics because large portions of the scene can be accessed from the closest road within few minutes or are of low elevation. In contrast, probabilities appear compressed for regions corresponding to long access time or high elevation because these areas represent only a small proportion of the study region. This stretch-and-compress effect is less evident in the model of the distance from the Pacific Coast, because surface differences for equal distance intervals are less important for this variable.

Table 1. Percentage of overall classification consistency and accuracy
(Band combination: TM1, TM2. TM3, TM4, TM5, TM7, NDVI, Brightness, Wetness)

	Classification Consistency in the Training Sites			Classification Accuracy of Independently Controlled Sites		
Land Cover Category	*n*	*equal priors*	*mod. priors*	*n*	*equal priors*	*mod. priors*
Pasture	27970	90.4	96.9	1763	86.3	95.0
Annual Crops	2592	88.8	97.1	0	---	---
Plowed Soil	2997	93.9	99.0	0	---	---
Sugar Cane	7624	94.2	98.8	482	67.6	82.8
Ornamental Plants	715	70.1	96.6	0		
Pineapple	1033	99.8	100.0	297	98.3	98.3
Shadowed Ornamental Plants	905	98.3	100.0	236	94.9	95.8
Mixed Agriculture	543	25.1	55.9	0	---	---
Bamboo	3322	65.9	93.5	0	---	---
Banana	22841	81.9	97.9	355	84.5	96.3
Fruits and Nuts Trees	4383	61.9	85.8	246	50.8	69.9
Coffee	4890	77.6	95.4	528	79.9	97.2
Oil Palm	3868	69.0	94.5	170	52.4	88.2
Palmito Palm	2680	79.6	90.1	0	---	---
Tropical Old-growth Forest	7819	45.4	82.0	1706	34.6	93.8
Logged Tropical Forest	8347	27.5	71.6	285	26.7	55.8
Initial SS(1)	1694	33.0	66.0	0	---	---
Intermediate SS(1)	3673	29.8	66.0	0	---	---
Advanced SS(1)	7514	31.1	68.6	296	57.8	77.4
Inundated Palm Forest	7201	64.2	96.0	352	85.2	97.7
Reforestation	7108	55.8	87.0	412	40.5	55.6
Montane Old-growth Forest	8109	83.7	99.1	2115	71.6	95.4
Initial Montane SS(2)	589	55.5	84.7	494	15.8	58.9
Advanced Montane SS(2)	669	50.8	63.8	618	30.3	73.9
Dwarf Subalpine Forest	267	67.7	67.5	444	11.9	48.2
PS(3)on Montane Landslides	82	89.2	98.8	0	---	---
Mangroves	2513	93.0	99.2	490	95.1	95.9
Subalpine Paramo	1920	99.1	99.8	1782	86.9	93.0
Bare Soil	1821	97.4	99.9	160	70.0	93.8
Urban Areas	742	96.6	100.0	160	48.8	84.4
Water	10426	99.9	100.0	754	99.9	100.0
Clouds	3589	100.0	100.0	963	100.0	100.0
Shadow	4020	99.7	99.9	98	89.8	98.9
Overall Accuracy		74.6	91.9		68.7	89.0
Kappa		0.73	0.91		0.66	0.88

(1) Secondary Succession of tropical lowland forests after Finegan's model (1996)
(2) Secondary Succession of montane forests after Kappelle's model (1995)
(3) Primary succession

Figure 3. Prior probability model over cloud-free land areas
(Sea and major clouds were masked out)

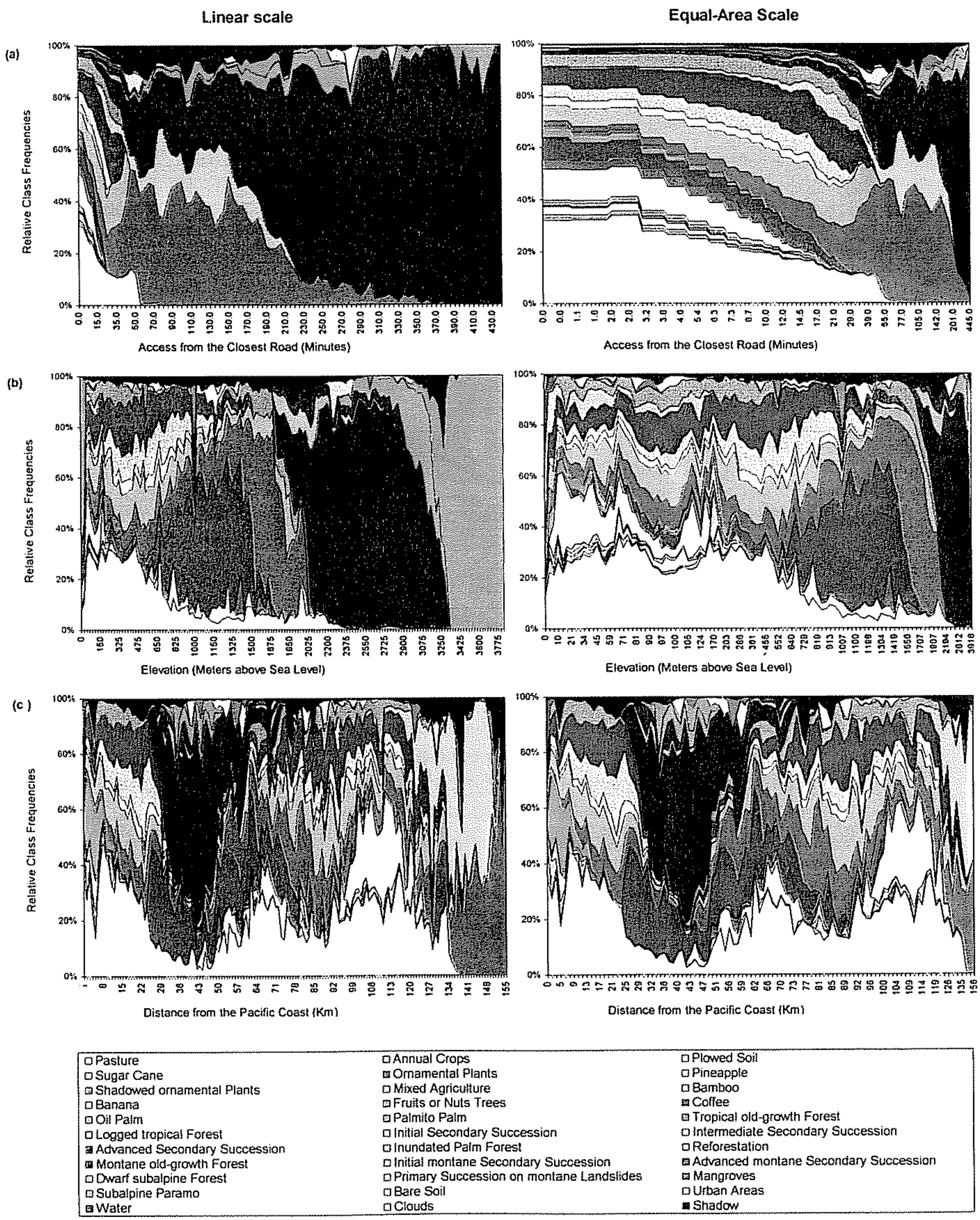
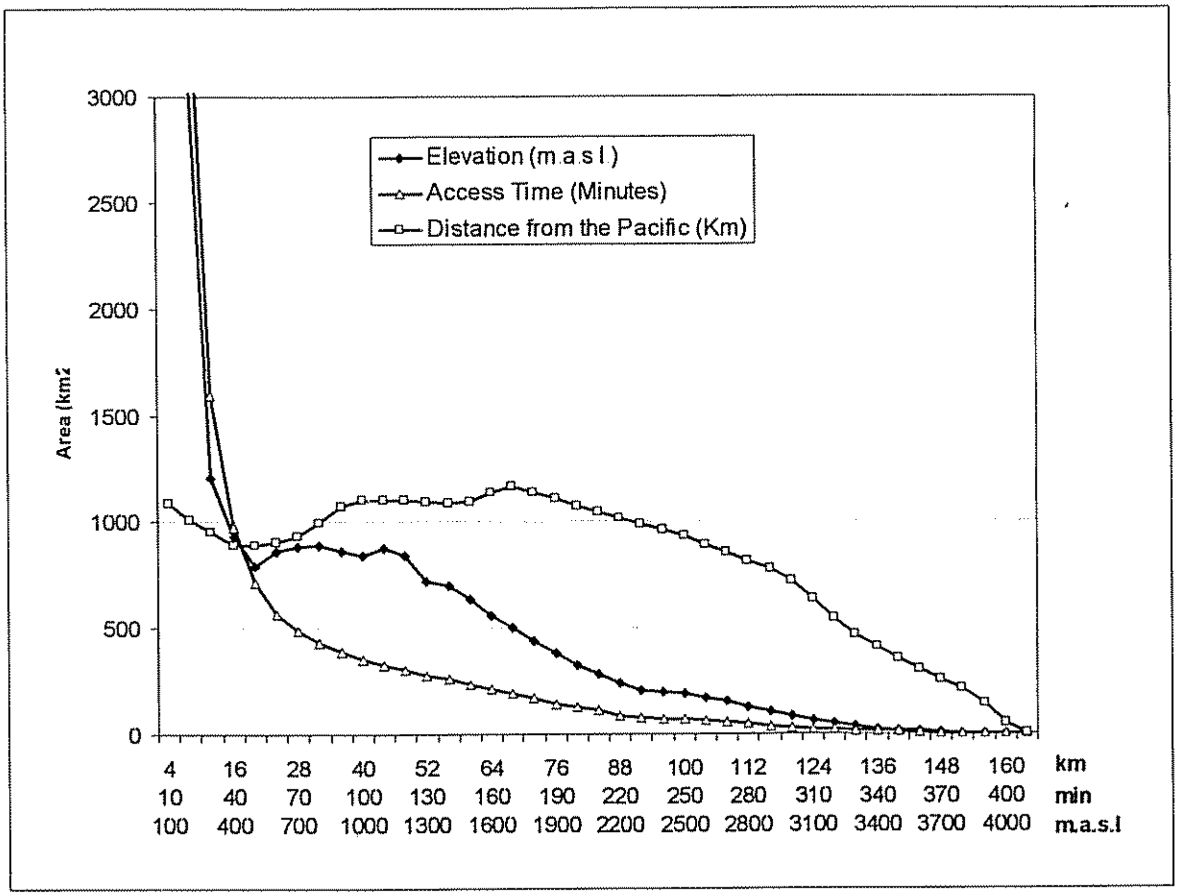


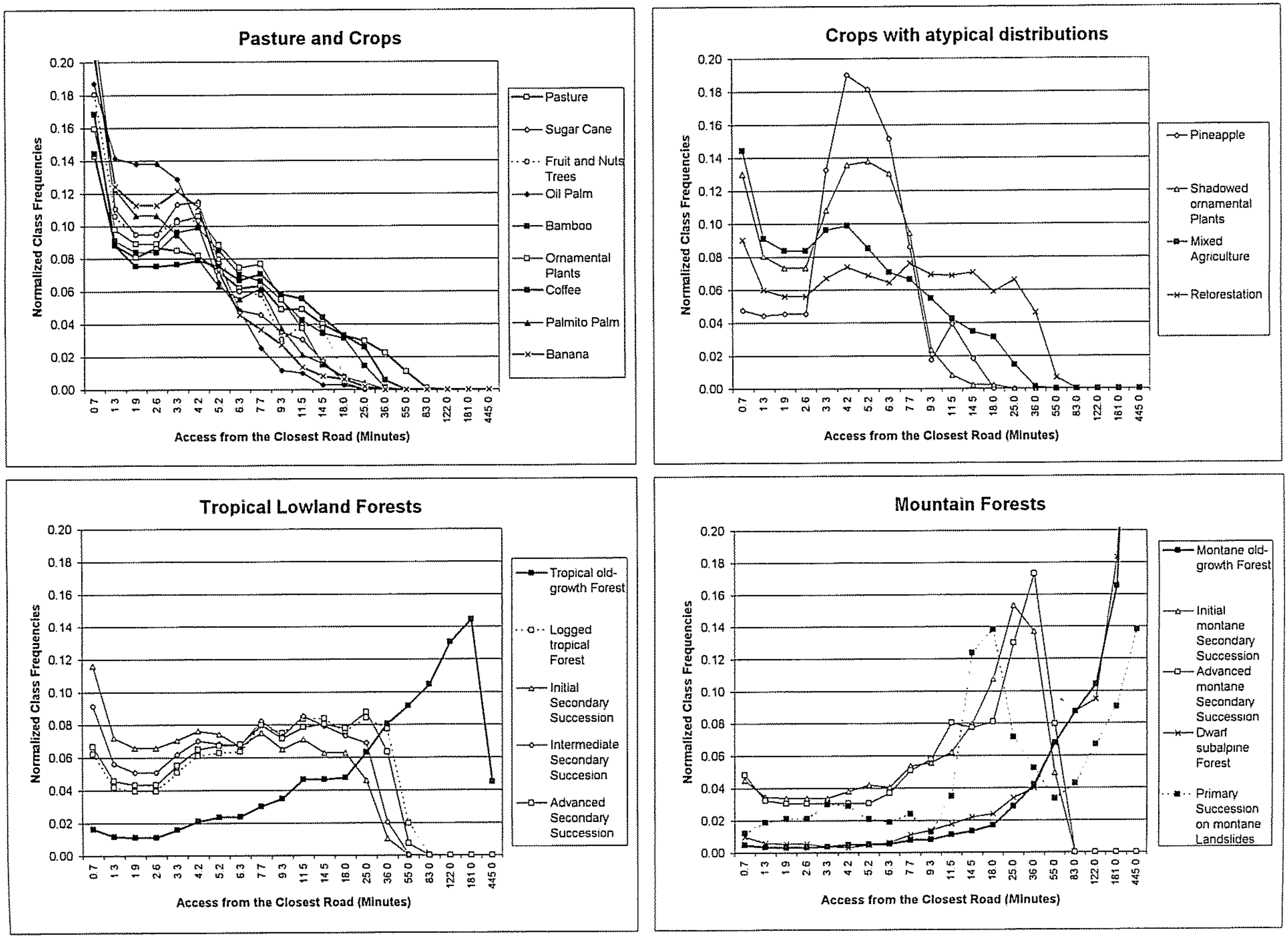
Figure 4. Arial distribution of the ancillary variables used to stratify the satellite image



The equal-area prior probability model for time of access indicates that about three quarters of the study area are dominated by cover types that were originated by human activities: pasture, crops, reforestation, secondary forests, and logged forests. Only in the less accessible quarter natural forests and other natural ecosystems still dominate.

Horizontally distributed prior probabilities in the time of access model indicate that this variable was less useful for stratification than the other two in about one quarter of the study area. However, time of access was the most effective ancillary variable for improving the discrimination of undisturbed forests from logged forests, secondary successions, and spectrally related crop categories such as shadowed coffee and mixed agriculture. There is a clear pattern that differentiates mature undisturbed forests from other human induced categories. In the normalized equal-area representation of time of access (figure 5), secondary successions and logged forests have similar distributions and are more frequent than old-growth forests in areas that can be accessed in less than 30-40 minutes. At access times longer than one hour, logging and secondary successions become very unlikely. Since farmlands are sometimes left fallow to restore soil fertility, secondary successions, especially younger successions, are the most frequent at the shortest times of access. Between 2 and 3 minutes the likelihood of secondary successions reaches a minimum, only to increase slightly up to 20 to 40 minutes, when a second maximum is reached. At even longer access times, the frequency of secondary successions and logged forests decreases progressively.

Figure 5. Normalized class distributions for time of access



As access time increases, the proportion of older successions increases, and that of young successions decreases Cropland and pastures appear thus to be abandoned more frequently and definitively at longer access times, probably because at long walking times benefits from land farming or cattle raising become marginal due to increased costs of transport.

Pasture and crop categories have a clear pattern of decreasing frequency from low to high values of access time. The decrease is much faster than for secondary successions, logged or planted forests. Only the distributions of pineapple and shadowed ornamental plants present a different pattern. The incompleteness of the digital road coverage used to model time of access is very likely responsible for this seemingly atypical pattern. As observed in Landsat imagery acquired between 1982 and 1996, large fields of pineapple and shadowed ornamental plants were established in the last decade. Some of the roads that give access to these new fields were constructed after the latest national cartography was completed. As a result, these categories appear to be more frequent in less well accessible areas even if, in fact, accessibility is very good. The bias in the time of access model caused by roads not present in the digital database can also be observed in the distribution of other crop categories, but is less evident for those categories that have a longer tradition in the study region and that are therefore close to older roads (e.j. coffee, sugar cane, and some banana fields). Mixed agriculture, which is mostly a land use form of small land-holdings and subsistence agriculture, might also be more frequent at longer access times because better accessible lands are used by larger farms that produce export crops.

Patterns of access time for forest categories in mountain regions are similar to those in the tropical lowland area. Undisturbed ecosystems (old-growth forest and dwarf subalpine forest) have very similar patterns and are mostly very difficult to access, while secondary successions are more abundant in areas that are easy to access. Primary successions, which are dominated by the pioneer *Alnus acuminata*, typically cover landslides incidentally generated either by road construction in steep slopes (peak at about 18 minutes of access time) or by natural disturbances such as earthquakes and extreme weather events.

Since most natural ecosystems and crops exist within specific elevation ranges, elevation above sea level helped to improve their discrimination, especially for those categories that have narrow elevation ranges (figure 6). Coffee, that is sometimes spectrally mixed with secondary successions and logged forests, dominates at an elevation range where these other categories are less frequent. This contributed to reduce confusion between these land cover categories.

Elevation was of little use for improving the discrimination within categories of secondary succession because there is no reason to expect a correlation of elevation with grow stages of secondary successions. However, old-growth forests are more frequent than secondary successions at higher elevations. This is because most protected area within the study region were established in mountain or foothill regions. Elevation was thus useful for improving the discrimination between old-growth forests and secondary successions, but not for the discrimination of categories within secondary successions. Elevation and time of access explain why tropical old-growth forests, logged forests and secondary successions were classified much better using modified priors probabilities.

