Land Use Zones and Land Use Patterns in the Atlantic Zone of Costa Rica

A Pattern Recognition Approach to Land Use Inventory at the Sub-Regional Scale, using Remote Sensing and GIS, applying an Object-Oriented and Data-Driven Strategy

Promotoren:

Dr. Ir. M. Molenaar
hoogleraar op het vakgebied landmeetkunde en teledetectie

Dr. Ir. J. Bouma
hoogleraar in de bodeminventarisatie en landevaluatie, speciaal gericht op de (sub)tropen

STELLINGEN

1. Land use at the sub-regional level can be described by means of a set of spatial objects. The object-oriented approach enables a structured description of land use and land use change, which is lacking in most present approaches to land use inventory. *(This thesis.)*

2. The possibilities for discriminating between land cover types only on the basis of spectral reflectance characteristics are limited. Major improvements to spectrally separate land cover types are not to be expected by increasing the spectral and spatial resolution of satellite imagery. The integration of remote sensing data with other types of data, employing a process of reasoning for information extraction whereby the proper use of rules of inference is of great importance, is more promising in this respect.

3. The interpretation of (remotely sensed) data for the extraction of information on terrain objects always refers to a particular context. The definition of context to consist of: (a) a thematic context that is reflected in the classification system and attribute structure of objects; (b) a spatial context that is reflected in the levels of aggregation and the used scale levels; (c) the purpose of mapping and the aspect of time, enables a meaningful description of the context and the structured use of context information.

4. Reconnaissance soil surveys do not provide data with sufficient detail to be relevant for evaluating land use with respect to the bio-physical land potentials at farm - or field level. On the other hand, the information that is provided is too specific for the evaluation of land use at sub-regional level. The goals and practice of the reconnaissance soil survey should be reconsidered.

5. Scientific development in the field of land use inventory studies and rural surveys is hampered by the lack of a central paradigm.

6. Land evaluation and farming system analysis (LEFSA) provide information on land use, through land and land use inventory and diagnosis of problems associated with land use. The function of information in the planning process and plan execution has to be defined as yet. This function can only be determined through analysis of the planning process.

7. With the technical capabilities of modern GIS systems for handling and processing spatial data, the interpretation of results and judgement of their validity has become a primary problem. The required expertise is generally lacking with the operators of GIS systems and with planners who might base their decisions on the information that is generated.

8. Zolang er 'oerrunderen' grazen in door onszelf gecreëerde natuur en transgene stieren volkomen natuurlijk ogen zal een duurzaam milìeu een illusie zijn.

Tom Lemaire (1970), Filosofie van het landschap, Uitgeverij AMBO.
Hans Achterhuis (1992), De illusie van groen. Over milieucrisis en de fixatie op techniek, Uitgeverij de Balie.

9. Wetenschappelijk onderzoek is eigenlijk een poging van wetenschappers om die situatie te scheppen waarin hun ideeen het duidelijkst bevestigd worden. Dit subjectieve aspect wordt vaak ontkend door het suggereren van objectiviteit. Echter juist de erkenning daarvan bevordert de objectiviteit.

Ilya Prirogine and Isabelle Stengers (1990), Orde uit Chaos. Uitgeverij Bert Bakker.

10. Wetenschap is een strijd geworden om geld, faciliteiten en plaats voor publicaties in tijdschriften. De condities die daarvoor gelden werken eerder verstikkend op de wetenschap dan dat deze bijdragen tot vernieuwende inzichten.

Henry H. Bauer (1992), Scientific literacy and the myth of the scientific method, University of illinois press.

11. Onderzoekers die in de tropen gestationeerd zijn hebben de neiging harder te werken en/of meer op vakantie te gaan. Beide zijn redenen om onderzoek in de tropen te bevorderen.

12. Hoe groter de afstand van waaraf wij het aards oppervlak waarnemen des te onbeduidender lijkt alles te worden wat zich op dat oppervlak afspeelt.

13. Computers lack a body to be intelligent.

Hubert L. Dreyfus (1979), What computers can't do. The limits of artificial intelligence, Harper & Row publishers, New York.

Stellingen behorende bij het proefschrift " Land use zones and land use patterns in the Atlantic zone of Costa Rica" van Jeroen Huising, te verdedigen op 3 mei 1993 in de aula van de Landbouw Universiteit Wageningen.

Jeroen Huising

Land Use Zones and Land Use Patterns in the Atlantic Zone of Costa Rica

A Pattern Recognition Approach to Land Use Inventory at the Sub-Regional Scale, using Remote Sensing and GIS, applying an Object-Oriented and Data-Driven Strategy

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen
op gezag van de rector magnificus,
dr. H.C. van der Plas,
in het openbaar te verdedigen
op maandag 3 mei 1993
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen

Cover illustration by Gwendolyn Huising

CIP-gegevens Koninklijke Bibliotheek, Den Haag

Huising, Jeroen

Land use zones and land use patterns in the Atlantic zone of Costa Rica: a pattern recognition approach to land use inventory at the sub-regional scale using remote sensing and GIS, applying an object-oriented and data-driven strategy / Jeroen Huising. - [S.l. : s.n.] Thesis Wageningen. - With ref. - With summary in Dutch.
ISBN 90-5485-106-6
Subject headings: land use / land cover / remote sensing / pattern recognition / GIS / object-oriented / land use evaluation.

Printed by Ponsen & Looijen, Wageningen

Abstract

This thesis describes an approach to land use inventory at the sub-regional scale in the Guacimo-Rio Jiménez-Siquirres (GRS) area in the Atlantic Zone of Costa Rica. Therefore, the concept of "land use zones" is introduced. The land use zone (LUZ) plays a central role in the definition of an observational methodology as well for structuring dynamics in land use. Land use is described in terms of the land use pattern (LUP). The LUP denotes the farming systems and land utilization types (LUTs) occurring within a land use zone.

This thesis formulates a methodology for the inventory of land use and land use change that is object-oriented and data-driven. "Object-oriented" means that land use is expressed in terms of a collection of objects (land use zones) with specific geometric and thematic characteristics. A classification system is developed so that each class contains land use zones with a characteristic thematic description, geometry, aggregation structure and dynamics. The handling of such complex object information requires that emphasis is put on the definition of a data model.