Figure 6. Normalized class distributions for elevation above sea level

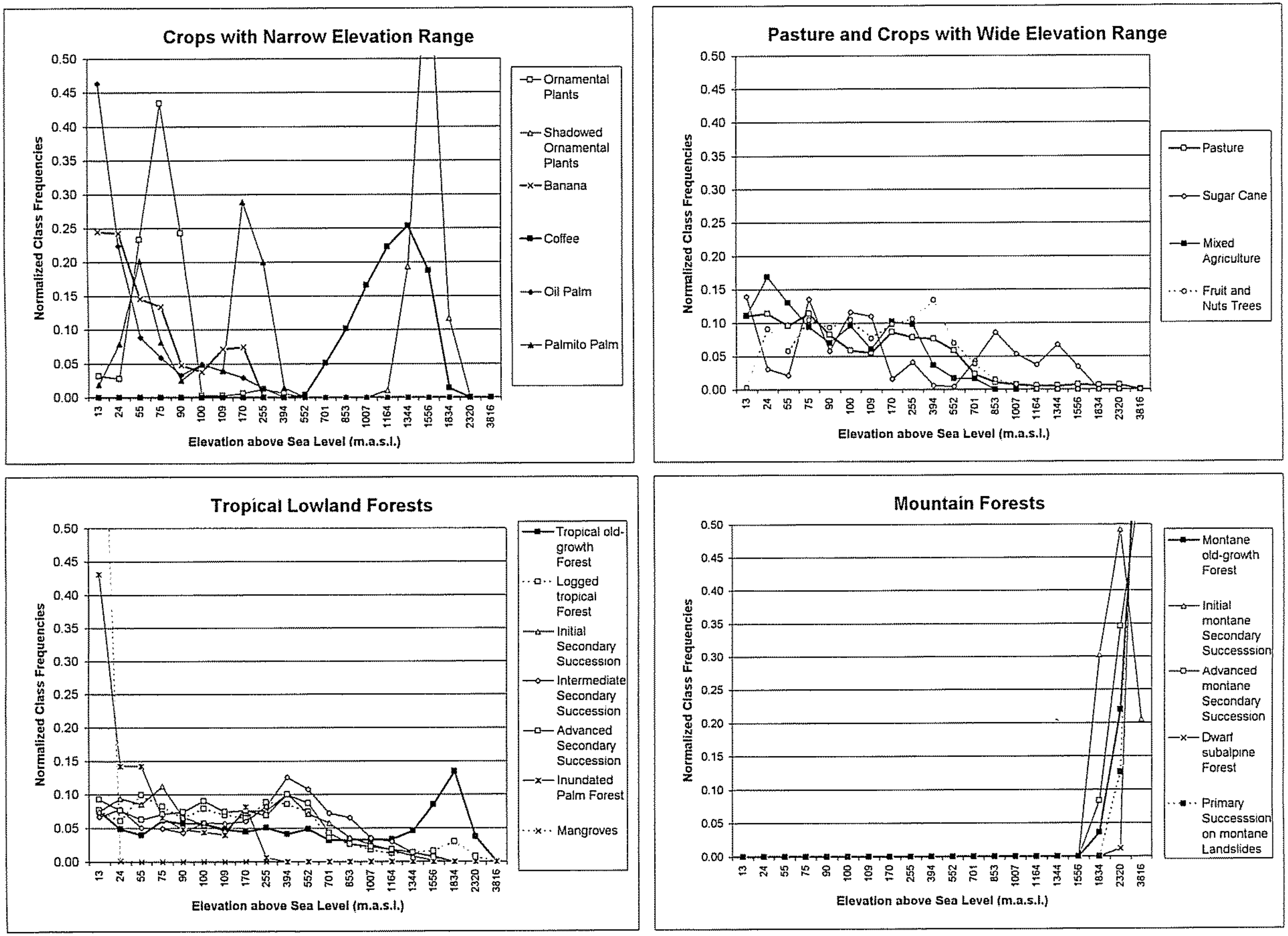
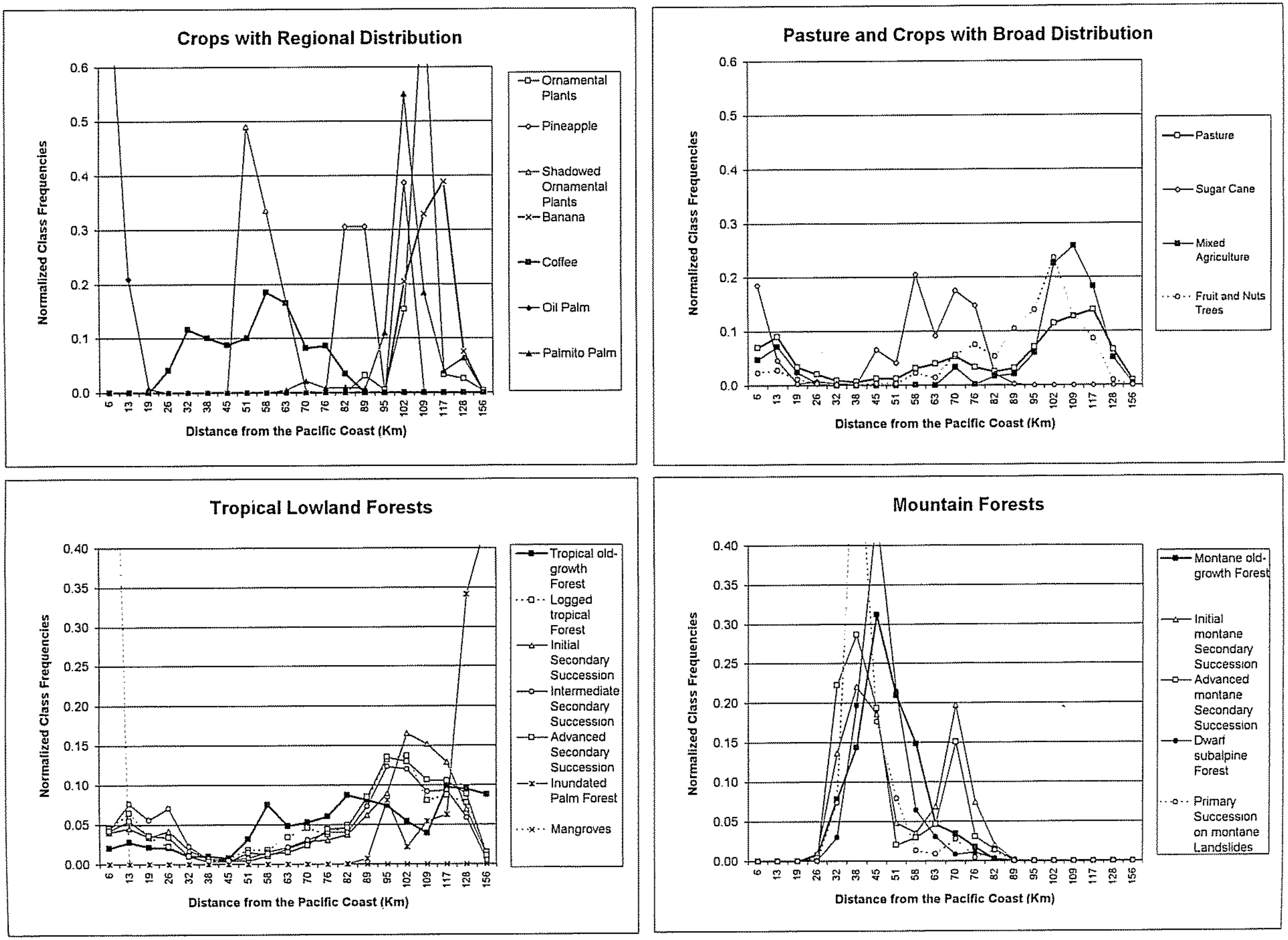


Figure 7. Normalized class distributions for distance from the Pacific Coast



The distance from the Pacific Coast was especially useful for improving the discrimination of cover categories with different patterns of horizontal distribution within the study region (figure 7). Some of these categories are spectrally similar and occur at comparable conditions of access an elevation (banana, oil palm, inundated palm forest).

As expected, different stages of tropical secondary succession have similar patterns of distance from the Pacific Coast. However, these patterns are slightly different from those of tropical old-growth forests, probably because there are regional concentrations of old-growth forests within the protected areas of the Central Volcanic Mountain Range. In mountains regions, secondary successions present two frequency peaks slightly on the left and on the right of the peak for undisturbed mountain forests. The left peak is explained by the greater abundance of secondary forests in the Pacific slope (Los Santos Reserve), which has been more severely degraded than other mountain forests of the Talamanca range. The upper Talamanca range explains the peak of undisturbed mountain old-growth forest. Right to it, the second peak of secondary successions, corresponds to the location of the Central Volcanic Mountain Range, where deforestation and land abandonment has been more advanced than in the mountain forests of the Talamanca Range.

The accuracy and scale of the available ancillary data layers affected the prior probability model. The fieldwork phase revealed that the road data were quite outdated, especially in regions where great abundance of secondary forests and tropical crops were found (Atlantic Zone). Better road data and a digital elevation model of higher resolution would probably have contributed to a more accurate model of time of access, and thus to a even better discrimination of cover types related to different accessibility conditions (e.g. old-growth, logged, and secondary forest types). Despite limitations in quality of the ancillary data used in this study, the classification with per-stratum modified priors resulted in much greater map accuracy than the traditional maximum likelihood classification with equal priors. Thus the method appears to be useful for improving the classification accuracy of large and complex landscapes with spectrally mixed land-cover categories.

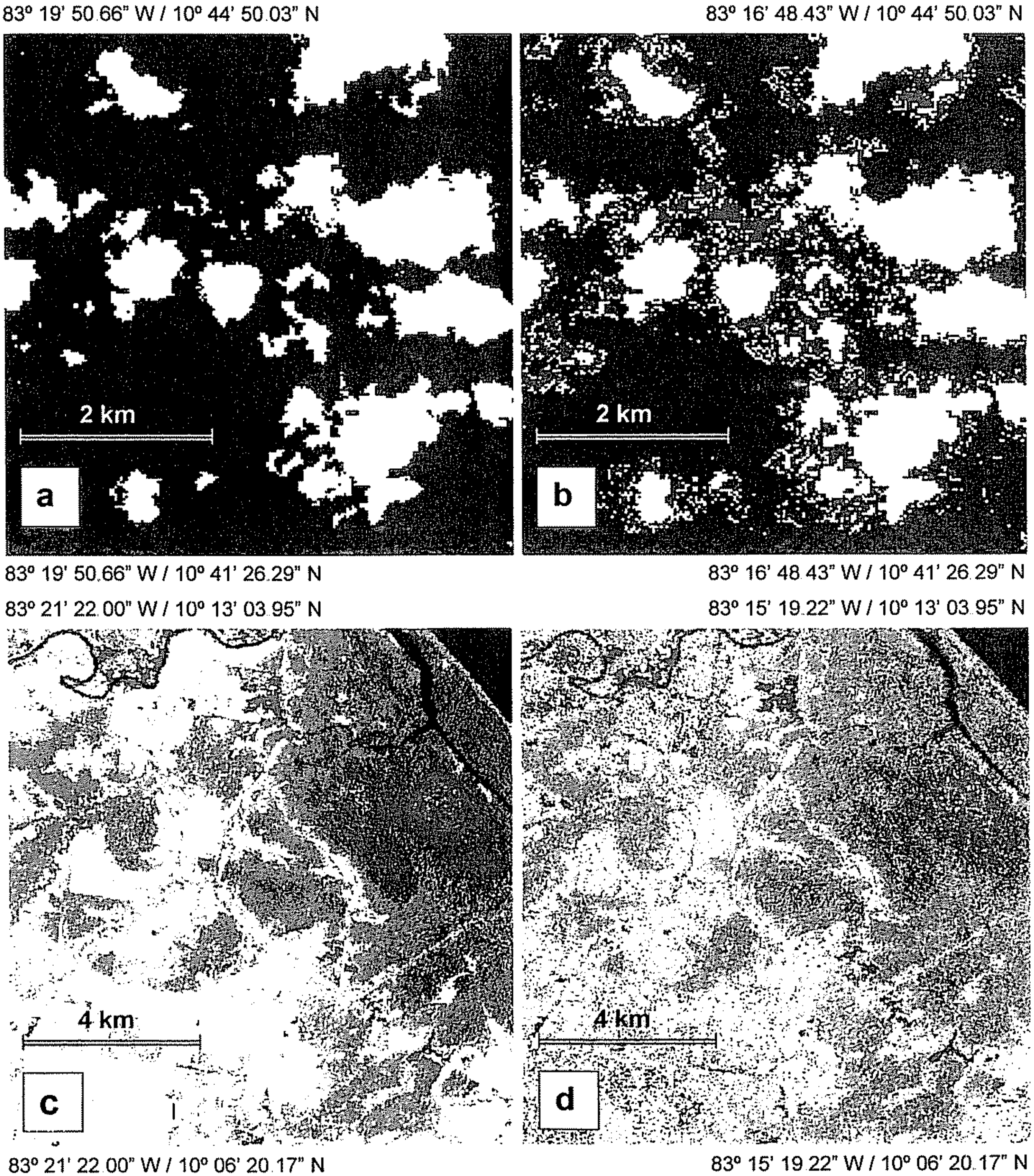
6. Discussion

In most remote sensing applications, only a subset of potential land-cover categories makes sense, ecological or otherwise, at a given pixel location. The proposed classification method is more accurate because all contextually unlikely classes, unless spectral evidence is strong, have little chance of classifying. Therefore, even if several land-cover categories are present in the map legend, at any given pixel location, the situation is similar to that of a classification with fewer categories. This favors correct class allocation. However, classification accuracy can only improve if image stratification was conducted carefully. Classification of no or too coarsely stratified images using unequal but invariant priors within the strata could even be less accurate to that of equal priors, because contextually impossible categories would be always present, and sometimes with increased probabilities. In addition, categories with low prior probabilities would have slim chance of classifying a pixel, even if the context would strongly favor their presence. Thus, fine image stratification is important not only to avoid artifacts at the boundaries of strata, but also to minimize the risk of biasing area and positional information.

Minimized classification of contextually impossible categories is clearly an advantage of classifying with locally modified prior probabilities. This becomes evident over ocean water, where only clouds, shadow and water signatures had significantly increased priors.

Figure 8. Visual comparison of classification results

Ocean water under semi-transparent clouds was correctly classified using modified priors (a), but was classified as urban area (red) or bare soil (yellow) with the equal priors classification (b).
The 'salt and pepper' effect is much greater in the output of the classification with equal priors (d). Contextually unlikely classes, such as fruit and nuts tree plantations (cyan) in inundated palm forests (purple), or sugar cane (red) in banana fields (yellow) are absent in the image classified with locally modified prior probabilities (c).



This resulted in a correct classification of clouds and water over the oceans, while unlikely cover types, like urban areas and bare soil, were classified in the border regions of clouds over ocean water with the traditional equal prior classification (figure 8a and 8b).

The 'salt and pepper' effect in the image classified with modified prior probabilities is also much reduced in comparison to the output of the traditional approach. This is especially true in the strata where only few categories are possible, such as areas with high elevation or difficult access. However, the difference is evident even in areas with low elevation and good access conditions, where several land-cover categories are likely to be present (figure 8c and 8d)

Misclassification in the image classified with locally modified priors is likely to occur in two situations: when a particular land-cover category exists over a wide range of ecological conditions, and in local contexts where several spectrally similar land-cover categories have approximately the same prior probability of existing. Both situations occur for secondary successions. Different growth stages of secondary successions have overlapping spectral signatures and exist in very similar geographic conditions. In addition, the range of ecological conditions where secondary successions exist is the same of that of other spectrally related categories, such as mixed agriculture, shadowed coffee, logged forests, tree plantations, and others. This made it particularly difficult to obtain high levels of classification accuracy for secondary successions. The choice of ancillary variables to stratify the scene and model the prior probabilities is obviously of great importance to minimize these situations. However, as experienced in this study, the availability of digital data will often determine the choice of ancillary variables to be considered. Nevertheless, for the stratification decision, it is important to give particular attention to land-cover categories that can occur in a wide range of ecological conditions and that are spectrally similar to other categories. In this study sugar cane and coffee were over- classified. Sugar cane because it appears from sea level to about 1600 m.a.s.l., and because several spectral signatures are required to adequately represent all phenological and harvesting stages present in sugar cane farms. Some of these stages were spectrally similar to other land-cover categories. Coffee because it is often grown in association with trees, which makes it spectrally similar to secondary forests, tree plantations, and mixed agriculture, which all exist in about similar conditions.

The contribution of ancillary variables to stratification and improved discrimination need to be evaluated separately. As in the case of time of access, a variable can be less important for stratification, but especially useful for improving the discrimination of spectrally similar categories.

In most study regions, the spatial distribution of cover categories is not random, and patterns of horizontal distribution can easily be recognized through simple visual interpretation. Ancillary variables with minimal or no biophysical meaning, such as the distance from the Pacific Coast in this study, can be useful to improve the discrimination of these categories. Such variables are easier to generate in digital format than more complex variables, such as time of access and elevation.

In areas where several land-cover categories are likely to occur (easy access, low elevation) several signatures have similar prior probabilities. This can result in some 'salt and pepper' effect. However, this effect has a different significance than in the output of the traditional approach. Regions of great mixture of pixels assigned to different classes are spectrally mixed <u>and</u> likely to harbor different classes. This conclusion can not be made when the classification is performed without modifying the priors on a contextual basis.

When the priors are constant and invariant, contextually impossible categories are present, which does not occur with locally modified priors For that reason, area estimates obtained through a classification performed using contextually modified class prior probabilities, and corrected with their corresponding estimates of omission and commission errors, are more likely to correspond to the true areas, than their equivalent estimates obtained from the traditional approach. This conclusion is important for ecological modeling using area estimates obtained from interpretation of satellite sensor data.

7. Conclusion

The methods presented in this article were successfully implemented in a tropical region of great complexity and over an area that largely surpass those studied with a similar approach by other researchers. It can be concluded that practical limitations of the classification method using modified prior probabilities have been successfully overcome. First, to address the issue of cost, the MD threshold method proved to be an effective alternative to stratified random sampling of training data for the estimation of priors. Second, the sample size obtained from this sampling technique was large enough to allow for a confident estimate of the prior probabilities, even if the number of strata and classes was large. The large sample size also minimized the potential bias of the prior estimates due to a non-random design of the training data. Third, the risk of incomplete and potentially biased prior estimates was minimized through a process for correcting the sampled class-frequencies using a set of ecological and geographical criteria. Finally, clustering the ancillary variable to stratify the scene appeared to be adequate to exploit the information content of these variables.

Software development is still required to mesh all the elements required to apply this classification strategy in an easy to use interface. Once such interface is available, analyst time and required training to apply the method will be reduced, thus allowing the method to be applied in most operational situations.

Acknowledgements

This research was supported by a grant from the Swiss Organization for Development and Cooperation (INTERCOOPERATION). The Landsat TM image was provided by the Foundation for the Development of the Central Volcanic Mountain Range (FUNDECOR), the digital cartographic data by the Costa Rican National Geographic Institute (IGN), and the Department of Geography at the University of Costa Rica (UCR). Independent ground data were generously provided by Eileen Helmer, FUNDECOR, and Dirk Ressel. Martin Artavia and Geoffrey Venegas made a fundamental contribution to the project by collecting training data in the field, and Mathias Tobler in Visual Basic programming. Finally, special thanks are due to the personnel at the GIS and Computer laboratory of the Tropical Agricultural Research and Higher Education Center (CATIE) for logistical support.

References

BOLAÑOS, R. A. M., and WATSON, V. C., 1993, Mapa ecológico de Costa Rica. (San José, Costa Rica: Centro Científico Tropical).

BOOTH, D. J., and OLDFIELD, R. B., 1989, A comparison of classification algorithms in terms of speed and accuracy after application of a post-classification modal filter. International Journal of Remote Sensing, 10, 1271-1276.

BOYD, D. S., FOODY, G. M., CURRAN, P. J., LUCAS, R. M., and HONZAK, M., 1996, An assessment of radiance in Landsat TM middle and thermal infrared wavebands for the detection of tropical forest regeneration. International Journal of Remote Sensing, 17, 249-261.

CAICCO, S. L., SCOTT, J. M., BUTTERFIELD, B., and CSUTI, B., 1995, A gap analysis of the management status of the vegetation of Idaho (U.S.A.). Conservation Biology, 9:498-511.

CIBULA, W. G., and NYQUIST, M. O., 1987, Use of topographic and climatological models in a geographiclal database to improve Landsat MSS classification for Olympic National Park. Photogrammetric Engineering and Remote Sensing, 54, 587-592.

CONESE, C., and MASELLI, F., 1993, Selection of optimum bands from TM scenes trough mutual information analysis. International Society for Photogrammetry and Remote Sensing. Journal of Photogrammetry and Remote Sensing, 48,2-11.

CONESE, C., MARACCHI, G., and MASELLI, F., 1993, Improvement in maximum likelihood classification performance on highly rugged terrain using principal component analysis. International Journal of Remote Sensing, 14, 1371-1282.

DAVIS, F. W., and DOZIER, J., 1990, Information analysis of a spatial database for ecological land classification. Photogrammetric Engineering and Remote Sensing, 56, 605-613.

DETWILER, R. P. and HALL, C. A. S., 1988, Tropical forests and the global carbon cycle. Science, 239, 42-47.

DIXON, R. K., BROWN, S., HOUGHTON, R. A., SOLOMON, A. M., TREXLER, M. C., and WISNIEWSKI, J., 1994, Carbon pools and flux of global forest ecosystems. Science, 263, 185-190.

DOWNTON, M. W., 1995, Measuring tropical deforestation: development of the methods. Environmental Conservation, 22, 229-240.

FINEGAN, B., 1996, Pattern and process in neotropical secondary rain forests: the first 100 years of succession. Tree, 11:119-124.

FLACK, J., November 1995, *"Interpretation of Remotely Sensed Data Using Guided Techniques"* http://www.per.dem.csiro.au/staff/FlackJulien/thesis/node2.html (2-Februar-2000).

FOODY, G. M., CAMPBELL, N. A., TRODD, N. M., and WOOD, T. F., 1992, Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. Photogrammetric Engineering and Remote Sensing, 58, 1335-1341.

FOODY, G. M., and CURRAN, P. J., 1994, Estimation of tropical forest extent and regenerative stage using remotely sensed data. Journal of Biogeography, 21, 223-244.

FOODY, G. M., and CURRAN, P. J., 1996, Identifying terrestrial carbon sinks: classification of successional stages in regenerating tropical forest from Landsat TM data. Remote Sensing of Environment, 55, 205-216.

FRANKLIN, S. E., and WILSON, B. A., 1992, A three-stage classifier for remote sensing of mountain environments. Photogrammetric Engineering and Remote Sensing, 58, 449-454.

GONG, P and HOWARTH, P. J., 1990, The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. Photogrammetric Engineering and Remote Sensing, 56, 67-73.

HARRISON, S., 1991, Population growth, land use and deforestation in Costa Rica, 1950-1984. Interciencia, 16, 83-93.

HELMER, E. H., 1999, The Landscape Ecology of Secondary Tropical Forest in Montane Costa Rica. PhD Thesis (Oregon: Oregon State University).

HILL, R. A., and FOODY, G. M., 1994, Separability of tropical rain-forest types in the Tambopata-Candamo Reserved Zone, Peru. International Journal of Remote Sensing, 15, 2687-2693.

HOLDRIDGE, L. R., GRENKE, W. C., HATHEWAY, W. H., LANG, T., and TOSI, J. A. Jr., 1971, Forest Environments in Tropical Life Zones: a Pilot Study (Oxford: Pergamont Press).

HOUGHTON, R. A., 1991, Tropical deforestation and atmospheric carbon dioxide. Climatic Change, 19, 99-119.

HUTCHINSON, C. F., 1982, Techniques for combining Landsat and ancillary data for digital classification improvement. Photogrammetric Engineering and Remote Sensing, 48,123-130.

INSTITUTO METEOROLÓGICO NACIONAL, 1987, Atlas Climatológico de Costa Rica. (San José, Costa Rica: Instituto Meteorológico Nacional).

JENSEN, J. R., 1996, Introductory Digital Image Processing: A Remote Sensing Perspective, 2nd edn (New Jersey: Printice Hall).

JENSEN, J. R., and TOLL, D. L., 1982, Detecting residential land-use developments at the urban fringe. Photogrammetric Engineering and Remote Sensing, 48, 629-643.

JOEN, B., and LANDGREBE, D. A., 1992, Classification with spatio-temporal interpixel class dependency contexts. IEEE Transactions on Geoscience and Remote Sensing, 30, 663-672.

KAPPELLE, M., 1995, Ecology of Mature and Recovering Talamancan Montane *Quercus* Forests, Costa Rica. Ph.D. Thesis, University of Amsterdam (Heredia, Costa Rica: National University, School of Environmental Sciences).

KLEINN, C., CORRALES, L., and MORALES, D., 2001, Forest area in Costa Rica: a comparative study of tropical forest cover estimates over time. Environmental Monitoring Assessment, 00:1-24.

KUMMER, D. M., and TURNER, B. L., 1994, The human causes of deforestation in Southeast Asia. BioScience 44, 323-328.

LEWIS, M. M., 1994, Species composition related to spectral classification in an Australian spinifex hummock grassland. International Journal of Remote Sensing, 15, 3223-3239.

MASELLI, F., CONESE, C., PETKOV, L., and RESTI, R., 1992, Inclusion of prior probabilities derived from a nonparametric process into the maximum likelihood classifier. Photogrammetric Engineering and Remote Sensing, 58, 201-207.

MASELLI F , CONESE, C , DE FILIPPIS, T , and ROMANI, R , 1995, Integration of ancillary data into a maximum-likelihood classifier with nonparametric priors. Journal of Photogrammetry and Remote Sensing, 50, 2-11.

MATHER, P M , 1985, A computationally-efficient maximum-likelihood classifier employing prior probabilities for remotely-sensed data. International Journal of Remote Sensing, 6, 369-376.

MAUSEL, P., WU,Y., LI, Y., MORAN, E. F., and BRONDIZIO, E. S., 1993, Spectral identification of successional stages following deforestation in the Amazon. Geocarto International, 4, 61-71.

MORAN, E. F., PACKER, A., BRONDIZIO, E. S., and TUCKER, J., 1996, Restoration of vegetation cover in the eastern Amazon. Ecological Economics, 18, 41-54.

MYERS, N., 1988, Tropical deforestation and remote sensing. Forest Ecology and Management, 23, 215-225.

NAGENDRA, H., and GADGIL, M., 1999, Satellite imagery as a tool for monitoring species diversity: an assessment. Journal of Applied Ecology, 36, 388-397.

SADER, S. A., WAIDE, R. B., LAWRENCE, W. T., and JOYCE, A. T., 1989, Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data. Remote Sensing of Environment, 28, 143-156.

SADER, S. A., POWELL, G. V. N., and RAPPOLE, J. H., 1991, Migratory bird habitat monitoring through remote sensing. International Journal of Remote Sensing, 12, 363-372.

SKIDMORE, A. K., and TURNER, B. J., 1988, Forest mapping accuracies are improved using a supervised nonparametric classifier with SPOT data. Photogrammetric Engineering and Remote Sensing, 54, 1415-1421.

SKOLE, D. L., CHOMENTOWSKI, W. H., SALAS, W. A., and NORBE, A. D., 1994, Physical and human dimentions of deforestation in Amazonia. BioScince, 44, 314-322.

SKOLE, D. L., and TUCKER C., 1993, Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988. Science, 260, 1905-1910.

STRAHLER, A. H., 1980, The use of prior probabilities in maximum likelihood classification of remotely sensed data. Remote Sensing of Environment. 10, 135-163.

SWAIN, P. H., and DAVIS, S. M., 1978, Remote Sensing: The Quantitative Approach (New York: McGraw-Hill).

TUOMISTO, H., RUOKOLAINEN, K., KALLIOLA, R., LINNA, A., DANJOY, W., and RODRÍGUEZ, Z., 1995, Dissecting Amazonian biodiversity. Science, 269, 63-66.

UNITED NATIONS ENVIRONMENTAL PROGRAM, 1995, Global Biodiversity Assessment. Edited by V. H Heywood (Cambridge: Cambridge Unversity Press).

VELDKAMP, E., WEITZ, A. M., STARISKY, I. G., and HUISING, E. J., 1991, Deforestation trends in the Atlantic Zone of Costa Rica: a case study. Land Degradation and Rehabilitation, 3,71-84.

Part III

Secondary and Logged Tropical Forests as Detected with Spectrally Enhanced Landsat TM Data and Modified Prior Probabilities

(Submitted for publication to Remote Sensing of Environment)

Part III

Secondary and Logged Tropical Forests as Detected with Spectrally Enhanced Landsat TM Data and Modified Prior Probabilities

LUCIO PEDRONI
Tropical Agricultural Research and Higher Education Center (CATIE), 7170 Turrialba, Costa Rica*

ABSTRACT

NDVI and Tasseled Cap are often claimed to enhance spectral separability of vegetation categories with different leaf biomass, canopy roughness, and species composition. Since secondary, logged, and undisturbed forests have different structure and composition, spectrally enhanced Landsat TM data are hypothesized to be useful for their discrimination. However, evidence supporting this hypothesis in tropical regions is weak, because good correlations between spectral data and data on biophysical properties of tropical forests were rarely observed. Instead, spatial patterns of forest occurrence might be related to geographical variables, such as ease for human access, elevation, and others. Inclusion of these non-spectral variables into the classification process might thus improve the discrimination of spectrally overlapping forest categories.

In a complex study region in central Costa Rica, a Landsat TM scene from March, 1996 was classified using 18 different combinations of spectral bands, NDVI and Tasseled Cap features. These classification experiments resulted in minor changes of the classification accuracy.

A Bayesian classification approach was then tested, in which the class prior probabilities were modified for 537 image strata representing different ecological and geographical conditions of the study region. The overall classification accuracy of this experiment was at least 17% superior to those that were obtained with the previous classifications. The land cover categories with the most important improvements in classification accuracy were the spectrally similar forest categories.

The hypothesis of spectral separability of tropical forest categories is discussed, and the potentials of methods that integrate spatial data with remote sensing data are demonstrated.

KEYWORDS: Secondary forests, Landsat TM, NDVI, Tasseled Cap, Prior Probabilities, Costa Rica

* lpedroni@catie.ac.cr, Tel: (++506) 556 15 30, Fax: (++506) 556 79 54

1. Introduction

The use of ancillary data has a great potential for improvement of the classification accuracy of vegetation types with little spectral separability. However, spectral data enhancement techniques have been used more frequently because they are easier to apply and because they have been helpful for the discrimination of broad land cover categories. Mapping of broad land cover categories, such as forest / non-forest, has had historical importance by increasing awareness about the problem of tropical deforestation (Woodwell *et al.*, 1987; Myers, 1988;. Sader and Joyce, 1988; Westman, *et al.*, 1989; Gilruth, *et al.*, 1990; Green and Sussmann, 1990; Hall *et al.* 1991; Campbell and Bowder, 1992; Grainger, 1993; Downton, 1995; Sohn *et al.*, 1999). However, this level of analysis of remotely sensed data portrayed an incomplete picture of tropical forest environments, since vegetation recovery through secondary succession, and forest disturbance caused by timber logging, have only rarely been addressed.

Secondary forests are thought to cover millions of hectares of abandoned land throughout the tropics (Dourojeanni, 1987; Brown and Lugo, 1990; Lugo and Brown, 1992; Fearnside and Guimaraes, 1996). While ecological research has made considerable progress for identifying patterns and processes of tropical secondary succession at the patch level (Saldarriga *et al.*, 1988; Janzen, 1988; Kappelle, 1995; Finegan, 1996), remote sensing could only provide incomplete and coarse information about forest succession and disturbance at the landscape level. Critical questions, such as the potential contribution of tropical secondary forests in offsetting global carbon dioxide emissions, have thus not yet been answered conclusively (Brown, *et al.*, 1996). If secondary and logged forests are erroneously classified as undisturbed old-growth forests, carbon pools in tropical forests are likely to be overestimated, while carbon sinks might be underestimated. More accurate data on forest succession and disturbance at the landscape level are also required to address the conservation of biological diversity (MacArthur and Wilson, 1967; Westman *et al.*, 1989), the restoration of soils, the management of water resources, and the sustainability of tropical agricultural systems. The remote sensing community is thus asked to abandon traditional forest / non-forest mapping and to develop methods that allow discrimination of forests categories that have distinctive ecological function, structure, and composition. Unfortunately, mapping secondary, logged, and undisturbed tropical forests using Landsat TM data is a complex task because of the spectral similarity of these categories. For improving their discrimination, spectral data enhancement techniques have been used with mixed results (Sader *et al.*, 1989; Foody and Curran, 1994; Sader 1995; Helmer, 1999). Only more sophisticated data classification techniques coupled with high density ground data resulted more consistently in improved classification accuracy of secondary forests (Brondizio *et al.*, 1994; Hill and Foody, 1994; Moran *et al.*, 1996; Foody *et al.*, 1996). If large areas are to be studied, methods that require stratified forest inventories and farm-level interviews are of little application because of cost constraints. Such working approaches were indeed an important facet of research that reported success in mapping secondary forests using Landsat TM data.

This paper presents the results of 18 classification experiments, in which a better discrimination of tropical secondary, logged, and old-growth forest was sought using two different strategies. The first strategy was based on the hypothesis that spectrally enhanced data were useful for improving the discrimination. The Normalized Difference Vegetation Index (NDVI) and the first three features of the Tasseled Cap transformation were computed and added to the spectral band of a Landsat TM image of central Costa Rica. The study region was then classified using different spectral band and index band combinations with the aim of identifying the band combination that would result in the more accurate discrimination.

The second strategy was based on the hypothesis that the discrimination of tropical forest categories can be improved using ancillary spatial data from a Geographical Information System (GIS). To test this hypothesis, the study region was stratified using the ancillary data, and a set of class prior probabilities was estimated for each stratum. The scene was then classified using a Bayesian approach, and the results compared with the previous experiments.

As a final experiment, training sites were selected in two ecological life-zones (*sensu* Holdridge *et al.*, 1971), and the spectral signatures were analyzed with the aim of understanding patterns of spectral response and the usefulness of spectral enhancement.

2. Background of the study

Efforts to map successional forest using remotely sensed data are not new, but the first attempts in the 80's were unsuccessful. Woodwell *et al.* (1986) and Woodwell *et al.* (1987) were unable to differentiate secondary growth from mature forest in Rondonia using MSS and AVHRR data. The failure was later attributed to the coarse spatial resolution of these sensor systems. Sader *et al.* (1990) suggested that discrimination of different forest types in the humid tropics might be possible using remotely sensed data with finer spatial resolution, such as Landsat TM or SPOT multispectral scanner data, which are also taken in spectral bands that are more suitable for vegetation studies. Results obtained by later studies that used such data supported this view (e.g. Roy *et al.* , 1991; Saxena *et al.*, 1992; Jusoff and D'Souza, 1996). However, limited discrimination capacity was not only caused by the spatial and spectral resolutions of data, but also by the digital classification procedures that were used. Most traditional classification techniques are unable to capture the whole information content of the spectral data. This has been demonstrated by Tuomisto *et al.* (1995), in a rain forest in Peru, where the visual interpretation of Landsat TM color composites produced more reliable information than the digital classification.

The computation of vegetation indexes, such as NDVI and Tasseled Cap, has often been proposed in remote sensing literature, as a simple strategy to extract more information from the spectral data. The use of such indexes was based on the underlying hypothesis that certain features present in the spectral data set were correlated with leaf biomass, canopy roughness, and other distinctive characteristics of the vegetation, but needed some type of enhancement practice to allow for digital discrimination. However, improved classification results using these techniques were not always reported. For example, in Puerto Rico and Costa Rica, Sader *et al.* (1989) found poor correlation between the NDVI calculated from Landsat TM data and forest successional stage and forest biomass. They concluded that NDVI was affected by sun-angle variations caused by topography and recommended the use of NDVI for low relief tropical forests. In a later study in two flat areas of Guatemala, Sader (1995) found that when the NDVI difference technique was applied, it was effective in stratifying major change categories, but not in effectively distinguishing among differences between the clearing of forest fallow and the clearing of older forest. In a research site in Ghana, Foody and Curran (1994) investigated the correlation between remotely sensed data from different sensor systems, vegetation indexes and forest biophysical properties measured in permanent plots. They concluded that 'further work to refine the relationships between remotely sensed radiation and biophysical properties related to regenerative state are required if remotely sensed data are to be used to locate and estimate the strengths of carbon sinks in tropical forests' (Foody and Curran, 1994, p. 240) While most research has focused on tropical lowland regions, Helmer (1999) used Tasseled Cap features in a mountainous region in central Costa Rica. She found that Tasseled Cap was useful to classify secondary forest sites on sunlit slopes, but not on shadowed ones. At her research site, high values of the third Tasseled Cap feature (the "Wetness" index) were correlated with old-growth forests, a result that was consistent with Steininger's findings in Brazil (1996) and with studies in the temperate region (Fiorella and Ripple, 1993; Cohen and Spies, 1992; Cohen *et al.*, 1995). However, Helmer admitted that the Kappa accuracies she obtained for different band and Tasseled Cap feature combinations did not differ significantly at the 95% confidence level. In her conclusion, she recommends stratifying by ecological zone and incorporating ancillary data for mapping forest successional stages in mountainous regions (Helmer, 1999, p. 45).

Most research that used spectral enhancement techniques such as NDVI and Tasseled Cap has been based on per pixel analysis techniques. Such classification techniques do not extract any information about the texture of the image data. However, textural variations in the image data might correlate with forest canopy roughness, a feature that is important to differentiate between old-growth forest types, that have emerging trees and canopy gaps, and secondary successions, that have a more homogeneous canopy texture (Finegan, 1996).

In a study site in Peru, Hill and Foody (1994) found that texture variations between different forest types could be spectrally enhanced using low-pass filters, therefore improving classification results. Low-pass filters reduced data variance and spectral overlap between classes. Texture variations were associated with forest structure (canopy height, smoothness, continuity, and tree density), but the authors acknowledged that the precise relationship between forest structural features and textural differences in the image data needed further investigation.

Nevertheless, extraction of textural information has been a common strategy of two groups of researchers that were successful in the discrimination of tropical secondary forest stages. The first group has been working several years in Brazil (Mausel *et al.* 1993; Brondizio *et al.*, 1994; Moran *et al.*, 1994a; Moran *et al.* 1994b; Li *et al.*, 1994; Moran *et al.*, 1996; Brondizio *et al.*, 1996; Tucker, *et al.*, 1998) and the second has been working in Bolivia, Brazil and Peru (Lucas *et al.*, 1993; Hill and Foody, 1994; Curran *et al.*, 1995; Foody *et al.*, 1996; Boyd *et al.*, 1996). Methods used by these researchers involved intensive field surveys, texture classifiers, and, sometimes, ancillary data. Texture classifiers, such as Extraction and Classification of Homogeneous Objects (ECHO), are spatial-spectral classifiers that work in a two stage process: first the scene is segmented into statistically homogeneous regions, and then, the data from each segment are classified in mass using the conventional maximum likelihood approach (Sholz *et al.*, 1977; Kartikeyan *et al.*, 1994; Kushwaha *et al.*, 1994).

Successful discrimination of tropical forest categories has thus been related to considerable achievements for conducting farm-level surveys, forest inventories, and classifier operations. However, spectral pattern and texture analysis may not be sufficient to accurately map tropical forest environments that are severely fragmented, house great ecological diversity, or that are topographically complex. These factors add spectral complexity to the data, but are often correlated to other non-spectral variables. Ancillary spatial data might thus be used to control landscape variables that are causing "spectral noise" or to include in the discriminant analysis information about the spatial distribution of land cover. This latter approach is based on hypotheses such as: secondary successions following farmland abandonment are spatially correlated with conditions of easy access for humans; undisturbed old-growth forests are more frequent in areas of difficult access; and logged forests occur only in areas where timber transportation is feasible and tree species are of market value.

In Costa Rica, an association between deforestation and climate, slope, soil fertility and infrastructure for human access to forests has been demonstrate by previous studies (Sader and Joyce, 1988; Veldkamp *et al.*, 1992). Correlation between deforestation and landscape variables were also found in Brazil (Stone *et al.*, 1991; Moran *et al.*, 1994a), Guatemala (Sader, 1995), Guinea (Gilruth *et al.*, 1995), Honduras (Ludeke *et al.*, 1990), Madagascar (Green and Sussman, 1990), Mexico (Dirzo and Garcia, 1992), in the Philippines (Kummer and Tuner, 1994) and other parts of the tropical region.

If information about the spatial relationship between land cover categories and ancillary variables can be made available on a per pixel basis, classification strategies that are able to exploit this information should produce better classification results than traditional classification approaches.

This type of use of ancillary data to improve the classification accuracy has been proposed for at least two decades (Swain and Davis, 1978; Strahler, 1980; Hutchinson, 1982). However, it appears that applications in tropical forest environments have been rare. Of course, some classification methods that use ancillary data have been shown to have little application because of their deterministic nature (pre-classification stratification, post-classification sorting) or because data distribution assumptions might be violated (e.g. addition of an elevation data layer to the multispectral data set before the classification). However, such problems are avoided with the use of a non parametric classification strategy involving geographical stratification and per stratum modification of class prior probabilities (Mather, 1985; Maselli *et al.*, 1995).

The increased use of GIS for data management, decision support, and scientific research is broadening the availability of digital databases about landscape variables in tropical countries. Classification techniques that are capable of using the information stored in these databases are thus becoming potentially usable.