For inventory purposes satellite imagery and aerial photos are used. The use of these materials involves pattern recognition. The "data-driven" approach in this case means that the classes to describe land use are not a-priori but inductive, i.e. they result from the inventory process. The data-driven approach is a strategy to gain insight in the sub-regional land use expressed in the land use patterns. The complex land use inventory process is unravelled into a number of sequentially ordered processing steps, described in the various chapters. This thesis consists of three parts.

The first part of this thesis defines LUZs as a tool for the inventory of land use and land use change. From the comparison of aerial photos of 1948-1952 and 1984 we learn that the LUZs, that belong to the agricultural area, have stable boundaries. This implies that the land use zone may serve as a reference area for monitoring land use change.

Data on farm size distribution and farming system composition of a number of zones were obtained by means of a farm survey. The data show significant differences in the farming system composition, on the basis of which we define the land use patterns. That these clear differences occur indicates that the LUZ serves as a spatial unit for the land use inventory at the sub-regional level. The differences in LUP relate to differences observed in land cover composition and farm size distributions. This relation indicates that information on LUP may be inferred from composite land cover and farm size characteristics, so that satellite imagery and aerial photos can be used as tools for land use inventory, if the proper rules for interpretation are applied.

The correspondence between LUP and composite land cover implies that change in LUP may be inferred from change in land cover composition, under the condition that the geometric characteristics do not change. Change in land cover composition of LUZs between 1986 and 1990 was investigated using satellite imagery. Clear trends in land use change were observed, when the proper interpretation rules are applied. These trends were a decrease in the area for the cultivation of maize and in pasture land, and an increase in the area for banana and macadamia production and reforestation. Besides changes in area of crops, a change in the condition of pastures and banana plantations could be indicated.

The second part describes the pattern recognition process. This concerns the identification and classification of the LUZs. First, the stratification of the GRS area into sub-regions is described. The spatial pattern, which is determined by the size, form and arrangement of the agricultural fields, is used as a key to the aerial photo interpretation and as a criterion to identify the LUZs.

Once the LUZs are identified, their field size characteristics and the land cover composition are determined. A procedure is described for the per pixel land cover classification. This procedure will guide the image analyst in the complex task of defining a set of training classes with statistical properties suitable for the maximum likelihood classification. Emphasis is on the training phases. The procedure presented here makes use of supervised as well as unsupervised approaches.

Special attention has been paid to the definition of LUZ classes. Statistical methods are used to identify and define the different patterns as a key to classification. Field size characteristics of the LUZs were determined. One-way analysis of variance and multiple comparison were used to evaluated differences in mean field size. This resulted in the definition of five classes for mean field size.

A hierarchical cluster analysis was performed to evaluate the difference in land cover composition between the LUZs. To derive the relevant groups of corresponding LUZs from the results (represented by a dendrogram) a critical distance is defined. The critical distance denotes the minimum distance at which LUZs (or groups of LUZs) are considered significantly different with respect to their land cover composition. The critical distance reflects accuracy of data on land cover composition, which is determined by the accuracy of the land cover classification and the geometric accuracy of the land cover map and the LUZ map. The resulting composite land cover classes provide information on land use of the LUZs.

Land use information is obtained by interpretation of the land cover composition and field size characteristics of the LUZs. This interpretation involves the transformation of data classes into information categories by using mapping rules (also termed decision rules). Mapping rules assign a conditional class label to an object, whereby the condition refers to a particular context. The mapping involves complex decisions. Insight in the complex decision structure is gained by putting the decision rules in hierarchical order. The result is a decision tree for the classification of LUZs in terms of LUPs. The decision tree leads to stepwise classification of LUZs. The decision tree provides a formalized description of the decisions in the classification of the LUZs.

In Part Three the land use in the GRS area is evaluated with respect to bio-physical land potentials. The LUZ map and the physiographic soil map were combined. The soil unit boundaries and the land use zone boundaries corresponded to a high degree. But this does not mean that land use is in agreement with the (bio-physical) land potentials. Results show that 18 % of the GRS area is at risk of land degradation, while 51 % of the area is considered to have potential for more intensive use. Expert judgement is used to determine the suitability of the soil types for specific land utilization types (LUTs). However, the exact position of the soil type or the LUT cannot be determined at a sub-regional scale, with units being composite in nature. This introduces a fundamental uncertainty with respect to the statements on land use suitability. The study, therefore, has an exploratory character. The figures denote expectations.

In the last chapter the variation in banana yield within one plantation (representing a particular LUZ) is investigated. Soil survey data explained 67 % of the variation. Combining Landsat-TM and soil data did not provide a better estimation of yields. The explained variation remained 67 %.

Preface

The main part of this thesis deals with land use inventory. The original idea had a broader scope. It incorporated also soil suitability assessment and agro-ecological evaluation as part of a GIS for land use planning. This thesis shows that land use inventory in itself is a topic large enough to spend almost 5 years on. The soil, as such, plays a minor role in this thesis, though I do think that my experience as a soil scientist in soil survey strongly shaped my attitude towards land use inventory. This is reflected, for example, in the use of aerial photos. With respect to land use inventory emphasis has been primarily on data acquisition. In the beginning attention was focused on the use of satellite imagery, to be shifted later to data modelling and object classification.

Priorities change during the research as consequence of insights gained in the regional setting and in the subject itself. Problems are encountered and this further steers the line of investigation. Doing research is in fact a kind of struggle to come to grips with the subject of interest, to fulfil conditions required for the research and to put everything on paper. The struggle was harder and took more time than I had imagined. At this point I am relieved to conclude this research, but at the same time anxious to further explore the possibilities the theory offers. I hope the reader may acknowledge the importance of the problems discussed in this thesis and can make use of my experience to his or her own advantage.

I want to thank my two supervisors. Prof. Martien Molenaar for his confidence. The discussions we had I found very stimulating and rewarding. He had to read the various draft versions of this thesis. His comments were very valuable. Prof. Johan Bouma thoroughly read and commented on the various versions of my many chapters. His critical attitude inspired me to put an extra effort in improving quality of the output, at times when enthusiasm failed me. Also his contribution to the chapters in Part Three of this thesis are very much appreciated. Their support and guidance also in the early phase of the project is thankfully acknowledged.
My companions at the department of Surveying, Photogrammetry and Remote Sensing I thank for the pleasant atmosphere and support during the time I spent at the department. Especially I want to mention John Stuiver who installed the necessary software in Costa Rica and helped me to overcome hardware problems. His laughter soothed the many frustrations I experienced at that time. He also introduced me to ARC-INFO and other software that I needed once returned to Wageningen.
Part of my work I carried out at the department of Soil Science and Geology. I want to thank all who were involved in one way or the other in my research. Especially I want to name Dr. M.A. Mulders, who arranged BCRS funding for the purchase satellite imagery and the production of the poster. He also assisted me in doing my first field work in Costa Rica. The BCRS is acknowledged for their financial support.

My research was part of a larger multi-disciplinary research project in the Atlantic Zone of Costa Rica (the Atlantic Zone Programme), a programme of cooperation between the agronomic centre for research and training (CATIE) in Costa Rica, the Wageningen Agricultural University (WAU) and the costarican ministry of agriculture (MAG). Much gratitude I owe to the people of the project in Costa Rica for their support and friendship: Fernando Cambronero, Luis Quiros, Quillermo Valverde, Olga Carvajal and Celia Alfaro. My colleagues Henk Waaijenberg, Willem Wielemaker and Andre Nieuwenhuysen. Dr. Jan Wienk, the initial coordinator of the programme, helped to realize the necessary conditions for my work. He and Liesbeth van der Ziel are thanked for their hospitality. The same gratitude I want to express to Hans Bronkhorst, who succeeded Jan Wienk as coordinator, and his wife Joke.

At CATIE, Turrialba most of the image processing work was done. I want to thank all at the department 'centro computo', especially Javier Saborio, the system manager, for all the support he gave me in using the Data General and later the Compacq. Hilda Piaggio helped me using SAS.

It was a pleasure to work with the MSc. students that participated in my research: Sytze de Bruin, Chris Stiggelbout, Fred Stolle, Rob Hootsman, Paul Rümkens, Ed Veldkamp and Rob Tan.

Don Juan Ernesto Schroeder and doña Heide accommodated me while I was doing my field work in the north. I enjoyed their company.

Mirdy Naeff did a marvellous job correcting the syntax and spelling of my english and making redactional improvements to the text. We had long (nightly) session going over the first part of the thesis. Drs. Nancy Smith-van Weesep I owe for checking and improvements on my english of the second part.

Willemien Brooijmans had to endure my many moments of frustration and anger, she was also there to share moments of satisfaction and joy. She opened my eyes for many things in Costa Rica which I otherwise would not have seen and contributed as such to the success of my endeavour.

TABLE OF CONTENT

TABLES

FIGURES

PART III

APPENDICES

EXPLANATION OF TERMS

Data class : Data source-specific class or cluster, defined from relationships in a particular data space. Also named 'attribute values class' when the object attribute refers to a particular data source.

Faces (photo interpretation) : surfaces on the aerial photo, homogeneous in tone and texture, recognisable through change in tone or texture or through linear features delimiting the faces.

Farm system : Decision making unit comprising the farm household, cropping and livestock systems, that transforms land, capital and labour into useful products that can be consumed or sold (Fresco *et al.*, 1990).

Farming system : Class of similarly structured farm systems (Fresco *et al.*, 1990).

Farming Systems Analysis : Diagnosis and analyses of farm level variables, concerning farming and land use, covering both ecological and socio-economic aspects (Fresco *et al.*, 1990).

Geographical Information System : A (computerized) system providing a structure for the description of spatial data and tools for storing, retrieving, transforming and displaying of geographical data, describing objects from the real world (Burrough, 1986).

Information categories : classes of information bearing a direct relevance in the user situation. Inferences about the information classes are drawn from the collection of data classes.

Land Cover : Vegetation and artificial constructions covering the land (Rhind & Hudson, 1980).

Land Evaluation : The process of assessment of land performance when used for specific purposes, involving the execution and interpretation of surveys and studies of landforms, soils, vegetation and climate and other aspects of land in order to identify and make comparison of promising kinds of land use in terms applicable to the objectives of the evaluation (FAO, 1976).

Land Use : Man's activities on the land, which are directly related to the land (Clawson & Stewart, 1965).

Land Use Pattern (LUP): The LUP describes the land use of a land use zone. The LUP denotes the land utilization type(s) and farming system(s) occurring within a LUZ.

Land Use Zone (LUZ): A geographical unit (or object) with a particular land use pattern and a dynamic behaviour, expressed by change in the land use characteristics and by change in the land use zone boundaries. The LUZ provides a geographical basis for the description of land use and land use change at sub-regional level.

Land Utilization Type (LUT): Crop, crop combination or cropping system with a specified technical and socio-economic setting (FAO, 1983). In this thesis the LUT refers to land use at field level.

Land Evaluation and Farming Systems Analyses (LEFSA): A conceptual model for land use planning, integrating the land evaluation and farming systems analyses approaches.

Object class : Kind or type of object.

Spatial pattern : An arrangement of (spatial) components which has more meaning than a simple aggregation of the components. Within the context of the present study the components refer to agricultural fields.

Superclass : Generalization of object classes; many object classes can belong to one superclass.

LIST OF ABBREVIATIONS

AP	-	Agricultural Penetration
APs	-	Aerial Photos
APTD	-	Average Pairwise Transformed Divergence
AZ	-	Atlantic Zone
BB	-	Bare soil and Built-up area (land cover class)
CATIE	-	Centro Agronómico Tropical de Investigación y Enseñanza
CLC	-	Composite Land Cover
CLCC	-	Composite Land Cover Class
FDS	-	Formal Data Structure
FOR	-	Forest (land cover class)
FS	-	Field Size/ Farm Size
FSA	-	Farming Systems Analysis
GIS	-	Geographical Information System
GRS	-	Guacimo-Rio Jiménez-Siquirres
IDA	-	Instituto de Desarollo Agrario
LC	-	Land Capability Class
LCC	-	Land Cover Class
LCCC	-	Land Cover Class Composition
LE	-	Land Evaluation
LEFSA	-	Land Evaluation and Farming Systems Analyses
LU	-	Land Use
LUP	-	Lans Use Pattern
LUZ	-	Land Use Zone
LUT	-	Land Utilization Type
MAG	-	Ministerio de Agricultura y Ganaderia
PAS	-	Pasture (land cover class)
PMU	-	Physiographic Mapping Unit
SI	-	Satellite Imagery
TM	-	Thematic Mapper
USGS	-	United States Geological Survey
WAU	-	Wageningen Agricultural University
WA	-	Wooded Area (land cover class)

GENERAL INTRODUCTION

LAND EVALUATION AND FARMING SYSTEM ANALYSIS.