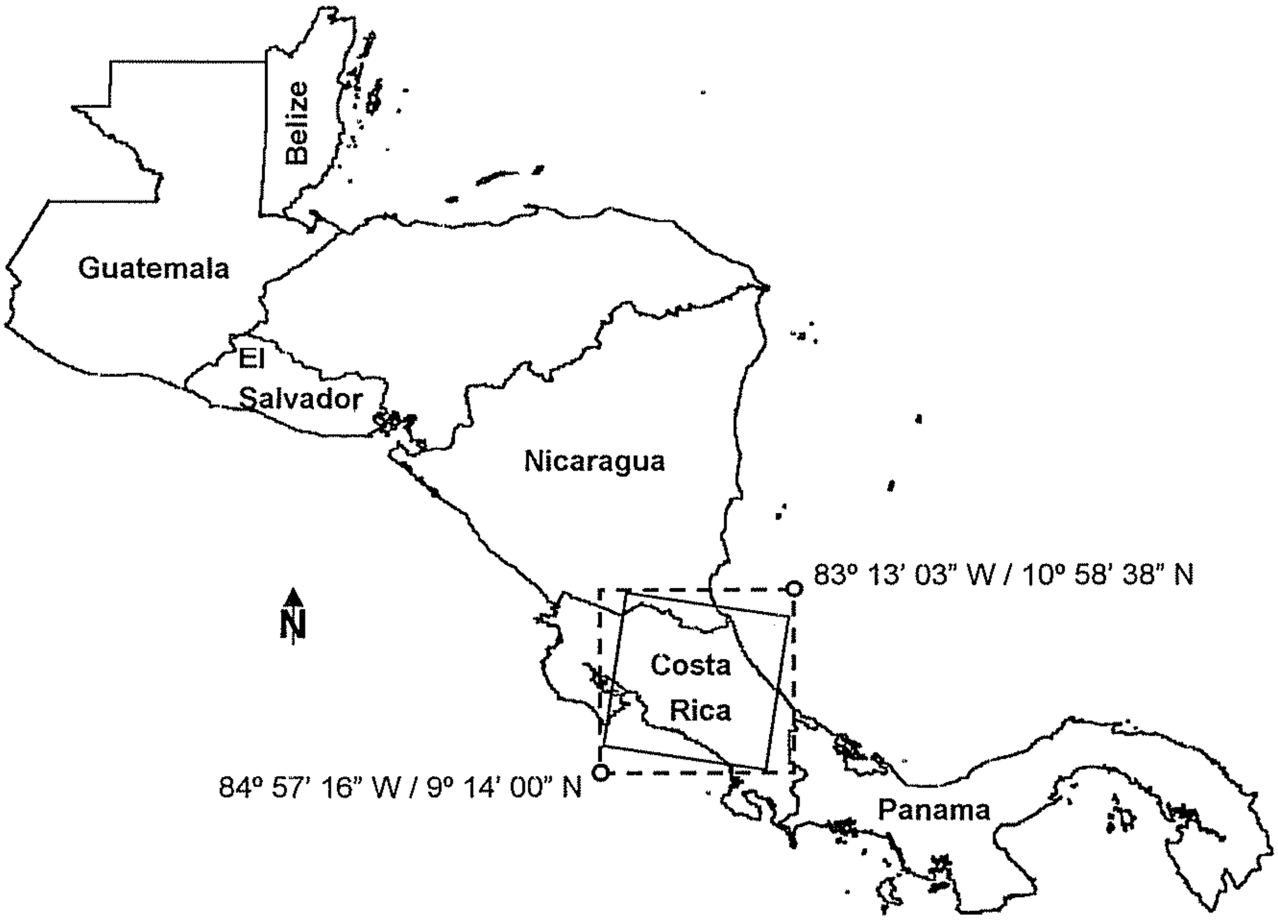
3. Materials and methods

3.1 Study region

As an example of large and complex tropical region, the entire Landsat TM scene (path 15 and row 53) of the central portion of Costa Rica (Figure 1) was chosen as the study area. The scene includes a ground area of 30,950 km^2, of which 5,433 km^2 (17.5%) are covered by water, and the rest is land with elevations ranging from 0 to 3,825 meters above sea level. The image was taken on March 1996, on a day of the dry season with relatively low cloud cover for this part of the World (approximately 13.2%). The image was of great spectral complexity because of the diverse ecological, topographical, climatical, and land-use conditions it represented: 12 life-zones and 11 transition-zones (*sensu* Holdridge *et al.*, 1971) were present within the scene according to the Ecological Map of Costa Rica (Bolaños and Watson, 1993). By comparison, the central and eastern United States combined have only 11 life-zones among them (Sawyer and Lindsey, 1964). Trade winds and the mountains of the Central Volcanic and Talamanca ranges create extremely variable climatic conditions in the study region, with average yearly precipitation ranging from 1,400 mm yr^{-1} to more than 7,000 mm yr^{-1} (IMN, 1987). This climatic variability results in brown vegetation patterns in the Pacific area during the dry season, and almost evergreen vegetation in the Atlantic Zone and in the mountain regions.

Sun-angle variations due to the rugged topography, small to medium-sized land tenure, fragmented patterns of forest cover, frequently mixed forms of land-use, and clouds and haze added considerable spectral complexity to the scene.

Figure 1. Study area



3.2 Classification scheme of secondary forest

Classification schemes have to be mutually exclusive and exhaustive (Congalton, 1991). This is particularly difficult to achieve in a classification scheme of secondary vegetation, since each growth stage is a phase in a succession time-continuum from clean land to mature forest. In this research, qualitative descriptions of succession categories were adopted to classify the training sites, because quantitative ground data were not available and too costly to obtain.

The forest succession categories were defined using models of the succession process described in the ecological literature. Secondary forests found in the basal, premontane and low montane tropical regions (*sensu* Holdridge *et al.* 1971) were classified using the three-phase model (initial, intermediate, and advanced secondary succession) described by Finegan (1992 and 1996), who developed his model from research conducted in the Costa Rican Atlantic zone.

In mountain regions, secondary forests were classified using Kappelle's (1995) and Kappelle's *et al.* (1996) two-phase succession model (initial and advanced montane succession). Kappelle's model was developed from research conducted in the upper mountains of the Costa Rican Talamanca Range.

Other remote sensing research adopted similar classification schemes. Mausel *et al.* (1993) and Foody *et al.* (1996) used a three phase succession model to classify tropical secondary forests that was very similar to that described by Finegan. They found distinctive spectral patterns for each of the three succession categories. Helmer (1999) used Kappelle's model to classify secondary forests in a mountain region of Costa Rica, and achieved a good discrimination. Thus, it appeared likely that a good classification could be achieved using these models as a classification scheme.

3.3 Fieldwork and data processing

The satellite image was georeferenced to the northern Costa Rican national coordinate system using a Lambert Conformal Conic projection. The thermal band was eliminated from the spectral data set because its spatial resolution and contrast were considered to be too coarse for spectral discrimination. The blue band was eliminated in the first classification test, but its exclusion was found to reduce the classification accuracy even though the data in this band were affected by haze. Atmospheric corrections were not performed because of lack of data about atmospheric conditions at the date of image adquisition.

The Normalized Difference Vegetation Index (NDVI) and the first three features of the Tasseled Cap transformation were added to the reflective bands, because the review of the literature suggested that they might enhance certain spectral features of the cover types focussed in this study.

During the period from March, 1998, to March, 1999, 826 sites were visited in the field and their coordinates gathered in average mode using a non differential Global Positioning System (GPS) unit (Garmin 12 XL). The majority of the sites visited in the field, had had their geographic coordinates read previously from the georeferenced satellite image. The center points were selected in the computer screen for land cover patches that met the following criteria: they were located outside areas covered by clouds and shadow; they were easy to delimit from surrounding land cover types; they appeared to contribute for the completion of an exhaustive training data set of all spectral patterns observed in the image; they were not too difficult to access; and they were large enough for training statistics development. At each field location the composition and structure of the vegetation was briefly described using a form especially designed for this study. Because of time and budget restrictions, no quantitative measurements were taken of the vegetation. Field-level forest description and classification were carried out with assistance of trained personnel of the Tropical Agricultural Research and Higher Education Center (CATIE) that had had several years experience in measuring permanent plots of secondary and primary forest, e.j. in the framework of research conducted by Finegan (1992 and 1996). This ensured accurate field-level identification of succession forest, and consistence with the classification scheme.

The GPS coordinates of the field sites were superimposed onto the satellite image to locate and delimit the polygons required for signature extraction. The polygons were delimited using the region-growing option of ERDAS Imagine 8.3.1 software, and their boundaries were cross-checked with sketches of the sites and their surroundings, that had been made during ground data collection. Some sites, especially those at the earliest stages of succession, were eliminated from the training data set, when the visual interpretation of the image suggested that they had significantly changed since the image data were acquired. The training data set was then complemented with additional data obtained through visual interpretation of the satellite image, especially for evident land cover categories such as water, bare soil, clouds, shadow, and urban areas. The final set of training data included 979 spectral signatures representing variations in age, structure, composition, moisture, phenological status, elevation above sea level, and sun-angle illumination for 33 land cover categories.

The pixel vectors of selected training sites were then extracted to perform a statistical comparison of the spectral patterns of the different forest types using SAS software. Comparable sites were defined as those belonging to the same ecological life-zone *sensu* Holdridge *et al.* (1971) and with similar sun-angle exposition. The sun-angle stratification of the scene was performed using a digital elevation model shadowed with the sun-angle parameters of the date and hour in which the satellite image was acquired. Two ecological life-zones were selected for spectral analysis: the Tropical Wet Forest (TWF), and the Premontane Tropical Wet Forest Transition to Basal (PTWFTB). These two life-zones categories covered the largest areas within the study region, and included secondary forest sites investigated by Finegan (1992, 1996). Their selection ensured that enough spectral observations from field described sites were available for spectral pattern analysis.

17 maximum likelihood classifications assuming constant and invariant class prior probabilities were then performed using the spectral bands and different vegetation index combinations. The band and index combination that produced the best classification output for the forest categories was then classified again, using a maximum likelihood classification routine in which the class prior probabilities were modified.

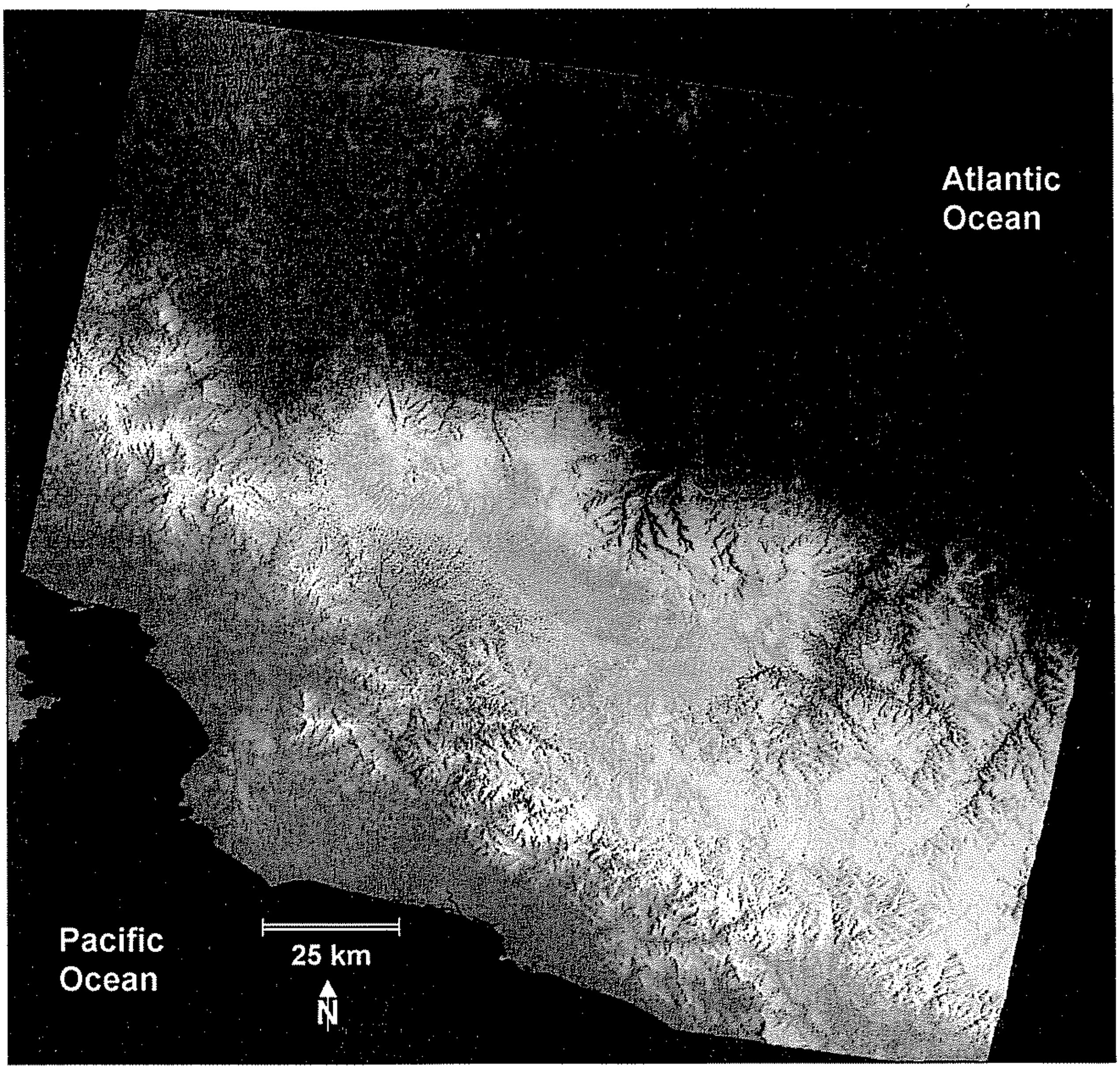
The class prior probabilities were modified according to the geographic context, that was described by three ancillary data sets representing models of: elevation above sea level; walking time required to access a pixel location from the closest road; and distance categories from the Pacific Coast and export banana fields (Figure 2). The three models were created from digital vector data at a 1 : 50,000 map scale, except for some portions of the contour line data set, that were available only at a 1 : 200,000 map scale. The data of these models were combined using ERDAS Imagine 8.3.1 ISODATA clustering algorithm, and the resulting 537 clusters were considered different geographical strata.

To estimate the class prior probabilities from sampled class frequencies within each of these strata, stratified random sampling in the field would have been impossible to achieve. Therefore, an alternative sampling procedure was used, that counted training data pixels as well as those from the best previous classification that with 95% confidence were of greatest spectral similarity (shortest Mahalanobis Distance) to the mean vectors of the training data. Using ARC/INFO macro language, the frequency of these pixels was then adjusted to fit the range of geographical and ecological conditions allowed for each land cover category according to the criteria of local experts and personal experience. The estimated class frequencies of each stratum were normalized so that they added to 1.0, and were then used as a model of class prior probabilities. With this procedure, information from the ancillary data set was included into the maximum likelihood decision rule probabilistically, thus avoiding the problems mentioned earlier related to determinism and data distribution assumptions.

The procedure used to estimate the class prior probabilities is described with more detail in another paper by the author of this research.

Figure 2. Geographic context of the study area

84°57'16"W / 10°58'38"N | 83°13'03"W / 10°58'38"N



84°57'16"W / 9°14'00"N | 83°13'03"W / 9°14'00"N

RGB composition of the three ancillary variables representing elevation above sea level, walking time required to access a pixel, and discrete continuous distance categories from the Pacific coast and export banana fields.

The results of the different classifications were compared to evaluate the effectiveness of the different spectral band and index combinations, and the usefulness of including information extracted from ancillary variables into the multi-spectral data classification process.

4. Results

The mean plots of the spectral data sampled at similar sun-angle illumination showed that spectral differences among forests with varying degrees of disturbance and at different stage of the succession were subtle (Figures 3).

The mean Digital Numbers (DN) were statistically different ($p < 0.05$) among forest categories for the same ecological life-zone (ANOVA and Tukey). However, they were also different within the same forest category between two ecological life-zones (t-test and Wilcoxon rank sum test).

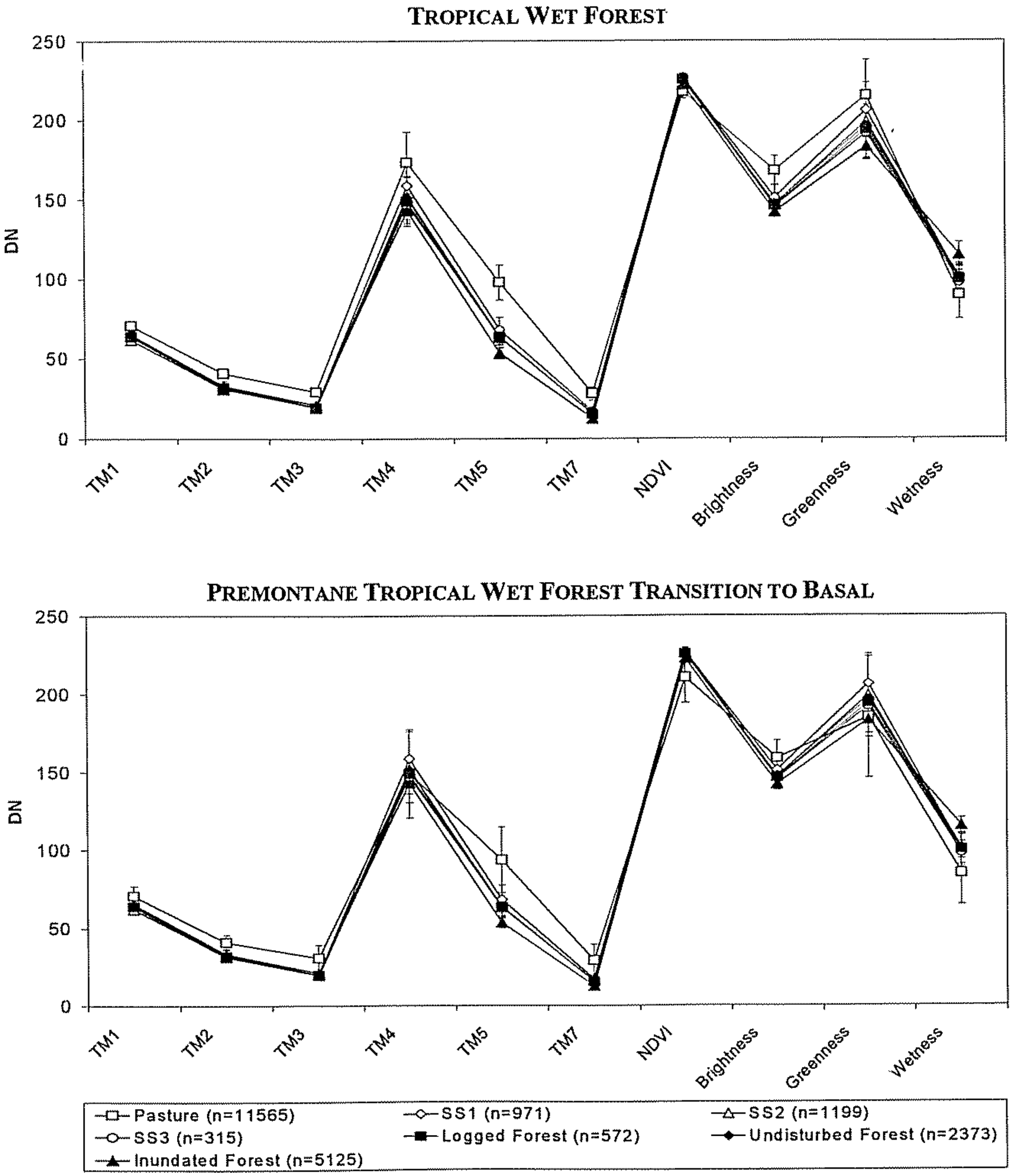
Forest categories were thus spectrally different between and within categories. However, the class spectral separability was not as good as suggested from the result of analysis of variance. The band and index mean values of any particular forest class were often within the standard deviation range of the means of the other classes. Only for inundated forests the mean DN of band 5 (middle IR) and the Wetness index were outside the standard deviation range of the other forest classes. These results were expected, since variations within individual forest categories were observed during fieldwork in terms of vegetation structure, composition, and humidity, and because the sample size used for spectral data analysis was large.

The underlying hypothesis of NDVI, that increased red absorption and near-IR reflectance is correlated with increased green leaf biomass was supported only for pasture and the forest categories as a group. In both life-zones, pastures showed less red absorption than the forest categories, while the differences between the forest categories were minimal. Near-IR reflectance decreased with succession from the initial to advanced stages, probably because of increased shadow and moisture. In pastures found in the TWF life-zone, the near-IR reflectance was higher than it was for that found for forest categories. The opposite was observed in pastures of the PTWFTB life-zone, which was consistent with the expectation of positive correlation between biomass and near-IR reflectance (Figure 4).

In both life-zones, near-IR reflectance increased following the sequence: advanced succession, logged forest, and primary forest, that is thought to correspond with increased leaf biomass, while visible reflectance values decreased from pasture toward intermediate forest succession stages and remained about constant for the other forest categories. Middle-IR values (especially band 5), that are generally related to water absorption, decreased in the sequence from pasture toward advanced forest succession. However, differences were minimal, except between pasture and initial succession stages.