In 1986 the Wageningen Agricultural University (WAU) entered into a cooperative research effort with the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) and the Minestério de Agricultura y Ganaderia (MAG). These institutions started a program for regional development studies, natural resource inventory and studies of ecological change in the Atlantic Zone of Costa Rica. The ultimate aim of the research is to provide relevant insights for agricultural planning purposes, directed to sustained land use and small farm development. The project aims at the development of a methodology for defining alternative land use scenarios, based on the Land Evaluation and Farming Systems Analysis (LEFSA) concept (Fresco *et al.*, 1990).

Within the framework of the Atlantic Zone research program attention is devoted to the application of satellite remote sensing and geographical information systems (GIS) for general purpose land cover and land use inventory and the monitoring of changes in this respect. It serves to provide information needed for the LEFSA sequence, although the LEFSA sequence was not yet defined in 1987, when the present study was initiated.

LEFSA is a conceptual model for land use planning. Land use planning refers to regional agricultural planning. It is a form of intermediate-level planning of regions with a view to bridging the gap between general macro-planning and specific project planning. It concerns the identification of priority areas, which can be understood in both a geographical and a thematic sense. Both land evaluation and farming systems analysis represent tools for land use planning. The LEFSA integrates both methods, in order to profit from the advantages of both approaches and to overcome the shortcomings intrinsic to each one. The initiative to reconcile procedures had previsouly been taken by Young (1986).

Land evaluation has a strong geographical orientation, through the mapping and description of the different types of land. It looks at potentials for the use of land, based on an evaluation of the bio-physical resources, with respect to selected types of land utilization (FAO, 1976). The shortcomings of many land evaluations are related to problems in integrating agronomic and socio-economic information. A constraint of the methodology itself is the lack of clear procedures for the selection of land use types (Fresco *et al.*, 1990; also remarked by Young, 1986).

Farming systems analysis focuses on determining present uses of land. In contrast, land evaluation puts the emphasis on future and potential uses. Farming system analysis diagnoses the present situation with regard to farming and land use, providing insight in possible and necessary improvements. Farming system analysis lacks the geographical orientation of the land evaluation approach. Its problems are largely related to the selection of the relevant areas. The land units serve to identify the areas where agricultural problems are subsequently diagnosed. These units, however, might be less suitable to denote areas with homogeneous land use characteristics. In such areas the farmers are part of similar agricultural systems, operating under comparable agro-ecological as well as socio-economic conditions. The integration of LE and FSA, as proposed in LEFSA, intends to preserve the strong geographic character of the LE and to use the FSA for diagnosis and identification of relevant land utilization types.

LEFSA recognizes the different emphasis on the hierarchical level at which both approaches work. LE focuses on the regional level in its reconnaissance task and on cropping system level in its detailed analysis, whereas FSA concentrates on the farm level. Therefore, hierarchical levels are identified that are acceptable to both methodologies. Starting from the highest level, these are: the regional level, the sub-regional level, the farm system and the farm sub-system (which consists of the household system, the cropping system and the livestock system).

These hierarchy levels are intended to be relevant for both the land evaluation and the farming system analysis. By implication, both the land use and the physiography or the soil can be described as separate systems or models at the different hierarchy levels. In the present investigation land use and soil have been treated as independent themes.

Both the crop system and farm system can be geographically and thematically defined as individual entities. This is less clear for the description of land use at sub-regional and higher levels. In the document mention is made of the possibility to employ land use and vegetation maps to which farm information from statistical sources or limited field work can be related. Such maps may be based on information from remote sensing, whereby "care should be taken that the different land use types and cropping and livestock systems can be identified within the land use/cover units" (Fresco *et al.* 1990). It is commented that "this might require changes in the way the land cover and land use classes are presently defined."

The latter quotation specifies the focus of the present study. The aim is to contribute to the LEFSA sequence by elaborating the sub-regional level with respect to the land use. To provide tools for the inventory of land use at the sub-regional scale, satellite remote sensing data and aerial photography are employed. The traditional tools for FSA comprise observations in the field and farm survey.

SATELLITE IMAGERY AND AERIAL PHOTOS FOR SUB-REGIONAL LAND USE INVENTORY

Since the launching of the first satellite for land observation in 1972 (ERTS1), remotely sensed data has been used extensively for the inventory of land cover and land use. However, there are some limitations to the use of satellite remotely sensed data. These were also experienced with respect to the land cover and land use inventory in the Atlantic Zone of Costa Rica. In brief:

- Remote sensing is best suited for the inventory of land cover, since this can be directly sensed. The relevance for land use inventory is limited. Also in a general purpose land cover classification, the classes are often hardly specific, further complicating the inference of land use information.
- The results of the satellite-based land cover classification often show limited reliability.
- Image interpretation is a rather unstructured, subjective activity. As a consequence the results have a somewhat 'ad hoc' character. The process is therefore difficult to describe and transfer and the results are uncertain. Also the procedure may be less efficient; much time is needed to obtain results of acceptable quality. The general-purpose classification of land cover requires insight in the occurrence of land cover types over a large area. It also requires insight in the possibilities for the recognition of these cover types through remote sensing. With the increasing need for timely data and given the changeable character of land cover and land use, this is considered to be a major limitation.

The present document addresses these questions, as well as the problems associated with the use of remote sensing (and GIS) for purposes being discussed.

For regional agricultural planning, information on the regional land use pattern is required. The above-mentioned problems indicate a discrepancy between the methods and techniques available and the information required. This is related to a discrepancy in scale level. A different general strategy for the inventory of land use is clearly needed. The methods (and techniques) should be adapted to the information requirements. Remote sensing serves as a possible data source. Integration with other data sources is needed to obtain the required land use information.

An important aspect of the approach to land use inventory presented here is its object-oriented character. The object refers to a geographic entity, associated with a specific level of aggregation. As such, scale dependencies can be accounted for. Also in land evaluation, we see that in recent years attention has focused, in recent years, on tailoring soil studies (or the use of soil survey data) to specific requirements. These study designs take into account the scale dependencies, as well as levels of detail and accuracy of the information required (Bouma *et al.*, 1986; Bouma, 1989).

GEOGRAPHICAL INFORMATION SYSTEMS AND THEORY

Land use is changeable, especially in regions like the Atlantic Zone of Costa Rica, where the colonization history is relatively young.
The capability for regular updating and monitoring of land use change is therefore an important requisite. With respect to regional agricultural planning it is important to visualize the effects of different land use scenarios and to be able to assess the effects of measures taken. It implies the use of a geographical information system (GIS).

The monitoring of land use change requires a systematic approach to land use inventory. Structured, objective and quantitative methods and measures are needed to enable the consistent execution of the inventory process and description of land use corresponding to different points in time or to different regions. It entails the design of a data model for description of land use and a clearly defined process structure for the inventory of land use. The system aspects are an important theme in the present research.

A geographical information system (GIS) provides a certain structure for the description of spatial data and tools for storing, retrieving, transforming and displaying of geographical data (Burrough, 1986). The geographical data describe objects from the real world.
The term GIS can be taken to denote a computer system with the technical capabilities and functions just cited. Alternatively, the term may denote an application-oriented GIS, providing a conceptual data model and prescribed (semi-automated) procedures for data processing relevant to the field of application, The term may also refer to a GIS in which a fully elaborated data model has been implemented, containing data on a specific area and for a specific application and with standard (or fully automated) procedures. In this thesis we are concerned with the development (or application) of a GIS for a specific application, namely land use inventory.

Geographical information systems are widely used at present. Problems are encountered with the exchange of data and with the linking of geo-information of different fields of application. (These problems were also experienced in the present investigation, specifically in linking the soil and physiographic data to the land cover and land use data.) A major problem is that the object definitions are always embedded in a particular context (i.e. field of application, scale, geographical region). Solutions have been sought in data standardization (i.e standard definitions of terrain objects) but have, as yet, not been very successful. This is observed and commented upon in an article by Molenaar (1991c) entitled "Object hierarchies, why is data standardization so difficult?". An explanation of the difficulties in data standardization would require an investigation of the structural aspects of data and object definition. That is, it presupposes the definition of theoretical information models for geographical information systems. The formal data structure (FDS) developed by Molenaar (1991a, 1991b) has been adopted as a principle for the structuring of the object description in the present study.

The data model only constitutes part of an information system, when this is taken to represent an arrangement of components (or subsystems) that *process* inputs into outputs. The information contained in a map (or GIS) cannot be seen independently from the process (and methods) to derive the information from the input data.

Also with respect to data processing, standardization of the procedure (defining a uniform process structure) is important to guarantee comparable and reproducible results. The definition of such a structured procedure (or process structure) requires insight in the data and information flow. In the information analyses the functional steps and the relation between these steps in the information process are identified. On the other hand the data structure and the methods and techniques for data processing are investigated. Decision rules determine the process logic (Essink and Romkema, 1989). In the context of the land use inventory, the information process refers to the inference of land use information from satellite imagery and aerial photos, applying knowledge on the regional context (see Figure 1 for the general structure).

Image interpretation (referring to both aerial photo and satellite imagery) is a rather complex process. Image interpretation involves seeing and understanding. This requires both the identification of image pattern elements (i.e. tone, color, size, shape, etc.) and the analysis and articulation of conceptual knowledge, based on diverse stereotype models and heuristic rules employed by expert interpreters (Agrialas and Harlow, 1990).

As is often the case in aerial photo or satellite image interpretation, the objects of interest are not defined a-priori, such that the interpretation can be directed to the recognition hereof. Indeed, the interpretation serves to identify the objects. In the case of the land use inventory the object as concept for modelling of land use even had to be introduced. In that event, the problem is to determine the relevant object characteristics, which requires insight in the role of the objects in the land use analysis (i.e. the context). When it is known which information the objects have to carry, the relevant object characteristics can be determined. From these the criteria for the object definition and recognition follow. Using imagery for the recognition of objects preseumes the translation of the object characteristics into image characteristics. The pattern recognition contains context-dependent decision moments. Lack of insight in the problematic sketched here makes image interpretation a very subjective activity.

With respect to remote sensing, systems for interpretation of satellite imagery have received much attention in the literature. The systems are generally based on a per pixel classification. In order to improve results the models not only contain algorithms for solving mathematical problems, but they also incorporate knowledge about the objects to be mapped. Making use of such knowledge represents the expert system approach. For an overview of the knowledge-based (or model-based) models, see Agrialas and Harlow (1990) with respect to remote sensing image interpretation, and Robinson and Frank (1987) as regards knowledge-based systems for GIS. A general conceptual model that would apply to a GIS for land use inventory is presented in Figure 1.

Two main tasks were set for the present investigation: first, the definition of (a structure for) the land use inventory process; and second, the application of the land use inventory process at sub-regional scale. A representative area of about 30 by 30 km was selected for the application and validation of the procedure. The area is named after the three villages that define the area: Guacimo, Rio Jiménez and Siquirres (GRS area). The question is whether land use can be quantified through a repeatable and transferable methodology.

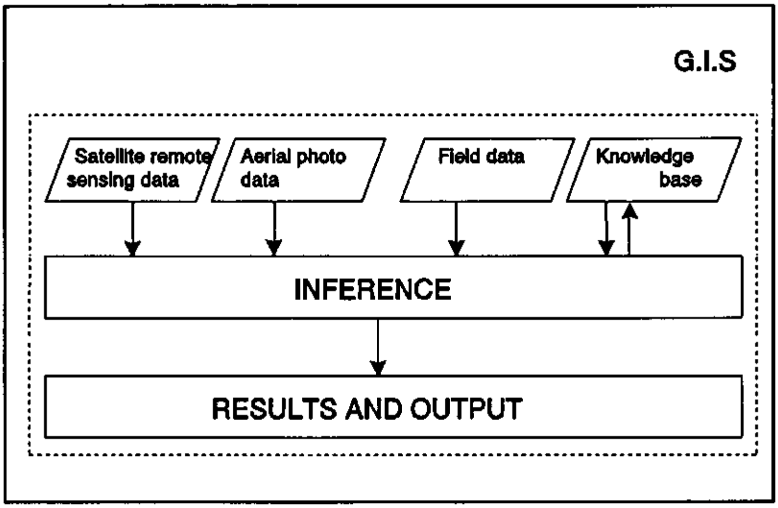


Figure 1 *General structure of a knowledge-based system.*

LAND USE INVENTORY AND ANALYSIS IN THE ATLANTIC ZONE OF COSTA RICA

For the regional inventory of land use in the Atlantic Zone of Costa Rica, the land use zone is defined as the central geographical unit (or spatial object). The land use zone concept assumes the existence of sub-regions with discriminating land use characteristics. The land use is described by the land use pattern, denoting the farming and cropping systems occurring within a land use zone and the prevailing socio-economic conditions. The units are intended to serve as target areas, composed of farmers operating under similar conditions and experiencing more or less the same problems and opportunities. That is, it serves to diagnose and analyze the agricultural system at sub-regional level and it identifies the relevant land use types for the land evaluation to be executed per unit identified. As such the units are assumed to provide a context for defining and evaluating land use scenarios.