The NDVI enhancement of the relationship between red and near-IR was useful for discriminating between pasture and forest categories. The latter had consistently higher NDVI values than pastures. However, NDVI was almost invariant among the forest categories, and was therefore useless for their discrimination.

Figure 3. Comparison of spectral mean plots in two life-zones



DN = Digital Number
TMx = Number of the spectral band of Landsat TM
SS1 = Initial Succession
SS2 = Intermediate Succession
SS3 = Advanced Secondary Succession
n = Number of pixels sampled

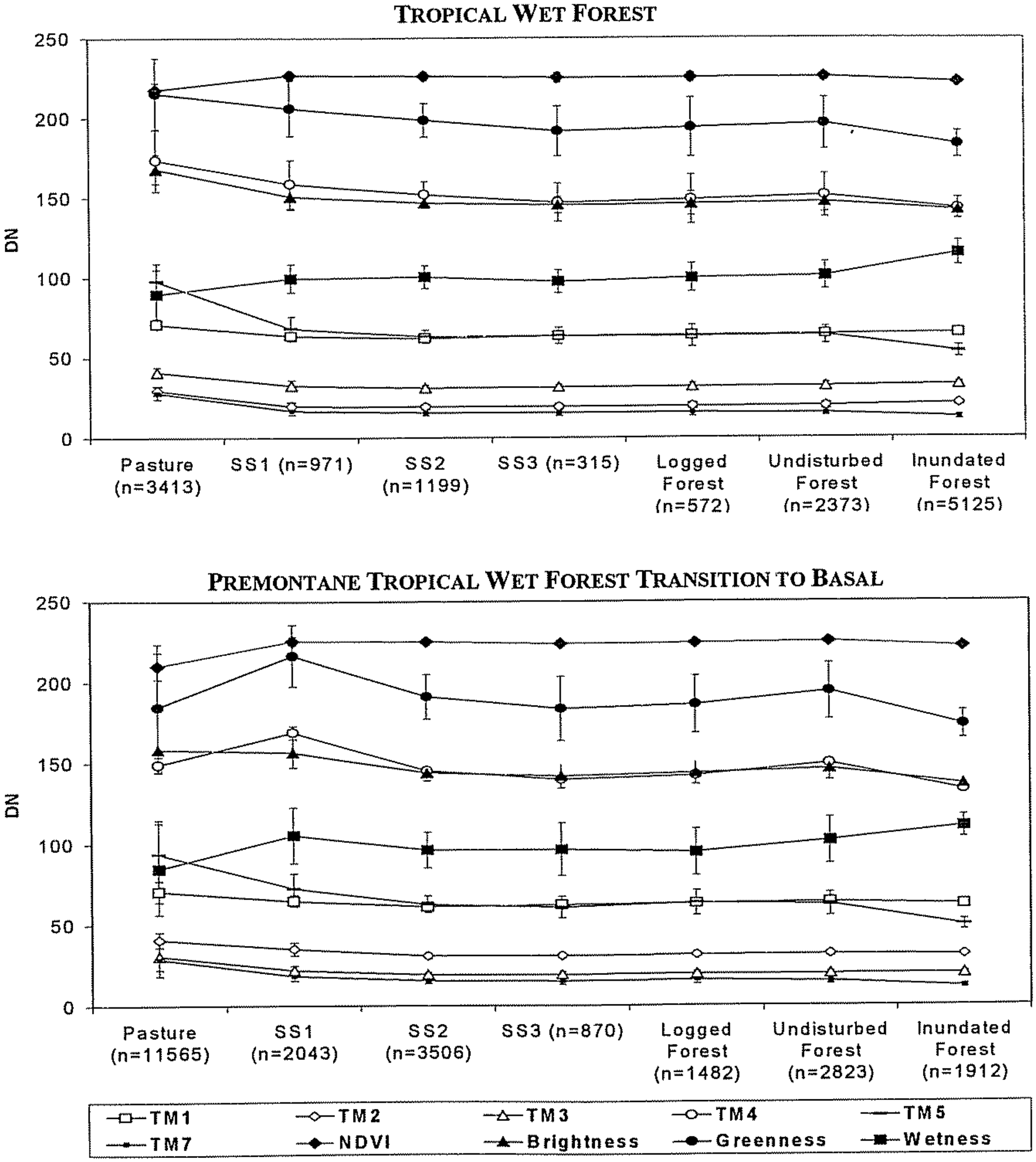
Tasseled Cap indexes appeared to be more useful for enhancing spectral differences between secondary succession stages, logged, and undisturbed forest categories. Brightness DN were higher for pasture and initial forest succession, approximately constant for the other forest categories, and lower for inundated palm forests. Greenness DN decreased from initial succession stages to advanced succession stages, increased following the sequence: advanced successions, logged forest, undisturbed old growth-forest; and were again lower for inundated palm forests. The peak in Greenness and Wetness DN for initial succession stages in the PTWFTB life-zone was most likely caused by the presence of very wet training sites in this category. Wetness DN were positively correlated with wetter forest sites, since they increased in the sequence: logged forest, undisturbed forest, inundated palm forest, and mangroves. Wetness DN are consistently lower for pastures, where moisture levels were generally lower than in forest categories. Wetness DN were negatively correlated with the middle-IR band 5, which confirmed that the third feature of the Tasseled Cap transformation was a good indicator of vegetation categories with increased water content.

The usefulness of spectral data enhancement using vegetation indexes for improving the discrimination of tropical forest categories was also assessed using a comparison of the classification accuracy obtained with the classification experiments. However, only 252 control sites (15,206 pixels) described in the field by other researchers could be made available for classification accuracy assessment of selected land cover categories. For that reason, the self-consistency of the classification of the 979 training sites (164,466 pixels) was calculated as well. This was considered sufficient to compare the performance of the different classification experiments.

Classification results showed that NDVI and Tasseled Cap indexes have some capacity to enhance spectral differences among forest categories and especially between pasture and forest. However, variations in classification accuracy were minimal compared to results obtained with the Bayesian classification (Table 1). In addition, only four band and index combinations (TM + NDVI, TM + Brightness, TM + NDVI + Brightness, TM + NDVI + Brightness + Wetness) produced superior overall classification consistencies than did the classification of the TM bands alone (Table 2). The band and index combination TM + NDVI + Brightness + Wetness was used for the classification with modified prior probabilities because it produced the highest average classification consistency for the forest categories, which are the land cover categories that were more difficult to discriminate.

From the evaluation of the classifications results, it was concluded that while NDVI and Tasseled Cap were of little advantage for increasing the classification accuracy of secondary, logged, and undisturbed tropical forest categories, the inclusion of information extracted from ancillary non-spectral variables was instrumental in achieving this goal.

Figure 4. Comparison of Mean Digital Numbers (DN) in two life-zones



DN = Digital Number
TMx = Number of the spectral band of Landsat TM
SS1 = Initial Succession
SS2 = Intermediate Succession
SS3 = Advanced Secondary Succession
n = Number of pixels sampled

Table 1. Percentage of overall classification consistency and accuracy
(Band combination: TM1, TM2 TM3, TM4, TM5, TM7, NDVI, Brightness, Wetness)

	Classification Consistency in the Training Sites			Classification Accuracy of Independently Controlled Sites		
Land Cover Category	*n*	*equal priors*	*mod. priors*	*n*	*equal priors*	*mod. priors*
Pasture	27970	90.4	96.9	1763	86.3	95.0
Annual Crops	2592	88.8	97.1	0	---	---
Plowed Soil	2997	93.9	99.0	0	---	---
Sugar Cane	7624	94.2	98.8	482	67.6	82.8
Ornamental Plants	715	70.1	96.6	0		
Pineapple	1033	99.8	100.0	297	98.3	98.3
Shadowed Ornamental Plants	905	98.3	100.0	236	94.9	95.8
Mixed Agriculture	543	25.1	55.9	0	---	---
Bamboo	3322	65.9	93.5	0	---	---
Banana	22841	81.9	97.9	355	84.5	96.3
Fruits or Nuts Trees	4383	61.9	85.8	246	50.8	69.9
Coffee	4890	77.6	95.4	528	79.9	97.2
Oil Palm	3868	69.0	94.5	170	52.4	88.2
Palmito Palm	2680	79.6	90.1	0	---	---
Tropical Old-growth Forest	7819	45.4	82.0	1706	34.6	93.8
Logged Tropical Forest	8347	27.5	71.6	285	26.7	55.8
Initial SS(1)	1694	33.0	66.0	0	---	---
Intermediate SS(1)	3673	29.8	66.0	0	---	---
Advanced SS(1)	7514	31.1	68.6	296	57.8	77.4
Inundated Palm Forest	7201	64.2	96.0	352	85.2	97.7
Reforestation	7108	55.8	87.0	412	40.5	55.6
Montane Old-growth Forest	8109	83.7	99.1	2115	71.6	95.4
Initial Montane SS(2)	589	55.5	84.7	494	15.8	58.9
Advanced Montane SS(2)	669	50.8	63.8	618	30.3	73.9
Dwarf Subalpine Forest	267	67.7	67.5	444	11.9	48.2
PS(3) on Montane Landslides	82	89.2	98.8	0	---	---
Mangroves	2513	93.0	99.2	490	95.1	95.9
Subalpine Paramo	1920	99.1	99.8	1782	86.9	93.0
Bare Soil	1821	97.4	99.9	160	70.0	93.8
Urban Areas	742	96.6	100.0	160	48.8	84.4
Water	10426	99.9	100.0	754	99.9	100.0
Clouds	3589	100.0	100.0	963	100.0	100.0
Shadow	4020	99.7	99.9	98	89.8	98.9
Overall Accuracy		74.6	91.9		68.7	89.0
Kappa		0.73	0.91		0.66	0.88

(1) Secondary Succession of tropical lowland forests after Finegan's model (1996)
(2) Secondary Succession of montane forests after Kappelle's model (1995)
(3) Primary succession

Table 2. Classification results using different band and index combination

	PTF	LTF	SS1	SS2	SS3	IPF	REF	MBP	MSS1	MSS2	DSAF	MPS	MANG	AVG F	FOR	Overall	Kappa
TM-TM1	32.4	17.9	18.5	17.7	20.6	62.3	47.2	79.0	24.7	30.0	61.3	62.7	89.5	41.7	76.3	69.09	0.668
TM	45.0	27.7	31.7	26.4	28.2	65.4	57.5	84.2	51.5	49.1	68.0	89.2	93.4	50.3	81.9	74.42	0.725
TM+b	45.5	27.9	32.2	26.8	28.7	64.8	57.1	84.5	52.5	49.7	69.4	89.2	93.8	50.5	82.3	*74.56*	0.726
TM+b+g	45.0	27.6	32.4	29.1	29.5	65.0	56.8	84.1	56.5	49.4	69.1	89.2	93.9	50.6	82.5	73.83	0.718
TM+b+g+w	45.4	28.6	29.8	28.7	30.1	64.6	54.2	84.0	54.6	51.9	69.6	88.0	94.1	50.4	82.4	73.61	0.716
TM+b+w	45.0	27.6	32.4	29.1	29.5	65.0	56.8	84.1	56.5	49.4	69.1	89.2	93.9	50.6	82.5	73.83	0.718
TM+g	44.8	28.0	28.2	26.8	28.8	65.5	55.2	83.8	50.7	51.1	68.5	90.4	93.6	50.0	82.0	73.91	0.719
TM+g+w	44.3	28.3	28.7	27.0	29.1	65.2	54.4	84.0	52.8	51.7	67.7	90.4	94.0	50.0	82.1	73.59	0.716
TM+ndvi	44.9	26.2	32.1	27.5	29.8	65.4	56.9	83.9	52.5	47.7	68.5	88.0	93.5	50.2	82.4	*74.44*	0.725
TM+ndvi+b	45.7	27.1	32.8	28.1	30.6	64.7	56.8	84.0	53.6	48.1	68.3	90.4	93.5	50.6	82.6	*74.55*	0.726
TM+ndvi+b+g	45.7	27.9	29.8	28.9	30.3	64.6	54.0	84.1	54.1	49.0	68.3	90.4	93.6	50.3	82.5	73.98	0.720
TM+ndvi+b+g+w	45.7	28.8	31.0	30.1	31.2	64.2	53.5	84.0	56.5	51.7	68.3	88.0	93.4	50.6	82.6	73.71	0.717
TM+ndvi+b+w	45.4	27.5	33.0	29.8	31.0	64.2	55.8	83.7	55.5	50.8	67.7	89.2	93.0	50.7	82.7	*74.60*	0.719
TM+ndvi+g	45.0	26.9	29.4	27.8	29.7	65.1	54.4	83.9	53.5	48.5	68.3	91.6	93.5	50.0	82.3	73.84	0.718
TM+ndvi+g+w	45.0	27.9	30.4	28.3	30.4	64.9	53.9	83.9	53.5	49.8	69.1	89.2	93.2	50.2	82.4	73.70	0.717
TM+ndvi+w	44.9	26.7	31.1	28.0	30.3	65.3	56.3	83.9	55.0	48.0	67.7	90.4	93.5	50.3	82.4	74.13	0.722
TM+w	44.6	27.9	30.9	27.7	28.6	65.5	57.1	84.0	55.1	49.9	68.3	90.4	93.9	50.4	82.2	74.12	0.722
Mod. Priors	82.0	71.6	66.0	66.0	68.6	96.0	87.0	99.1	84.7	63.8	67.5	98.8	99.2	82.3	96.5	91.98	0.914
max.	45.7	28.8	33.0	30.1	31.2	65.5	57.5	84.5	56.5	51.9	69.6	91.6	94.1	50.7	82.7	74.600	0.726
Mod. P. - max	36.3	42.8	33.0	35.9	37.4	30.5	29.5	14.6	28.2	11.9	-2.2	7.2	5.1	31.5	13.8	17.382	0.187
min.	44.3	26.2	28.2	26.4	28.2	64.2	53.5	83.7	50.7	47.7	67.7	88.0	93.0	50.0	81.9	73.587	0.716
Mod. P. - min	37.7	45.4	37.8	39.6	40.4	31.8	33.5	15.5	34.0	16.1	-0.3	10.8	6.2	32.3	14.7	18.395	0.198

TM	Landsat TM bands 1, 2, 3, 4, 5, 7	**g**	Tasseled Cap Greenness Index
NDVI	Normalized Difference Vegetation Index	**w**	Tasseled Cap Wetness Index
b	Tasseled Cap Brightness Index	**Mod. Prior**	Classification with modified prior probabilities
PTF	Undisturbed old-growth Tropical Forest	**MSS2**	Montane Advanced Secondary Succession
LTF	Logged Tropical Forest	**DSAF**	Dwarf Sub-Alpine Forest
SS1	Initial Secondary Succession	**MPS**	Montane Primary Succession
SS2	Intermediate Secondary Succession	**MANG**	Mangroves
SS3	Advanced Secondary Succession	**AVG F**	Average Forest Classification Self-Consistency
IPF	Inundated Palm Forest	**FOR**	Self-Consistency for lumped Forest Categories
REF	Reforestation	**Overall**	Overall Self-Consistency for all 33 LC-categories
MBP	Montane Old-growth Forest	**Kappa**	Overall Kappa of the training sites
MSS1	Montane Initial Secondary Succession		

5. Discussion

Classification results provided little evidence that spectral data enhancement techniques might be useful to improve the classification accuracy of secondary, logged, and undisturbed tropical forest categories in regions as large and complex as those of the selected study region.

The spectral patterns observed in training sites selected at similar sun illumination levels and in the same ecological life-zone suggest that ecologically meaningful succession categories have spatially variant spectral patterns, which is consistent with observed variations in vegetation structure, composition, and moisture within these categories. Such variations are well known from the ecological literature, which suggests that despite typical patterns of vegetative development, secondary succession stages, and thus very likely their corresponding spectral patterns, are strongly influenced by initial site conditions and land-use history.

In the humid tropics, vertebrates play an important role in seed dispersal from neighboring forests (Howe and Smallwood, 1993), remnant trees produce seeds or attract seed vectors (Guevara *et al.*, 1986), soil fertility influences growth rates (Uhl, 1987; Tucker *et al*, 1998), and land-use history determines capacity of soil seed banks to germinate, and the sprouting vigor of cut or crushed roots and stems (Uhl *et al.*, 1988). These initial site conditions affect the duration of the succession, species richness, and biomass accumulation rates. Because initial sites conditions can vary greatly within a region, forest age can not be used as a predictor of successional stage or to estimate any biophysical parameter such as biomass, tree height or species composition. Forest age is also difficult to determine in practical situations, because forest clearing and land abandonment result in a mosaic of secondary forests with each area representing a point in a succession time-continuum from cleared land to mature forest. From space-borne sensor systems, forest age is difficult to establish because frequent cloud-cover does not allow for continuous observation of the same point. Nevertheless, successional sequences are characterized by increases in leaf and wood biomass (Uhl, *et al.*, 1988; Brown and Lugo, 1990), canopy roughness (Foody *et al.*, 1996, Tucker, *et al.*, 1998), and changes in species composition (Finegan, 1992 and 1996). Such changes are thought to result in distinguishable patterns of spectral evolution of secondary successions. This hypothesis may hold true when observing the same forest patch over time. It may not necessarily apply when observing several forest patches at different locations and at the same time.

The question of what forest patch is observed, at when and where is relevant for drawing conclusions about patterns of spectral response. In areas with rugged terrain, variable moisture conditions, and advanced forest fragmentation (which implies presence of few large homogeneous forest patches and numerous small patches), extraction of spatially unbiased, normally distributed, and spectrally representative and separable training statistics can hardly be obtained. Signatures of selected homogeneous sites will not be representative of the true spectral nature of the cover types of interest. Analysis of spectral data obtained from such sites alone is likely to suggest spectral separability, which might be a correct conclusion for these sites, but not for the entire population of spectral patterns present in a large scene. In contrast, signatures extracted from randomly selected sites might be representative, but are more likely to show large variances and spectral overlap, a problem that the requirement of extracting the signatures from large training sites (at least 10 pixel per band according to Jensen, 1996) might increase. Analysis of such training data is less likely to produce separable spectral classes, but such data are representative of true spectral patterns.

In ecologically complex and fragmented landscapes, a fewer number of forest patches might be found, that exhibit the area and illumination requirements for adequate training sites and that correspond exactly with one and only one growth stage of the succession. More frequently, training sites of secondary forest categories will include some proportion of pixels belonging to other land cover categories, such as pasture, younger or older succession forest, and remnant trees. Variations in illumination caused by rugged topography might also increase the spectral variability of these sites. It is almost impossible to avoid including pixels which correspond to the canopy of large isolated trees in the training data for pasture, canopy openings in forested training sites, or early succession portions in sites that are predominately intermediate or advanced secondary successions.

In Costa Rica, large patches of homogeneous forest can rarely be found outside remaining fragments of undisturbed forest. This makes it particularly difficult to obtain spectral information for secondary and disturbed forest categoreis. On most farms, secondary forests, especially at the earlier stages, are a mix of small patches at different stages of development, with sometimes a single phase dominating the site. A similar observation can be made in logged forests that sometimes include patches of secondary succession where timber extraction has opened large portions of the forest canopy. In mature forest plantations it is not rare to observe invading secondary vegetation below the canopies of the trees, while in young, not well managed plantations, natural secondary vegetation might even dominate the site. Sites dominated by secondary vegetation might also include species that were established by the farm owner (e.g. *Coffea sp., Erithrina sp., Musa sp., Theobroma cacao, Macadamia sp., Citrus sp.*), and that continued to survive even after land abandonment. There also might be some remnant trees still standing from before the original forest was cleared for farming or grazing. The presence of such abandoned crops and mature trees alter the "typical" spectral response hypothesized for secondary successions.

The hypothesis that secondary successions evolve following a typical pattern of spectral change also implicitly assumes that there has been no human intervention in the abandoned sites. During fieldwork it was observed that intermediate and advanced secondary succession forests had been cleared a second time, but sometimes only partially, to utilize the tallest trees to shade new farming or grazing investments. Such sites are sometimes abandoned again, thus allowing a second wave of secondary succession to develop, while remnants of the first wave are still present.

Classification with modified prior probabilities allows for a certain level of control of the ecological and spectral complexity of a scene. The site specificity can be modeled and controlled through the stratification process, while the spectral complexity can be controlled through weighting the class spectra. In this way, pixels can only be assigned to classes that make ecological and geographical sense. Therefore, misclassification is possible only for site conditions where spectrally similar categories are equally likely to occur. With an adequate choice of ancillary variables, the number of strata where spectrally undistinguishable classes have similar prior probabilities can be minimized. In addition, within homogeneous site conditions, ecologically different vegetation categories have a greater chance of being spectrally dissimilar, since local factors that might alter biomass, species composition, and canopy roughness, and hence the spectral response, are less likely to be variant.

The accuracy by which the site conditions are modeled and the class prior probabilities are estimated, is of critical importance for improving land cover discrimination using the Bayesian classification approach. Finer stratification is certainly more suitable for controlling ecologically and spectrally relevant gradients, but the accuracy by which the prior probabilities can be estimated is negatively correlated with the number of strata. Further research is needed to establish the optimal relationship between the two.

7. Conclusions

The information content inherent in Landsat TM data is sufficient to map broad land cover categories, such as forested and non-forested areas. The spectral separability of such broad categories can be enhanced with NDVI and Tasseled Cap indexes under the assumption that spectral differences among categories are maintained throughout space. This assumption is not of general validity and can break down when the forest category includes subcategories such as deciduous and evergreen forest or secondary succession categories.

For detailed vegetation studies, such as the discrimination of tropical secondary forests, the assumption that vegetation categories maintain throughout space the same spectral patterns that were observed in training fields becomes less likely to be true. Ecological knowledge about the spatial variability of successional processes and the results of this research suggest that spectral patterns of succession categories might be complicated by specific site conditions, the history of site use, and other factors that are spatially variant. It appears, that the larger and complex the area to be studied, less likely it will be that secondary succession stages might be identified using spectral data alone.

Modeling the variance of site conditions in a GIS framework, appears thus to be a necessary complement of spectral patterns analysis. The Bayesian classification used in this research is an adequate approach to analyze spectral data and control site condition variability. However, the use of this method is certainly more time and resources demanding than traditional methods.

Nevertheless, the contribution that remote sensing research can make by providing refined data on land use and land cover change for inventories of GHG emissions, global change research, and other applications requires the adoption of more sophisticated data analysis methods.

Acknowledgements

This research was supported by a grant from the Swiss Organization for Development and Cooperation (INTERCOOPERATION). The Landsat TM image was provided by the Foundation for the Development of the Central Volcanic Mountain Range (FUNDECOR), the digital cartographic data by the Costa Rican National Geographic Institute (IGN) and the Department of Geography at the University of Costa Rica (UCR). Independent ground data were generously provided by Eileen Helmer, FUNDECOR, and Dirk Ressel. Martin Artavia and Geoffrey Venegas made a fundamental contribution by collecting training data in the field, and Mathias Tobler in Visual Basic programming. Finally, special thanks are due to the personnel of the GIS and Computer laboratory of the Tropical Agricultural Research and Higher Education Center (CATIE) for their logistical support.

References

Bolaños, R. A. M. and Watson, V. C., 1993. Mapa ecológico de Costa Rica. Centro Científico Tropical, San José (Costa Rica)

Boyd, D. S., Foody, G. M., Curran, P. J., Lucas, R. M. and Honzak, M., 1996. An assessment of radiance in Landsat TM middle and thermal-IR wavebands for the detection of tropical forest regeneration. International Journal of Remote Sensing, 17:249-261.

Brondizio, E. S., Moran, E. F., Mausel, P. and Wu, Y., 1994. Land use change in the Amazon Estuary: patterns of Caboclo settlement and landscape management. Human Ecology, 22:249-278.

Brondizio, E., Moran, E., Mausel, P. and Wu, Y., 1996. Land cover in the Amazon estuary: linking of Thematic Mapper with botanical and historical data. Photogrammetric Engineering and Remote Sensing, 62:921-929.

Brown, S. and Lugo, A. E., 1990. Tropical secondary forests. Journal of Tropical Ecology, 6:1-31.

Brown, S., Shvidenko, A. Z., Galinski, W., Houghton, R. A., Kasischke, E. S., Kauppi, P., Kurz, W. A., Nalder, I. A. and Roikov, V. A., 1996. WG2 Summary: Forests and the global carbon cycle: past, present, and the future role. In: M. J. Apps and D. T. Price (Editors), Forest Ecosystems, Forest Management and the Global Carbon Cycle. Springer-Verlag, Berlin and Heidelberg, pp. 199-208.

Campbell, J. B. and Bowder, J. O., 1992. SPOT survey of agricultural land uses in the Brazilian Amazon. International Arch. Photogrammetry Remote Sensing, Commission VII, 29:159-165.

Cohen, W. B. and Spies, T. A., 1992. Estimating structural attributes of Douglas-fir/western hemlock forest stands from Landsat and SPOT imagery. Remote Sensing of Environment, 41:1-7.

Cohen, W. B., Spies, T. A. and Fiorella, M., 1995. Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A. International Journal of Remote Sensing, 16:721-746

Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environemnt, 37:34-46.

Curran, P. J., Foody, G. M., Lucas, R. M. and Honzak, M., 1995. A methodology for remotely-sensing the stage of regeneration in tropical forests. In: P. M. Mather (Editor), TERRA-2 Understanding the Terrestrial Environment: Data Systems and Networks. Wiley, Chichester, pp. 189-202.

Dirzo, R. and Garcia, M. C., 1992. Rates of deforestation in Los Tuxtlas, a neotropical area in southeast Mexico. Conservation Biology, 6:84-90.

Dourojeanni, M., 1987. Aprovechamiento del barbecho forestal en áreas de agricultura migratoria en la Amazonia Peruana. Revista Forestal del Perú, 14:15-61.

Downton, M. W., 1995. Measuring tropical deforestation: development of the methods. Environemntal Conservation, 22:229-240.

Fearnside, P. and Guimaraes, W. M., 1996. Carbon uptake by secondary forests in Brazilian Amazonia. Forest Ecology and Management, 80:35-46.

Finegan, B., 1992. El potencial de manejo de los bosques húmedos secundarios neotropicales de las tierras bajas. Colección Silvicultura y Manejo de Bosques Naturales, No. 5. Tropical Agricultural Research and Higher Education Center (CATIE), Turrialba, Costa Rica.

Finegan, B., 1996. Pattern and process in neotropical secondary rain forests: the first 100 years of succession. Tree, 11:119-124.

Fiorella, M. and Ripple, W. J., 1993. Determining the successional stage of temperate coniferous forests with Landsat satellite data. Photogrammetric Engineering and Remote Sensing, 59:239-246.

Foody, G. M. and Curran, P., 1994. Estimation of tropical forest extent and regenerative state using remotely sensed data. Journal of Biogeography, 21:223-244

Foody, G. M., Palubinskas, G., Lucas, R. M., Curran, P. J. and Honzak, M., 1996. Identifying Terrestrial Carbon Sinks: Classification of successional stages in regenerating tropical forest from Landsat TM data. Remote Sensing of Environment, 55:205-216.

Gilruth, P. T., Hutchinson, C. F. and Barry, B., 1990. Assessing deforestation in the Guinea highlands of West Africa using remote sensing. Photogrammetric Engineering and Remote Sensing, 56:1375-1382.

Grainger, A., 1993. Rates of deforestation in the humid tropics: estimates and measurements. The Geographical Journal, 159:33-44.

Green, G. M. and Sussman, R., 1990. Deforestation history of the eastern rainforests of Madagascar from satellite images. Science, 248:212-215.

Guevara, S., Purata, S. E. and Van der Maarel, E., 1986. The role of remnant trees in tropical secondary succession. Vegetatio, 66:77-84.

Hall, F. G., Botkin, D. B., Strebel, D. E., Woods, K. D. and Goetz, S. J., 1991. Large-scale patterns of forest succession as determined by remote sensing. Ecology, 72:628-640.

Helmer, E. H., 1999. The Landscape Ecology of Secondary Tropical Forest in Montane Costa Rica. Ph.D. Thesis, Oregon State University, Oregon.

Hill, R. A. and Foody, G. M., 1994. Separability of tropical rain-forests types in the Tambopata-Candamo reserved zone, Peru. International Journal of Remote Sensing, 15: 2687-2693.

Holdridge, L. R., Grenke, W. C., Hatheway, W. H., Lang, T. and Tosi, J. A. J., 1971. Forest Environments in Tropical Life Zones: A Pilot Study. Pergamont Press, Oxford.

Howe, H. F. and Smallwood, J., 1993. Ecology of seed dispersal. Annual Review of Ecological Systems, 13:201-228.

Hutchinson, C. F., 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. Photogrammetric Engineering and Remote Sensing, 48:123-130.

Instituto Meteorológico Nacional, 1987. Atlas climatológico de Costa Rica. IMN, San José (Costa Rica).

Janzen, D. H., 1988. Management of habitat fragments in a tropical dry forest: growth. Annuals of the Missouri Botanical Garden, 75:105-116.

Jensen, J. R., 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. 2nd edition, Printice Hall, Upper Saddle River, New Jersey.

Jusoff, K. and D'Souza, G., 1996. Quantifying disturbed hill dipterocarp forest lands in Ulu Tembeling, Malaysia with HRV/SPOT images. ISPRS Journal of Photogrammetry and Remote Sensing, 51:39-48.

Kappelle, M., 1995. Ecology of Mature and Recovering Talamancan Montane *Quercus* Forests, Costa Rica. Ph.D. Thesis, University of Amsterdam/School of Environmental Sciences, National University, Heredia (Costa Rica).

Kappelle, M., Geuze, T., Leal, M. E. and Cleef, A. M., 1996. Successional age and forest structure in a Costa Rican upper montane *Quercus* forest. Journal of Tropical Ecology, 12:681-698.

Kartikeyan, B., Gopalakrishna, B., Kalubarme, M. H. and Majumder, K. L., 1994. Contextual techniques for classification of high and low resolution remote sensing data. International Journal of Remote Sensing, 15:1037-1051.

Kummer, D. M. and Turner, B. L., 1994. The human causes of deforestation in southeast Asia. BioScience, 44:323-328.

Kushwaha, S. P. S., Kuntz, S. and Oesten, G., 1994. Applications of image texture in forest classification. International Journal of Remote Sensing, 15:2273-2284.

Li, Y., Mausel, P., Wu, Y. and Brondizio, E., 1994. Discrimination between very advanced secondary succession and moist mature forest near Altamira, Brazil, using TM data. ASPRS Proceedings, 350-364.

Lucas, R. M., Honzak, M., Foody, G. M., Curran, P. J. and Corves, C., 1993. Characterizing tropical secondary forests using multi-temporal Landsat sensor Imagery. International Journal of Remote Sensing, 14:3061-3067

Ludeke A. K., Maggio, R. C. and L. Reid, L. M., 1990. An analysis of anthropogenic deforestation using logistic regression and GIS. Journal of Environmental Management, 31:247-257.

Lugo, A. E. and Brown, S., 1992. Tropical forests as sinks of atmospheric carbon. Forest Ecology and Management, 54:239-255.

MacArthur, R. H. and Wilson, E. O., 1967. Theory of Island Biogeography. Princeton Univ. Press, Princeton, New Jersey.

Maselli, F., Conese, C., De Filippis, T. and Romani, R., 1995. Integration of ancillary data into a maximum-likelihood classifier with nonparametric priors. Journal of Photogrammetry and Remote Sensing, 50:2-11.

Mather, P. M., 1985. A computationally-efficient maximum-likelihood classifier employing prior probabilities for remotely-sensed data. International Journal of Remote Sensing, 6:369-376.

Mausel, P., Wu, Y., Li, Y., Moran, E. F. and Brondizio, E. S., 1993. Spectral identification of successional stages following deforestation in the Amazon. Geocarto International, 4:61-71

Moran, E. F., Brondizio, E. and Mausel, P., 1994a. Secondary succession. National Geographic Research and Exploration, 10:458-476.

Moran, E. F., Brondizio, E., Mausel, P. and Wu, Y., 1994b. Integrating amazonian vegetation, land-use, and satellite data. Bioscience, 44:329-338.

Moran, E. F., Packer, A., Brondizio, E. and Tucker, J., 1996. Restoration of vegetation cover in eastern Amazon. Ecological Economics, 18:41-54.

Myers, N. 1988. Tropical deforestation and remote sensing. Forest Ecology and Management, 23:215-225.

Roy, P. S., Ranganath, B. K., Diwakar, P. G., Vohra, T. P. S., Bhan, S. K., Singh, I. J. and Pandian, V. C., 1991. Tropical forest type mapping and monitoring using remote sensing. International Journal of Remote Sensing, 12:2205-2225.

Sader, S. A. and Joyce, A. T., 1988. Deforestation rates and trends in Costa Rica: 1940-1983. Biotropica, 20:11-19.

Sader, S. A., 1995. Spatial characteristics of forest clearing and vegetation regrowth as detected by Landsat Thematic Mapper imagery. Photogrammetric Engineering and Remote Sensing, 61:1145-1151

Sader, S. A. , Waide, R. B., Lawrence, W. L. and Joyce, A. T., 1989. Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data. Remote Sensing of Environement, 28:143-156.

Sader, S. A., Stone, T. and Joyce, A. T., 1990. Remote sensing of tropical forests: an overview of research and applications using non-photographic sensors. Photogrammetric Engineering and Remote Sensing, 56:1343-1351.

Saldarriga, J.G., West, D. C., Tharp, M. C. and Uhl, C., 1988. Long term chronosequence of forest succession in the upper Rio Negro of Colombia and Venezuela. Journal of Ecology, 76:938-958.

Sawyer, J. and Lindsey, A., 1964. The Holdrige bioclimatic formations of the Eastern and Central United States. Proceedings of the Indiana Academy of Sciences, 73:105-112.

Saxena, K. G., Tiwari, A. K., Porwal, M. C. and Menon, A. R. R., 1992. Vegetation maps, mapping needs and scope of digital processing of Landsat Thematic Mapper data in tropical region of south-west India. International Journal of Remote Sensing, 13:2017-2037.

Scholz, D., Russell, J., Lindenlaub, J. and Swaun, P., 1977. A case study using ECHO for analysis of multispectral scanner data. The Laboratory for Applications of Remote Sensing, Purdue University, Purdue, North Carolina.

Sohn, Y., Moran, E. and Gurri, F., 1999. Deforestation in North Central Yucatan. Photogrammetric Engineering and Remote Sensing, 5:947-958.

Steiniger, M. K., 1996. Tropical secondary forest regrowth in the Amazon: age, area and change estimation with Thematic Mapper data. International Journal of Remote Sensing, 17:9-27

Stone T. A., Brown I. F. and Woodwell, G. M., 1991. Estimation, by remote sensing, of deforestation in central Rodonia, Brazil. Forest Ecology and Management, 38:291-304.

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

Swain, P. H. and Davis, S. M., 1978. Remote Sensing: The Quantitative Approach, McGraw-Hill, New York.

Tucker, J. M., Brondizio, E. S. and Moran, E. F., 1998. Rates of forest regrowth in eastern Amazonia: a comparison of Altamira and Bragantina regions, Pará state, Brazil. Interciencia, 23:64–73.

Tuomisto, H., Linna, A. and Kalliola, R., 1995. Use of digitally processed satellite images in studies of tropical rain vegetation. International Journal of Remote Sensing, 15:1595-1610.

Uhl, C., 1987. Factors controlling succession following slash-and-burn agriculture in Amazonia. Journal of Ecology, 75:377-407.

Uhl, C., Bushbacher, R. and Serrao, E. A. S., 1988. Abandoned pastures in eastern Amazonia. I. Patterns of plant succession. Journal of Ecology, 76:938-958.

Veldkamp E., Weitz, A. M., Staritsky, I. G. and Huising, E. J., 1992. Deforestation trends in the Atlantic zone of Costa Rica: a case study. Land Degradation and Rehabilitation, 3:71-84.

Westman, W. E., Strong, L. L. and Wilcox, B. A., 1989. Tropical deforestation and species endangerment: the role of remote sensing. Landscape Ecology, 3:97-109.

Woodwell, G M., Houghton, R A., Stone, T and Park, A., 1986. Changes in the areas of forests in Rodonia, Amazon Basin, measured by satellite imagery. In: J. Trabalka and D. Reichle (Editors), The Changing Carbon Cycle: A Global Analysis. Springer-Verlag, New York, pp. 242-257.

Woodwell, G. M., Houghton, R. A., Stone, T., Nelson, R. and Kovalick, W., 1987. Deforestation in the tropics: new measurements in the Amazon Basin using Landsat and NOAA Advanced Very High Resolution Radiometer Imagery. Journal of Geophysical Research, 92:2157-2163.

Part IV

Appendix

CONTENTS

Table 1. Classification consistency of 979 training sites (164,446 pixels) using classification with modified prior probabilities

(Band combinations: TM1, TM2, TM3, TM4, TM5, TM7, NDVI, Brightness, Wetness)

Values are percentages

Reference data

Classified data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Pasture	96.9	2.0	0.6	0.6	0.1	0.0	0.0	7.3	2.0	1.0	2.1	0.2	2.3	4.7	0.4	2.1	6.5
2. Annual Crops	0.2	97.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0
3. Plowed Soil	0.0	0.2	99.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4. Sugar Cane	0.0	0.2	0.4	98.8	0.0	0.0	0.0	0.2	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0
5. Ornamental Plants	0.0	0.0	0.0	0.0	96.6	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6. Pineapple	0.0	0.0	0.0	0.0	0.1	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7. Shadowed Ornamental Plants	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8. Mixed Agriculture	0.1	0.0	0.0	0.0	0.0	0.0	0.0	55.9	0.0	0.1	0.3	0.0	0.1	0.0	0.0	0.4	0.3
9. Bamboo	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	93.5	0.0	0.0	0.0	0.0	3.7	0.0	0.0	1.1
10. Banana	0.4	0.0	0.0	0.0	0.5	0.0	0.0	2.8	0.1	97.9	0.1	0.0	0.0	0.0	0.0	0.1	2.7
11. Fruits and Nuts Trees	0.3	0.2	0.0	0.2	0.7	0.0	0.0	1.7	0.4	0.3	85.8	0.1	0.1	0.2	1.1	2.0	1.3
12. Coffee	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	95.4	0.0	0.0	0.0	0.2	1.0
13. Oil Palm	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	94.5	0.0	0.0	0.2	0.0
14. Palmito Palm	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	2.4	0.0	0.0	0.0	0.0	90.1	0.0	0.0	0.3
15. Primary Lowland Forest	0.1	0.0	0.0	0.0	0.3	0.0	0.0	0.9	0.0	0.0	0.9	0.4	0.2	0.0	82.0	5.7	1.4
16. Logged Forest	0.3	0.0	0.0	0.0	0.3	0.0	0.0	8.2	0.1	0.1	2.9	0.7	0.9	0.0	6.9	71.6	4.3
17. Young Succession	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.6	0.8	0.3	1.0	0.4	0.1	0.9	0.3	1.2	66.0
18. Intermediate Succession	0.2	0.0	0.0	0.0	0.3	0.0	0.0	8.1	0.0	0.1	1.7	1.1	0.2	0.1	1.6	1.5	5.7
19. Advanced Succession	0.2	0.0	0.0	0.0	0.1	0.0	0.0	7.4	0.2	0.2	2.1	0.7	0.9	0.0	4.4	10.0	4.9
20. Inundated Palm Forest	0.1	0.0	0.0	0.0	0.3	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.0	0.0	0.6	0.5	0.5
21. Reforestation	0.4	0.2	0.0	0.0	0.3	0.0	0.0	6.3	0.6	0.1	2.2	0.8	0.8	0.3	0.6	3.8	4.2
22. Primary Mountain Forest	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.1	0.0
23. Initial Mountain Succession	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
24. Advanced Mountain Succession	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25. Dwarf Mountain Forest	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26. Primary Mountain Succession	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27. Mangroves	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
28. Paramo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
29. Bare Soil	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
30. Urban Areas	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31. Water	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
32. Clouds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33. Shadow	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Column total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Number of pixels	27270	2592	3007	7675	733	1032	905	633	3383	23022	4172	4958	3926	2811	7868	8884	1575