The land use zone can be considered an aggregation of agricultural fields (as far as the agricultural lands are concerned). The land use zone may consist of many farms, a single farm, or a part of a farm when the corresponding area is large enough to be recognized at the sub-regional level and when the agricultural activities of the farm or enterprise are differentiated.

In Part 1 the land use zones are described and their relevance for the inventory of land use and land use change on a regional scale is assessed. The discussion It concerns the pragmatics of the identified objects. The first chapter gives a structure for the description of spatial objects. In the subsequent chapters, the thematic and behavioral aspect of the objects are investigated in relation to land use and land use change. All concern the GRS area. An important theme in this first part is the interpretation of the object's image characteristics with respect to the land cover and land use characteristics.

In Part 2, the strategic and methodological aspects are addressed. The strategy consists of the following parts:

- The stratification (or image segmentation) process, in which the land use zone boundaries are determined. The geographical units are determined by means of aerial photo interpretation, whereby the spatial pattern is the most important criterion. A pattern (or structure) is attributed by an expert to an arrangement of components, which has more meaning than a simple aggregation of the components (Agrialas and Harlow, 1990). Concerning land use inventory, two related types of patterns are distinguished, namely the spatial pattern (e.g. certain arrangement of fields) and the land use pattern (e.g. certain combination of land use types), whereby the former is an expression of the latter. Aside from being a mere aggregation of elementary objects, the land use zone is justified as a geographical entity in itself, by its specific land use history and land ownership pattern.
- The determination of the land use zone characteristics. The land use zones are characterized by elements derived from aerial photos and land cover composition. To obtain information on land cover, use is made of satellite imagery. Analyses of the object characteristics indicate the degree to which the land use zones can be mutually discriminated. On this basis, the aerial photo interpretation is evaluated and accepted or rejected.
- The classification of the land use zones. This consists of classification of the land use zones on the basis of their image characteristics. First the appropriate classes are defined, taking account of the discriminating power of the object characteristics. In second instance, these classes are to be translated to the relevant categories describing the land use patterns of the land use zones.

The structuring of the inventory and classification process requires a clear distinction between the object's image characteristics, which are important for object recognition and characterization, and the object's land use characteristics, which are inferred from the image characteristics. The inference is described as a separate activity.

To make an inventory of land use, a data-driven approach was selected. This means that the classes are not defined a-priori but are generated by the procedure. This approach is the strategic answer to situations in which exact information requirements cannot be specified and in which there is limited availability of materials. This approach is also an answer to the image interpretation dilemma described earlier.

In an introductory chapter to Part 2 the land use zone approach is described in general terms. It is justified as an object-oriented approach in terms of developments in the use of satellite imagery and aerial photos for land cover and land use inventory.

Part 3 deals with land use evaluation, taking bio-physical land characteristics into consideration. When the present research was initiated, soil mapping of the Atlantic Zone was already under way. A case study is presented which makes use of this soil data and of the land use zone data for evaluating land use at the sub-regional scale. The evaluation serves to identify priority areas with respect to possible over-exploitation of the natural resources and with respect to possibilities for improved use of the natural resource.

In addition, a study is presented concerning the variation in banana yield as explained by remote sensing and soil survey data. This study is included to demonstrate the use of remote sensing and soil information at a more detailed level.

ATLANTIC ZONE PROFILE

This condensed profile of the Atlantic Zone was compiled from Hall (1984), Silva (1982), Waaijenberg (1990) and Wielemaker (1990). The information on land cover was obtained from our own data sources.

Geomorphology and soils

The northern Atlantic Zone (AZ) is located in the north-east of Costa Rica. It constitutes the transition of the central volcanic mountain range to the Caribbean Sea.

Three major landforms can be distinguished from the south-west to the north-east:

- The sloping areas of the central mountain range with lava and lahar deposits of andesitic composition. At higher elevations these deposits are covered with volcanic ashes.
- Slightly inclined plains at the foot of the volcanoes with fine grained fluvio-laharic deposits.
- The coastal plain with fine-textured to peaty deposits and inundated depressions. Remnants of Tertiary and early Quaternary volcanism with deposits of basaltic composition, strongly dissected and with deep weathered soils are also found.

The soils in the area belong predominantly to the Andosols and Inceptisols. Their fertility status depends largely on the age of the deposits, which varies strongly for the first two of these three landscapes. The younger deposits are covered with nutrient-rich and non-acid soils, whereas the older deposits are covered with nutrient-poor and acid soils. The younger lava and lahar deposits on and near the volcanoes can be extremely stony. The soils of the coastal plain are fertile but often suffer from impeded drainage. Very recent fluvial deposits are shallow and sandy, limiting agricultural use.

Climate

The AZ has a humid tropical climate. The mean annual temperature is 26°. Daily temperatures vary little, and differences in day and night temperature are small (a maximum difference of about 12°).

The mean yearly precipitation varies between 3000 and 6000 mm/yr, with mean monthly precipitation generally between 300 and 700 mm. The months of February, March and April represent a relatively dry period with mean monthly precipitation values between 100 and 300 mm.

Agricultural history and present use

The colonization of the AZ began with the establishment of some cacao plantations around Matina in the 17th century. Before then, the AZ was populated by a few indigenous groups. Plantation worker were recruited from these groups. Later negro slaves were imported to work on the plantations.

The late colonization of the AZ might be explained by the adverse climatic conditions, the presence of dense forest cover and other less favorable conditions in the Atlantic Zone. The colonization proceeded very slowly in the beginning; by 1820, Matina had only 34 inhabitants.

The colonization process gained momentum with the construction of a harbor at Limon and the construction of the railway (1872-1890) for the benefit of the coffee export. The coffee was produced around Cartago. First the line Limon-Guapiles was built; later, the branch from Siquirres to Cartago was constructed. In this period, a start was made with the cul-tivation of banana. The product was already exported before the end of the century. The period is marked by the foundation of the United Fruit Company, which was to dominate banana production in the area for four decades to come.

The beginning of the 20th century showed a strong increase in the production of banana, accompanied by further extension of the railway infrastructure. Production came to a standstill in the 1930s as consequence of the Panama disease, which attacked the roots of the plants, and of labor unrest. Cacao, maize and cassava became important crops after the demise of the banana plantations. The initiative for the cultivation of cacao seems to have come from the banana companies. The area of its cultivation is often associated with former banana plantations. At the end of the 1950s, we see the return of banana cultivation, with new varieties resistant to the Panama disease. Production increased steadily thereafter, accompanied by improvements of the road infrastructure. This period is also marked by the immigration of many people from other parts of the country, seeking jobs at the plantations or looking for other opportunities to make a living. Those with financial resources bought farms; others occupied new lands still under forest at the fringe of the reclaimed areas.

The workers at the plantations were often employed on a temporary basis. When leaving the plantations many tried to settle nearby as farmers. In the 1970s, the pressure on the land increased and larger estates were invaded. These estates were generally not used for agricultural purposes. Conflicts were settled by the intervention of the Institute for Agrarian Development (IDA), buying up the estates and distributing the land.

At the end of the 1980s a new agrarian policy was announced, seeking to reduce subsidies on basic grains and increase the cultivation of crops for export. The last years have shown an increase in the cultivation of ornamental crops and flowers and food crops like macadamia. Most of these crops, however, are not feasible alternatives for the small farmer. Land use in the region is thus characterized by strong and rapid change, which find expression above all in the rapid deforestation of the area. The location of land brought into use, land claims and grants, and the settlements were all strongly related to the route of the railway. It has resulted in a pattern of parcels and fields positioned in a direction perpendicular to the railway. This pattern can still clearly be recognized.

Table 1 presents data on land cover of the AZ. This data was obtained on the basis of a 1986 satellite image (Landsat-TM). The largest part of the Atlantic Zone is still occupied by forest. The forests are found in the coastal area and on the slopes of the volcanoes of the central mountain range. The percentage presented in Table 1 somewhat overestimates the forest cover, because the Landsat-TM scene includes a forested part of Nicaraguan territory.

Of the unforested area, most is covered by 'wooded area'. This includes tree crops like cacao, citrus and macadamia (the latter two are not very important), homesteads, wooded pastures and other wooded areas like riverbanks. Pastures occupy a large part of the area, especially if we consider that part of the 'wooded area' also corresponds to grassland and that the 'grass vegetation on inundated or poorly drained lands' often forms part of the pasture land. Banana production is by far the most important economic activity in the region at present, though it occupied (only) 9.6 % of the unforested area in 1986. Livestock breeding takes a second place.

Table 1. *Land cover distribution in the Atlantic Zone*

The Atlantic Zone land cover distribution: Forested and deforested area:	
Forest	46.4 %
Unforested	53.6 %
Total area of the scene	569560 ha.
Land cover distribution within the deforested area:	
Banana	9.6 %
Pasture land	24.7 %
Unvegetated lands	13.6 %
Wooded area	27.3 %
Secondary Vegetation	12.9 %
Grass vegetation of lands inundated or poorly drained	7.0 %
Other	4.9 %

Land cover and land use was inventoried for the northern Atlantic Zone. The development of the methodology and testing of the validity of the approach, of which the next chapters give account, was done in a part of the Atlantic Zone considered representative for the area in total. The area has been called the Guacimo-Rio Jiménez-Siquirres (GRS) area, after the three most important villages in the area. The location of the GRS area is shown in Figure 1.

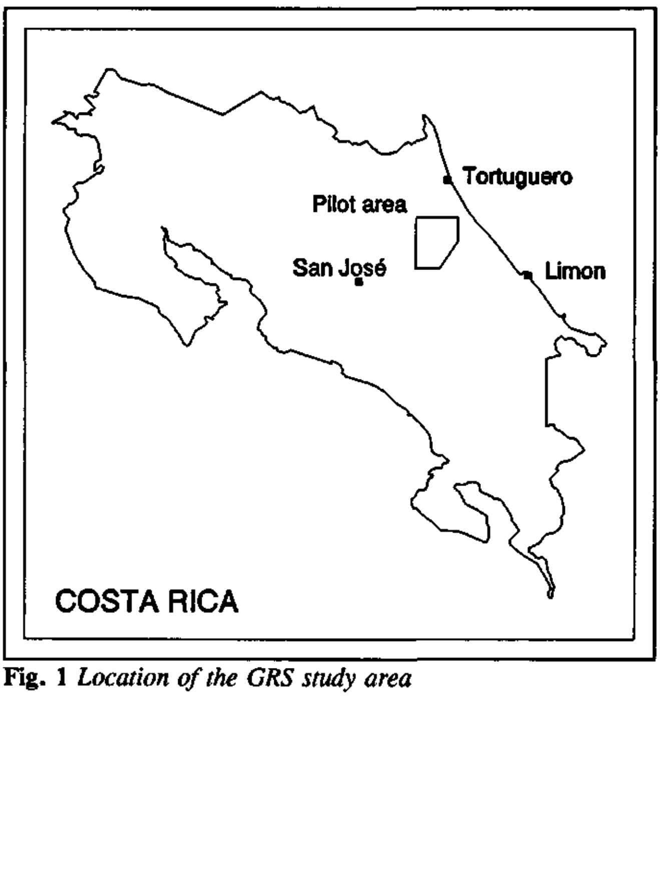


Fig. 1 *Location of the GRS study area*

PART I

LAND USE ZONES TO DESCRIBE LAND USE AND LAND USE CHANGE AT THE SUB-REGIONAL LEVEL

CHAPTER 1

A DATA STRUCTURE FOR DESCRIBING TERRAIN OBJECTS

INTRODUCTION

The main purpose of Part 1 of this thesis is to evaluate land-use zones as meaningful entities for inventorying and mapping land use and land-use change on a sub-regional scale. An inventory has been made of land-use in the Guacimo-Rio Jimenez-Siquirres area of Costa Rica. The results are presented as a land-use map in Appendix D.