Classified data	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1. Pasture	3.9	2.0	0.1	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2. Annual Crops	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3. Plowed Soil	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
4. Sugar Cane	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5. Ornamental Plants	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6. Pineapple	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7. Shadowed Ornamental Plants	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8. Mixed Agriculture	0.8	0.5	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9. Bamboo	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10. Banana	1.0	0.4	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11. Fruits and Nuts Trees	2.2	1.8	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12. Coffee	0.9	0.2	0.0	0.7	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13. Oil Palm	0.4	0.7	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14. Palmito Palm	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15. Primary Lowland Forest	2.7	5.1	2.0	0.9	0.2	0.2	0.6	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16. Logged Forest	5.1	8.1	0.6	2.4	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17. Young Succession	2.3	1.6	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18. Intermediate Succession	66.0	5.6	0.2	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19. Advanced Succession	9.8	68.6	1.1	2.1	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
20. Inundated Palm Forest	0.7	0.7	96.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
21. Reforestation	4.2	4.4	0.1	87.0	0.0	1.0	0.3	0.5	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1
22. Primary Mountain Forest	0.0	0.0	0.0	0.1	99.1	6.4	32.6	27.7	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
23. Initial Mountain Succession	0.0	0.0	0.0	0.1	0.1	84.7	2.6	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24. Advanced Mountain Succession	0.0	0.0	0.0	0.0	0.4	5.6	63.8	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25. Dwarf Mountain Forest	0.0	0.0	0.0	0.0	0.1	0.8	0.0	67.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26. Primary Mountain Succession	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27. Mangroves	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.2	0.0	0.0	0.0	0.0	0.0	0.0
28. Paramo	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0
29. Bare Soil	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0
30. Urban Areas	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0
31. Water	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0
32. Clouds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0
33. Shadow	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.0	0.0	99.9
Column total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Number of pixels	3803	7714	7284	6635	7554	607	933	372	83	2532	1918	1823	742	10427	3584	4009

Overall accuracy: 91.9%; Kappa accuracy: 0.91

Table 2. Classification consistency of 979 training sites (164,446 pixels) using equal prior maximum likelihood classification

(Band combinations: TM1, TM2, TM3, TM4, TM5, TM7, NDVI, Brightness, Wetness)

Values are percentages

Reference data

Classified data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Pasture	90.4	6.8	1.6	2.7	0.3	0.0	0.0	7.3	13.2	2.0	5.2	1.5	7.6	7.9	1.3	3.5	11.7
2. Annual Crops	1.2	88.8	1.8	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.1	0.0	0.0	0.0	0.1
3. Plowed Soil	0.3	2.0	93.9	0.5	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
4. Sugar Cane	0.9	1.2	2.0	94.2	0.8	0.0	0.0	0.0	0.1	1.6	0.5	0.2	0.2	0.0	0.0	0.1	1.1
5. Ornamental Plants	0.0	0.0	0.0	0.1	70.1	0.0	0.0	0.6	0.0	1.2	0.3	0.0	0.3	0.0	0.1	0.0	0.6
6. Pineapple	0.0	0.0	0.3	0.0	0.1	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7. Shadowed Ornamental Plants	0.0	0.0	0.0	0.0	0.0	0.0	98.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8. Mixed Agriculture	0.1	0.0	0.0	0.0	0.5	0.0	0.0	25.1	0.3	0.7	0.8	0.1	0.3	0.0	0.2	0.9	0.5
9. Bamboo	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.3	65.9	0.0	0.1	0.1	0.0	7.8	0.1	0.0	2.0
10. Banana	0.6	0.0	0.0	0.5	19.0	0.0	0.0	7.6	0.0	81.9	4.5	0.2	3.3	0.0	1.1	1.0	1.7
11. Fruits and Nuts Trees	1.3	0.5	0.0	0.5	4.0	0.0	0.0	15.6	1.7	3.6	61.9	2.8	2.1	0.5	4.7	8.8	10.1
12. Coffee	0.4	0.0	0.0	0.1	0.0	0.0	0.0	1.7	1.3	0.1	1.9	77.6	0.1	0.6	1.8	1.4	2.0
13. Oil Palm	0.9	0.2	0.0	0.1	2.0	0.0	0.0	5.5	0.1	4.6	2.1	0.1	69.0	0.0	6.9	8.5	3.9
14. Palmito Palm	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.3	0.0	0.2	0.1	0.0	79.6	0.0	0.0	1.9
15. Primary Lowland Forest	0.2	0.2	0.0	0.0	0.1	0.0	0.0	2.4	0.1	0.8	2.0	2.0	3.3	0.1	45.4	12.2	3.7
16. Logged Forest	0.2	0.0	0.0	0.1	0.4	0.0	0.0	4.3	0.4	0.5	4.6	1.0	3.8	0.1	10.4	27.5	6.3
17. Young Succession	0.6	0.0	0.0	0.2	0.1	0.0	0.0	3.6	1.0	0.6	2.1	1.2	0.8	2.2	1.6	2.3	33.0
18. Intermediate Succession	0.6	0.1	0.0	0.1	0.3	0.0	0.0	10.4	0.4	0.4	2.6	2.1	1.2	0.1	4.7	7.2	7.7
19. Advanced Succession	0.3	0.0	0.0	0.0	0.1	0.0	0.0	9.0	0.4	0.2	4.1	3.6	4.0	0.0	9.4	15.0	5.4
20. Inundated Palm Forest	0.1	0.0	0.0	0.0	1.5	0.0	0.0	0.2	0.0	1.4	1.2	0.0	1.6	0.0	2.8	1.4	0.5
21. Reforestation	0.6	0.2	0.0	0.1	0.5	0.0	0.0	6.3	2.8	0.3	5.3	5.9	2.3	1.0	3.9	8.6	7.6
22. Primary Mountain Forest	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	1.8	0.7	0.0
23. Initial Mountain Succession	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	1.0	0.0	0.0	1.4	0.4	0.2
24. Advanced Mountain Succession	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.1	0.0
25. Dwarf Mountain Forest	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0
26. Primary Mountain Succession	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.1	0.0	0.0
27. Mangroves	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.1	0.0
28. Paramo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
29. Bare Soil	0.0	0.0	0.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
30. Urban Areas	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31. Water	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
32. Clouds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33. Shadow	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Column total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Number of pixels	27270	2592	3007	7675	733	1032	905	633	3383	23022	4172	4958	3926	2811	7868	8884	1575

Classified data	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1. Pasture	5.3	3.0	1.5	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0
2. Annual Crops	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3. Plowed Soil	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.8	0.0	0.0	0.0
4. Sugar Cane	0.1	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0
5. Ornamental Plants	0.1	0.1	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6. Pineapple	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7. Shadowed Ornamental Plants	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.9	0.0	0.1	0.0	0.1
8. Mixed Agriculture	1.1	0.7	0.1	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9. Bamboo	0.0	0.1	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10. Banana	1.0	1.2	8.6	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11. Fruits and Nuts Trees	9.3	8.1	4.7	7.7	0.1	0.5	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12. Coffee	3.8	2.1	0.0	2.4	0.2	4.9	0.1	0.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13. Oil Palm	2.1	7.9	6.5	2.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14. Palmito Palm	0.0	0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15. Primary Lowland Forest	9.3	10.0	5.0	3.3	3.4	11.9	8.3	1.9	1.2	2.1	0.0	0.0	0.0	0.0	0.0	0.1
16. Logged Forest	10.5	15.5	2.7	4.6	0.3	0.7	0.9	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
17. Young Succession	4.6	2.9	0.1	2.6	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18. Intermediate Succession	29.8	8.7	1.5	7.4	0.2	0.7	0.2	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
19. Advanced Succession	13.4	31.1	3.6	5.5	0.2	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20. Inundated Palm Forest	0.9	1.9	64.2	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0
21. Reforestation	8.2	6.4	0.4	55.8	1.3	6.6	3.8	3.5	0.0	2.6	0.0	0.0	0.0	0.0	0.0	0.1
22. Primary Mountain Forest	0.1	0.0	0.0	0.6	83.7	8.4	30.5	19.4	1.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0
23. Initial Mountain Succession	0.1	0.2	0.0	0.2	2.0	55.5	3.9	5.4	1.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0
24. Advanced Mountain Succession	0.0	0.0	0.0	0.3	7.6	5.1	50.8	1.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
25. Dwarf Mountain Forest	0.0	0.0	0.0	0.1	0.4	4.0	1.2	67.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26. Primary Mountain Succession	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.0	89.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0
27. Mangroves	0.1	0.1	0.4	0.1	0.0	0.0	0.0	0.0	0.0	93.0	0.0	0.0	0.0	0.0	0.0	0.0
28. Paramo	0.0	0.0	0.0	0.0	0.4	1.2	0.0	0.0	1.2	0.0	99.1	0.0	0.0	0.0	0.0	0.0
29. Bare Soil	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	97.4	2.3	0.0	0.0	0.0
30. Urban Areas	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	96.6	0.0	0.0	0.0
31. Water	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	99.9	0.0	0.0
32. Clouds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0
33. Shadow	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.1	0.0	0.0	0.0	99.7
Column total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Number of pixels	3803	7714	7284	6635	7554	607	933	372	83	2532	1918	1823	742	10427	3584	4009

Overall accuracy: 74.6%; Kappa accuracy: 0.73

Table 3. Classification accuracy of 252 control sites (15,206 pixels) using classification with modified prior probabilities

(Band combinations: TM1, TM2, TM3, TM4, TM5, TM7, NDVI, Brightness, Wetness)

Values are percentages

Reference data

Classified data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1. Pasture	95			0.62		0	0			0	5.28	0.38	1.76		0	0.35			0.34	0	4.13	0.19	4.05	0	0		0.2	0.73	1.88	0	0	0	0
2. Annual Crops	0.45			0.21		0	0			0	0	0	0		0	0			0	0	0.24	0	0.4	0	0		0	0	0	0	0	0	0
3. Plowed Soil	0			0		0	0			0	0	0	0		0	0			0	0	0	0	0	0	0		0	0	3.75	1.25	0	0	0
4. Sugar Cane	0.51			82.8		1.35	0			0	0	1.14	0		0.06	0			0	0	0	0	0.4	0	0		0	0.06	0	1.88	0	0	0
5. Ornamental Plants	0			0.41		0.34	0			0.56	0.41	0	0		0	0			0	0	1.7	0	0	0	0		0	0.06	0	0	0	0	0
6. Pineapple	0			0		98.3	0			0	0	0	0		0	0			0	0	0	0	0	0	0		0	0	0	0	0	0	0
7. Shadowed Ornamental Plants	0			0		0	95.8			0	0	0	0		0	0			0	0	0	0	0	0	0		0	0	0	0	0	0	0
8. Mixed Agriculture	0			0		0	0			0.28	0	0	0		0	0			0.34	0	0.24	0	0	0	0		0	4.99	0	0	0	0	0
9. Bamboo	0			0		0	0			0	0	0	0		0	0			0	0	0	0	0	0	0		0	0	0	0	0	0	0
10. Banana	0.23			0		0	0			96.3	0.41	0	0		0	0			0	0	0	0	0	0	0		0	0	0	0	0	0	0
11. Fruits and Nuts Trees	0.45			0.21		0	0			0	69.9	0.19	0.59		0	1.4			1.69	0	0.49	0	0	0	0		0	0.06	0	0	0	0	0
12. Coffee	1.25			1.66		0	0			0	0	97.2	0		0.18	0.7			0.34	0	8.98	0	17	0	0		0	0	0	1.25	0	0	0
13. Oil Palm	0			0		0	0			0	0	0	88.2		0	0			0	0	0	0	0	0	0		0	0	0	0	0	0	0
14. Palmito Palm	0.62			0		0	0			0	1.22	0	0		0	0			0	0	0.49	0	0	0	0		0	0	0	0	0	0	0
15. Primary Lowland Forest	0.11			0		0	0			0	3.25	0	1.76		93.8	21.1			0.68	2.27	1.94	0.47	6.48	0	0		3.27	0	0	0	0	0	1.02
16. Logged Forest	0			0.21		0	0			0	5.28	0	1.18		0.06	55.8			7.09	0	6.8	0	0.2	0	0		0	0	0	0	0	0	0
17. Young Succession	0.11			6.02		0	0			2.25	1.63	0	3.53		0	1.05			1.01	0	2.18	0	0	0	0		0	0	0	0	0	0	0
18. Intermediate Succession	0			7.68		0	0			0.56	4.47	0.19	0		0	2.81			9.8	0	2.67	0	0	0	0		0	0	0	0	0	0	0
19. Advanced Succession	0.06			0		0	0			0	2.85	0	1.76		0	10.5			77.4	0	1.21	0	0	0	0		0	0	0	0	0	0	0
20. Inundated Palm Forest	0			0		0	0			0	1.22	0	0		3.4	0			0	97.7	2.67	0	0	0	0		0	0	0	0	0	0	0
21. Reforestation	0.17			0.21		0	0			0	4.07	0.57	1.18		0.35	6.32			1.35	0	55.6	0	1.21	0	0		0	0.17	0	0	0	0	0
22. Primary Mountain Forest	0			0		0	0			0	0	0	0		0	0			0	0	0	95.4	4.45	22.2	43.5		0	0.34	0	0	0	0	0
23. Initial Mountain Succession	1.02			0		0	0			0	0	0.19	0		0.12	0			0	0	4.13	0.99	58.9	1.46	2.25		0	0.34	0	0	0	0	0
24. Advanced Mountain Succession	0			0		0	0			0	0	0	0		0	0			0	0	0.97	1.84	2.43	73.9	5.41		0	0.17	0	0	0	0	0
25. Dwarf Mountain Forest	0			0		0	0			0	0	0	0		0	0			0	0	0	0.09	0	1.13	48.2		0	0.06	0	0	0	0	0
26. Primary Mountain Succession	0			0		0	0			0	0	0	0		0	0			0	0	5.58	0.09	4.45	0.81	0.68		0	0	0	0	0	0	0
27. Mangroves	0			0		0	0			0	0	0	0		0	0			0	0	0	0	0	0	0		95.9	0	0	0	0	0	0
28. Paramo	0			0		0	0			0	0	0	0		0	0			0	0	0	0	0	0.49	0		0	93	0	0	0	0	0
29. Bare Soil	0			0		0	0			0	0	0	0		0	0			0	0	0	0	0	0	0		0	0	93.8	8.75	0	0	0
30. Urban Areas	0			0		0	0			0	0	0.19	0		0	0			0	0	0	0	0	0	0		0	0	0	84.4	0	0	0
31. Water	0			0		0	0.42			0	0	0	0		0	0			0	0	0	0.05	0	0	0		0	0	0.63	0	100	0	0
32. Clouds	0			0		0	0			0	0	0	0		0	0			0	0	0	0.76	0	0	0		0	0	0	2.5	0	100	0
33. Shadow	0			0		0	3.81			0	0	0	0		1.99	0			0	0	0	0.14	0	0	0		0.61	0	0	0	0	0	99
Column total	**100**	**0**	**0**	**100**	**0**	**100**	**100**	**0**	**0**	**100**	**100**	**100**	**100**	**0**	**100**	**100**	**0**	**0**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**0**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Number of pixels	**1763**	**0**	**0**	**482**	**0**	**297**	**236**	**0**	**0**	**355**	**246**	**528**	**170**	**0**	**1706**	**285**	**0**	**0**	**296**	**352**	**412**	**2115**	**494**	**618**	**444**	**0**	**490**	**1782**	**160**	**160**	**754**	**963**	**98**

Overall accuracy: 89.0%; Kappa accuracy: 0.88

Table 4. Classification accuracy of 252 control sites (15,206 pixels) using equal prior maximum likelihood classification
(Band combinations: TM1, TM2, TM3, TM4, TM5, TM7, NDVI, Brightness, Wetness)

Values are percentages

Reference data

Classified data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1. Pasture	86.3			2.07		0	0			1.13	3.25	2.08	4.71		0.12	0.7			1.35	0.85	2.91	0	3.04	0	0.68		0	4.43	0	3.13	0	0	0
2. Annual Crops	0.62			0.62		0	0			0	0	0	0		0.06	0			0	0	0.24	0	0	0	0.23		0	0	0	1.25	0	0	0
3. Plowed Soil	0			0		0	0			0	0	1.33	0		0	0			0	0	0	0.28	0	0	0		0.2	0	2.5	14.4	0	0	0
4. Sugar Cane	2.27			67.6		1.35	0			3.66	1.22	0.95	0		0.18	0			0	0	0.97	0.09	0.4	0	0		0	0.17	2.5	15.6	0	0	0
5. Ornamental Plants	0			10.2		0.34	0			0.28	0.81	0	0		0.06	0			0	0.28	1.7	0	0	0	0		0	0	0	0	0	0	0
6. Pineapple	0			0		98.3	0			0	0	0.19	0		0	0			0	0	0	0	0	0	0		0	0	0	0.63	0	0	0
7. Shadowed Ornamental Plants	0			0		0	94.9			0	0	0	0		0	0			0	0	0	0	0	0	0		0	0	0	0	0.13	0	7.14
8. Mixed Agriculture	0.11			0		0	0			0	1.63	0	0		0	0.7			1.01	0	0	0	0	0	0		0	1.35	0	0	0	0	0
9. Bamboo	0.23			0		0	0			0	0.81	0	0		0	0			0	0	0	0	0	0	0		0	0	0	0	0	0	0
10. Banana	0.23			16.6		0	0			84.5	5.28	0.95	16.5		1.88	0			0	6.25	1.94	0	0	0	0		0	1.07	0	0	0	0	0
11. Fruits and Nuts Trees	0.96			0.41		0	0			1.13	50.8	4.17	4.71		13	7.37			6.76	0.57	2.67	0.38	0	0	0		0	1.96	0	0	0	0	0
12. Coffee	5.05			0		0	0			0	2.03	79.9	0		0.06	1.05			4.39	0	28.4	0.76	21.5	0.49	1.13		0	1.68	0	1.25	0	0	0
13. Oil Palm	0.96			1.45		0	0			6.48	0.81	0	52.4		6.92	2.11			1.69	1.7	0.24	0	0	0	0		0	0.34	0	0	0	0	0
14. Palmito Palm	1.59			0		0	0			0	0	0	0		0	0			0	0	0.24	0	0	0	0		0	0	0	0	0	0	0
15. Primary Lowland Forest	0.11			0		0	0			0	4.07	0.38	4.71		34.6	14.7			2.03	2.84	5.1	4.02	24.9	6.47	0.45		1.22	0.06	0	0	0	0	1.02
16. Logged Forest	0			0.21		0	0			0.85	10.2	0.38	7.65		7.85	26.7			7.43	0	2.67	0.38	0.2	0.16	0		0	0	0	0	0	0	0
17. Young Succession	0.51			0.21		0	0			1.41	1.63	0.95	4.12		0.18	2.11			2.03	0	0.97	0	0.61	0	0		0	0	0	0	0	0	0
18. Intermediate Succession	0.4			0		0	0			0.56	6.1	1.52	0.59		2.29	10.9			10.1	0.57	0.73	0.05	1.21	0.16	0		0	0	0	0	0	0	0
19. Advanced Succession	0.11			0		0	0			0	3.66	0.38	1.18		10.4	19.6			57.8	1.14	2.91	0.09	4.25	0	0		0	0	0	0	0	0	0
20. Inundated Palm Forest	0			0		0	0			0	2.44	0	0		10.7	1.4			0	85.2	0.97	0	0	0	0		0.82	0.06	0	0	0	0	0
21. Reforestation	0.4			0.62		0	0			0	5.28	4.17	3.53		6.15	12.6			5.07	0.57	40.5	1.51	1.21	1.13	2.25		1.63	0.56	0	0	0	0	0
22. Primary Mountain Forest	0			0		0	0			0	0	0	0		0.06	0			0	0	2.18	71.6	10.5	40.1	39.9		0	0.34	0	0	0	0	2.04
23. Initial Mountain Succession	0.11			0		0	0			0	0	0.95	0		0.12	0			0	0	1.94	2.08	15.8	8.58	7.66		0	0.34	0	0	0	0	0
24. Advanced Mountain Succession	0			0		0	0			0	0	0	0		0	0			0	0	1.21	9.46	5.06	30.3	30.2		0	0.39	0	0	0	0	0
25. Dwarf Mountain Forest	0			0		0	0			0	0	0	0		0	0			0	0	0.24	7.8	0.81	8.09	11.9		0	0.22	0	0	0	0	0
26. Primary Mountain Succession	0			0		0	0			0	0	0.57	0		0	0			0	0	1.21	0.38	7.69	2.43	0.45		0	0.11	0	0	0	0	0
27. Mangroves	0			0		0	0			0	0	0	0		4.16	0			0.34	0	0	0	0	0	0		95.1	0	0	0	0	0	0
28. Paramo	0			0		0	0			0	0	1.14	0		0	0			0	0	0	0.43	0.4	2.1	5.18		0	86.9	0	0	0	0	0
29. Bare Soil	0			0		0	0			0	0	0	0		0	0			0	0	0	0	0	0	0		0	0	70	13.1	0	0	0
30. Urban Areas	0			0		0	0			0	0	0	0		0	0			0	0	0	0.14	0	0	0		0	0	23.8	48.8	0	0	0
31. Water	0			0		0	0.85			0	0	0	0		0	0			0	0	0	0	1.42	0	0		0	0	1.25	0	99.9	0	0
32. Clouds	0			0		0	0			0	0	0	0		0	0			0	0	0	0.28	0	0	0		0	0	0	1.88	0	100	0
33. Shadow	0			0		0	4.24			0	0	0	0		1.17	0			0	0	0	0.24	1.01	0	0		1.02	0	0	0	0	0	89.8
Column total	100	0	0	100	0	100	100	0	0	100	100	100	100	0	100	100	0	0	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100
Number of pixels	1763	0	0	482	0	297	236	0	0	355	246	528	170	0	1706	285	0	0	296	352	412	2115	494	618	444	0	490	1782	160	160	754	963	98

Overall accuracy: 68.7%; Kappa accuracy: 0.66

Classification scheme of secondary forests and its hypothesized relationship with spectral patterns

According to Finegan (1992) and Finegan (1996), where soil is not extremely degraded and seed sources are nearby, secondary succession in the humid neotropical lowlands may be described in terms of three phases:

Phase 1: During the first phase (initial succession) the abandoned land is colonized by herbaceous and pioneer shrubs that in the neotropics occupy the site for 2 to 3 years (5 in Eastern Amazonia, according to Moran *et al.*, 1996). Short-lived and shade intolerant pioneer tree species generally establish themselves at the beginning of the successional process, but do not dominate the site at this early stage of succession. In the Atlantic Zone of Costa Rica, the first years of succession can be dominated by *Phytolacca riviniana*, *Piper auritum* or *Veronia patens*. However, several other species can also be present at this stage: Moran *et al.* (1996) found up to 88 species (trees, shrubs and herbaceous) at their sample sites in Eastern Amazonia. In these sites, grasses and herbs, such as *Desmodium canum*, *Elephantopus mollis*, and *Acalypha arvensis*, were found spatially clumped and at high densities, while palms and pioneer shrubs, such as *Lantana camara*, *Wulffia baccata*, *Cyperus flavida*, and *Orbygnia phalerata*, were more evenly distributed. Irregular spatial distribution patterns of grasses and shrubs were also observed in this study.

The development of secondary vegetation changes the spectral response of abandoned pastures: according to Mausel *et al.* (1993), greater vegetation density and increased chlorophyll absorption make reflectance values in the visible portion of the spectrum lower than in clean pasture. The near-IR values increase because of mesophyll reflectance, while the middle-IR values decrease because of increased water absorption.

Phase 2: In the second phase of succession (intermediate succession) the first phase's plant community is replaced by a new tree community composed of a few dominant species that became established during the first phase. Typical tree species from this phase belong to the genera *Cecropia, Heliocarpus* and *Ochroma* (America), *Macaranga* and *Musanga* (Africa and Asia) and *Trema*, which is common on all three continents (Finegan, 1996). In less than 10 years, the height of the trees can reach 20 meters. In Costa Rica this phase terminates after approximately 10 years (Finegan and Sabogal, 1988a and 1988b), but in other regions (French Guyana) the tree population only begins to decline after 20 or more years (Sarrailh *et al.*, 1990). The second phase tree community disappears with increasing age because it can not regenerate under its own shade. During this phase, and sometimes from the beginning of the succession, a new community of tree species colonize the site. This community will then dominate the third phase of the succession.

According to Mausel *et al.* (1993), the visible reflectance values in this second phase of succession are similar to those from the first phase, but there is a higher green / red ratio because of increased leaf biomass (chlorophyll red / blue absorption and green reflectance). More shadow and water are present at the site as vegetative development progresses, and this decreases near-IR and especially middle-IR reflectance values (Mausel *et al.*, 1993).

Phase 3: The third phase of succession (advanced succession) begins with the growth of a tree population that is more shade tolerant and long-living than the declining second phase plant community. This new tree population can reach 25-30 cm of Breast High Diameter (BHD) after 10-15 years, and 50 cm of BHD after 25 years (Finegan and Sabogal, 1988a). Tree height can be the same as that of primary forests in less than 30 years. While several families can be found almost exclusively in primary forests (*Leguminosae*, *Moraceae*, *Lauraceae*, *Annonaceae*, in the neotropics, and *Chrysobalanceae*, *Sapotaceae*, and *Myristicaceae* in South America) few families are typical to advanced succession, among them are *Vochysiaceae* and *Tiliaceae*. Most of the dominant species in this phase belong to the genera *Rollinia* (*Annonaceae*), *Cordia* (*Boraginaceae*), *Guazuma* (*Ulmaceae*), *Stryphnodendron* (*Leguminosae*), and *Spondias* (*Anacardiaceae*). Frequent species are: *Didymopanax morototoni* (*Araliaceae*), *Goupia glabra* (*Celastraceae*), *Jacaranda copaia* (*Bignoniaceae*), *Laetia procera* (*Flacourtiaceae*) and *Simaruba amara* (*Simarubaceae*) (Finegan, 1996). Little is known about the destiny of advanced secondary succession, however, the species that dominate this phase of succession are generally different, and not as shade tolerant, as those found in primary forests. This leads to the speculation, that at some point in the future, the advanced secondary succession tree population will start to decline and to open the way to more shade tolerant species, which may, or may not, be the same as the species in primary forests. Moran *et al.* (1996) found mature forest species in third phase succession forests such as: *Neea floribunda, Cenostigma tocantinum*, and *Bertholletia excelsa*. It is likely that as they grow older, more primary forest species will be found in advanced secondary successions.

According to Mausel *et al.* (1993), the spectral characteristics of advanced succession forests approach those of mature forests: visible band reflectance is lower, and green / red ratio higher compared with early stages of succession. The IR band values continue to decline, probably as a result of increased shadow, compared with younger successional forests. However, primary forests should have more variable spectral responses than advanced succession forests because of large amounts of shadow and spectral traps caused by their complex multilayered and gapped vegetation structure.

This three-phase succession model described above provided the classification scheme for the field-level discrimination of succession forest sites in the basal, premontane and low montane tropical regions (*sensu* Holdridge *et al.*, 1971). During ground data collection, important differences were observed between successional sites belonging to the same ecological phase of the succession in terms of vegetation abundance, spatial distribution patterns of dominant species, dominant species composition, and site humidity. Differences in terms of species composition, growth rates, and time-frame, were also reported in other studies (Foody *et al.*, 1996; Uhl, 1987), and are probably not uncommon even for successional processes that occur at nearby sites. For example, at their study site near Manaus, Brazil, Foody *et al.* (1996) distinguished two distinct successional pathways. The taxa that composed each phase of the two pathways were different, especially in the early stages of succession. Sites that had been cleared and used briefly as pasture were dominated by *Cecropiaceae* species, whereas *Clusiaceae*, *Flacourtiaceae*, and *Melastomataceae* species were found on sites that had been cleared and burned, then used as pasture for several years. In this study, vegetation inventories of the training sites were not carried out. Therefore, compositional sub-categories of successional growth stages could not be separated objectively. Nevertheless, to prevent mixing spectral patterns of differently featured successional pathways, the signatures of the training sites were not merged prior to classification.

Outside the humid lowlands, differences in the succession processes are evident not only at the species level, but also in processes and patterns as well. For that reason, two additional secondary succession models were adopted for the Costa Rican dry zone and for the mountain regions. In the <u>dry forest zone</u>, seed dispersal and site colonization depend more on wind than vertebrates (Janzen, 1988; Sabogal, 1992). Wind dispersed seeds fall close to their parent trees (Harper, 1977; Howe and Smallwood, 1982), and this makes tree colonization of abandoned pastures that are far from forest patches or remnant trees difficult and slower than in the humid tropics. According to Janzen (1988), who investigated forest succession in the dry zone of Costa Rica, the first colonizing tree population of abandoned dry zone pastures is composed of long-lived wind-dispersed species. Therefore, the succession process in the dry zone lack an equivalent to the second phase of the succession model present in the humid low-land tropics

A two-phase succession model is also described by Kappelle (1995) and Kappelle *et al.* (1996), who studied secondary successions in the high mountains of the Talamanca Range in Costa Rica. Above 2000 meters elevation, species that are typical for even higher elevations can be found in early successional stages. In addition, species typical of mature forests, such as *Quercus*, *Weinmannia*, *Drymis*, *Phoebe*, *Ocotea* and others, can be found even in young successional forests. These species already dominate the second phase of mountain secondary succession.

Successional forests in the mountain region of the study area were classified following the two-phase model of Kappelle, while forests corresponding to the Janzen's model were not found, because the study area did not include the Costa Rican dry forest zone.

References cited

Finegan, B., 1992. El potencial de manejo de los bosques húmedos secundarios neotropicales de las tierras bajas. Colección Silvicultura y Manejo de Bosques Naturales, No. 5. Tropical Agricultural Research and Higher Education Center (CATIE), Turrialba, Costa Rica.

Finegan, B., 1996. Pattern and process in neotropical secondary rain forests: the first 100 years of succession. Tree, 11:119-124.

Finegan, B., and C. Sabogal, 1988a. El desarrollo de sistemas de producción sostenible en bosques tropicales húmedos de bajura. Un estudio de caso en Costa Rica. Parte 1. El Chasqui, 17:3-24

Finegan, B., and C. Sabogal, 1988b. El desarrollo de sistemas de producción sostenible en bosques tropicales húmedos de bajura. Un estudio de caso en Costa Rica. Parte 2. El Chasqui, 18:16-24.

Foody, G. M., G. Palubinskas, R. M. Lucas,P. J. Curran, and M. Honzak, 1996. Identifying Terrestrial Carbon Sinks: Classification of successional stages in regenerating tropical forest from Landsat TM data. Remote Sensing of Environment, 55:205-216

Harper, J. L., 1977. Population Biology of Plants. Academic Press, New York.

Holdridge, L. R., W. C. Grenke, W. H. Hatheway, T. Lang, and J. A. J. Tosi, 1971. Forest Environments in Tropical Life Zones: A Pilot Study. Pergamont Press, Oxford.

Howe, H. F., and J. Smallwood, 1993. Ecology of seed dispersal. Annual Review of Ecological Systems, 13:201-228.

Janzen, D. H., 1988. Management of habitat fragments in a tropical dry forest: growth. Annuals of the Missouri Botanical Garden, 75:105-116.

Jensen, J. R., 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. 2nd edition, Printice Hall, Upper Saddle River, New Jersey.

Kappelle, M., 1995. Ecology of Mature and Recovering Talamancan Montane *Quercus* Forests, Costa Rica. Ph.D. Thesis, University of Amsterdam/School of Environmental Sciences, National University, Heredia (Costa Rica).

Kappelle, M., T. Geuze, M. E. Leal, and A. M. Cleef, 1996. Successional age and forest structure in a Costa Rican upper montane *Quercus* forest. Journal of Tropical Ecology, 12:681-698.

Mausel, P., Y. Wu, Y. Li, E. F. Moran, and E. S. Brondizio, 1993. Spectral identification of successional stages following deforestation in the Amazon. Geocarto International, 4:61-71.

Moran, E. F., A. Packer, E. Brondizio, and J. Tucker, 1996. Restoration of vegetation cover in eastern Amazon. Ecological Economics, 18:41-54.

Sabogal, C., 1992. Regeneration of tropical dry forests in Central America, with examples from Nicaragua. Journal of Vegetation Science, 3:407-416.

Sarrailh, J., *et al.*, 1990. Études sur la mise en valeur de l'écosysthème forestier guzanais après déboisement. Bois et Forêts des Tropiques, 219:79-97.

Uhl, C., 1987. Factors controlling succession following slash-and-burn agriculture in Amazonia. Journal of Ecology, 75:377-407.

The hypothesis of spectral separability: a critique

Vegetation classification using Landsat TM data is based on the theory that chlorophyll absorbs red radiation, leaf mesophyll reflects near infra-red (IR) radiation, moisture reduces near-IR and especially middle-IR reflectance, and shadows reduce reflectance in all wavelengths, especially longer wavelengths. Increased green leaf biomass should therefore be negatively correlated with remotely sensed radiation in the red wavelengths, and positively correlated with the near-IR wavelengths (Jensen, 1996).

Since leaf biomass is generally positively correlated with wood biomass, increased red absorption and near-IR reflectance should be correlated with high biomass vegetation types, such as forests (Foody and Curran, 1994). NDVI, that expresses the contrasting patterns of green vegetation in the red and near-IR wavelength portions of the spectrum, should therefore be particularly useful in detecting green vegetation and green leaf biomass. However, chlorophyll content and mesophyll structure are different among species, and may not correlate well with wood biomass, for example in tropical regions that exhibit deciduous and evergreen forest types during the dry season. Increased shadow cast by emerging trees, spectral traps caused by canopy gaps, and increased humidity resulting from recent local rainfall or particular local soil conditions can reduce near-IR reflectance as well, but are not necessarily correlated with vegetation biomass. The hypothesized relationship between remotely sensed red and near-IR radiation and vegetation biomass could therefore vary in space and time. Because even in the humid tropics some tree species drop some or all of their leaves during the dry season (e.g. *Gmelina arborea*), the assumption is also invalid after the wet season. This is particularly important for remote sensing research in the humid tropics, since cloud-free Landsat TM imagery is available only during the dry season. Scenes with exceptionally little cloud coverage in a tropical region that is usually covered by clouds might have been acquired during a period of exceptional drought. This may imply particular water stress conditions for the vegetation, and possibly altered spectral patterns, at least for those sites that do not have a naturally or artificially augmented water supply. Some stress conditions, such as drought, pests, disease, and leaf fall caused by increased wind, fire, and other reasons, can also substantially alter the leaf / wood biomass relationship and the corresponding spectral response in the visible and IR portions of the spectrum. Nevertheless, the spectral classification of forest categories relies on the assumption that spectral patterns observed at the patch level (e.g. in the training sites) are representative of the entire population of spectral data for these categories within a particular scene.

Ecological theory describes the successional process in terms of discrete phases that have distinctive structural and compositional characteristics (Saldarriga *et al.*, 1988; Janzen, 1988; Kappelle, 1995; Finegan, 1996). Such discrete phase models are appealing for mapping purposes, since they relate ecological phase transitions to increases in leaf and wood biomass (Uhl, *et al.*, 1988; Brown and Lugo, 1990), canopy roughness (Foody *et al.*, 1996, Tucker, *et al.*, 1998), and changes in species composition (Finegan, 1992 and 1996) that should result in different patterns of spectral response. However, the nature of secondary succession is transitional both within and between succession categories. Spatial variations in composition and structure between forest patches at the same successional phase have been shown to have different spectral responses even between sites located closely together (Foody *et al.*, 1996). In studies covering large tropical regions, it is unlikely that sufficient ground data can be made available to classify within class variations in ecologically meaningful and spectrally separable sub-categories. In addition, spectral changes of secondary forests should be continuous, rather than abrupt, if they relate to increases in leaf biomass, wood biomass, canopy roughness, and changes in species composition. Sharp spectral boundaries between successional categories are therefore unlikely to exist, even if the average mean values of the spectral data observed for the training sites may be statistically different. However, abrupt changes in the spectral response of a particular secondary forest site might occur after a few days of drought, local rainfall, fire, and other factors that are not related to the features observed for ecological classification of the vegetation.

Sharp boundaries between successional stages are also difficult to draw in the field, especially when looking at the patch size required for signature extraction. The patch size required for the development of training statistics for Landsat TM data is of 5.68 ha following the rule of thumb of 10 pixels per band (Jenzen, 1996). Under the conditions found in central Costa Rica, rather than homogeneous sites, secondary vegetation patches of that size often present a mixture of different secondary growth stages and other land cover categories. Their spectral complexity is sometimes increased by the presence of unevenly distributed remnant trees from the original forest, abandoned herbaceous and ligneous species established by the landowner, living fences, and, in rugged terrain, variations in sun illumination. Training statistics developed from such sites are too noisy to be useful for discriminating secondary succession stages. On the other hand, signatures extracted from training sites selected for their atypical homogeneity of the vegetation and topography are unlikely to represent the true spectral data distributions present in the whole scene.

Despite these problems, a high level of consistency can be achieved in the ecological classification of secondary vegetation sites at the field level. Trained personnel observe different features of the vegetation to classify the site, and with some practice, visual field level classification can be achieved consistently. In contrast, spectral data describe only one feature of the forest. They may, or may not, be correlated with ecological variables observed for field-level vegetation classification (Foody and Curran, 1994). Few studies have managed to record detailed ground data as well as remotely sensed data on the same date, to test for a correlation between the two data sets (Box, *et al.*, 1989; Cook, *et al.*, 1989; Foody and Curran, 1994; Tucker, *et al.*, 1998). When this did occur, correlations observed were generally weak (Foody and Curran, 1994). But most remote sensing research in tropical regions assumes that spectral differences among forest categories are: inherent to the categories, consistently maintained throughout space, and accurately captured by the remote sensor. The use of spectral data enhancement techniques, such NDVI and Tasseled Cap transformation, is based on these three assumptions. Therefore, these techniques might be useful for remotely sensed data sets – and more likely in small subsets (which is often the case of hypothesis testing in remote sensing research) – where these assumptions hold true. How these assumptions can be verified before classification has as yet not been addressed. However, if the categories of interest do not have distinctive spectral patterns, or their signatures can be separated only under particular site conditions, spectral data enhancement techniques would not be useful for classifying the entire data set without previous scene stratification and signature weighting

References cited

Box, E. O., B. N. Holben, and V. Kalb, 1989. Accuracy of the AVHRR vegetation index as a predictor of biomass, primary production and net CO2 flux. Vegetatio, 80:71-89.

Brown, S. and A. E. Lugo, 1990. Tropical secondary forests. Journal of Tropical Ecology, 6:1-31.

Cook, E. A., L. R. Iverson, and R. L. Graham, 1989. Estimating forest productivity with Thematic Mapper and biogeographical data. Remote Sensing of Environment, 28:131-141.

Finegan, B., 1992. El potencial de manejo de los bosques húmedos secundarios neotropicales de las tierras bajas. Colección Silvicultura y Manejo de Bosques Naturales, No. 5. Tropical Agricultural Research and Higher Education Center (CATIE), Turrialba, Costa Rica.

Finegan, B., 1996. Pattern and process in neotropical secondary rain forests: the first 100 years of succession. Tree, 11:119-124.

Foody, G. M., and P. J. Curran, 1994. Estimation of tropical forest extent and regenerative state using remotely sensed data. Journal of Biogeography, 21:223-244.

Foody, G. M., G. Palubinskas, R. M. Lucas, P. J. Curran, and M. Honzak, 1996. Identifying Terrestrial Carbon Sinks: Classification of successional stages in regenerating tropical forest from Landsat TM data. Remote Sensing of Environment, 55:205-216.

Janzen, D. H., 1988. Management of habitat fragments in a tropical dry forest: growth. Annuals of the Missouri Botanical Garden, 75:105-116.

Jensen, J. R., 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. 2nd edition, Printice Hall, Upper Saddle River, New Jersey.

Kappelle, M., 1995. Ecology of Mature and Recovering Talamancan Montane *Quercus* Forests, Costa Rica. Ph.D. Thesis, University of Amsterdam/School of Environmental Sciences, National University, Heredia (Costa Rica).

Saldarriga, J.G., D. C. West, M. C. Tharp, and C. Uhl, 1988. Long term chronosequence of forest succession in the upper Rio Negro of Colombia and Venezuela. Journal of Ecology, 76:938-958.

Tucker, J. M., E. S. Brondizio, and E. F. Moran, 1998. Rates of forest regrowth in eastern Amazonia: a comparison of Altamira and Bragantina regions, Pará state, Brazil. Interciencia, 23:64–73.

Uhl, C., R. Bushbacher, and E. A. S. Serrao, 1988. Abandoned pastures in eastern Amazonia. I. Patterns of plant succession. Journal of Ecology, 76:938-958.