A land-use map is a model of the real world. This model is relevant only in the specific context for which it has been designed. Of more general interest is the structure of the model, which we expect to be applicable to land-use modeling in other regions as well.

A model is described in terms of objects (or entities) and in terms of the relations between objects. In a geographic context, these objects represent terrain objects (or spatial objects). Geographic Information Systems (GIS) are used to store and analyze the data on terrain objects (Burrough, 1986).

Attempts to link different GIS have been disappointing and efforts to tackle this problem through standardization of object description have failed. Molenaar (1991c) attributes these failures to the context-dependent nature of object definitions. Further attempts to solve these integration problems have led to the investigation of spatial data structures. The importance of a data structure for describing terrain objects has been stressed by various authors (Peuquet, 1984; Molenaar 1991a,b).

This first chapter presents a data structure to describe terrain objects. It provides a framework for modeling terrain objects and it explains the terms and concepts that the reader will encounter in the subsequent chapters. The data structure presented here is based on the Formal Data Structure (FDS) defined by Molenaar (1989a, 1991b) and Molenaar and Janssen (1991). Only minor changes have been made.

A FORMAL STRUCTURE FOR DESCRIBING TERRAIN OBJECTS

Included in the descriptions of terrain objects are the following characteristics:

- Geometric structure;
- Thematic content;
- Dynamic behavior.

Also included in the descriptions are the relations between the objects. Besides the topological relation, three other types of relation can be defined. These are:

- Classification hierarchies;
- Aggregation hierarchies;
- Associations between the objects.

The dynamic behavior of the objects refers to the changes in these characteristics over time. Such changes require an updating of the objects' representations in a GIS.

Object-Oriented Terrain Description

There are three components in the basic structure for describing the terrain objects in a GIS (Fig. 1.1). These are:

- An object identifier;
- Geometric data;
- Thematic data.

Behavior can be added as a fourth component, but we shall not be commenting upon this further, as it is expressed in terms of the object's geometric or thematic characteristics.

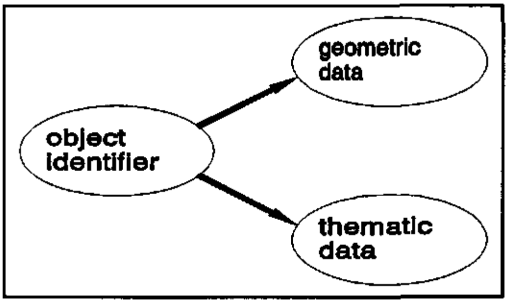


Fig. 1.1 *Basic structure for the representation of objects (after Molenaar, 1991a).*

Depending on the context, an object can be described as an area object, as a line object, or as a point object. (A town, for example, can be represented either as an area object or as a point object, depending, among other things, on the scale of the map.) Objects are represented in one of two geometric structures. The current systems of processing geographic information are either vector-based or raster-based. They imply a vector-to-raster conversion, or vice versa, when the objects that are presented in different geometric structures are to be combined or related.

The description of geometric structures is outside the scope of this disssertation. For a description of a formal data structure for vector maps, see Molenaar (1989a and 1991a).

The concept of the Land-Use Zone (LUZ) was developed to enable an inventory of land use in the Atlantic Zone (AZ) of Costa Rica (CR). The LUZ defines the central geographic object (or unit) for describing that land use. The LUZ is characterized by a certain land-use pattern and its boundaries (or geometric data) are determined through aerial-photo interpretation, a process in which the spatial pattern is the most important interpretation key. A spatial pattern is an arrangement of superficies with a homogenous land cover (e.g. fields).

Object Classes

Thematic data refer to the attributes of an object. Objects are related if they belong to the same object class, which is defined as either a common attribute structure or corresponding attribute values. All objects in an object class have a common attribute structure. In Figure 1.2, we see that one object class can contain many objects.

Take, for example, the object class "rivers". Relevant attributes of rivers include depth, width, current velocity, flow rate, and maximum tonnage. All the objects classified as rivers have these attributes. Only the values differ. The possible attribute values come from the attribute domain (Fig. 1.3). Similarly, we could define an object class "roads", also with its own attribute structure.

In principle, there is no difference between an object class that is defined through a common attribute structure and one that is defined through corresponding attribute values.

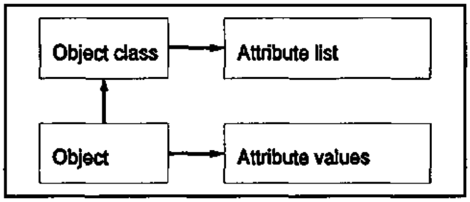


Fig. 1.2 *Class structure of objects (Molenaar, 1991a)*

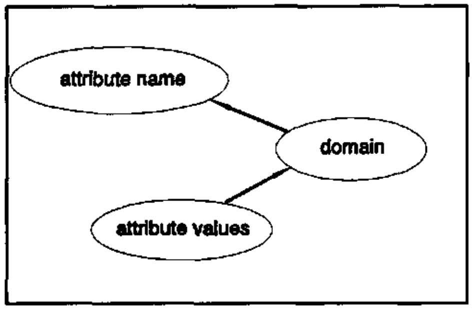


Fig. 1.3 *The relation attribute value-atribute domain-attribute name (Molenaar, 1991a).*

For example, we could classify rivers as either small or large, on the basis of their flow rates. At this level of generalization, the value domain would specifies the two possible class values. The former object class "rivers" becomes a superclass representing a higher level of generalization than the classes "small rivers" and "large rivers". This shows that we can build classification hierarchies by evaluating the attribute values at different generalization levels.

Object classes with partially corresponding attribute structures can be grouped into a superclass. Each level of the classification hierarchy has its own attribute structure and its own set of attributes whose values are evaluated. The object classes inherit the attribute structure and the attribute values of their superclass. We can, for example, define a superclass "waterways" that contains "rivers" and "canals" at the next level down. Canals include all water ways with no flow rate. Rivers, however, do have a flow rate, an attribute that allows us to divide them into either "large rivers" or "small rivers". At this level, we can determine the object attributes (e.g. large rivers are characterized e.g. by their maximum allowed tonnage, an attribute that might not be relevant for small rivers).

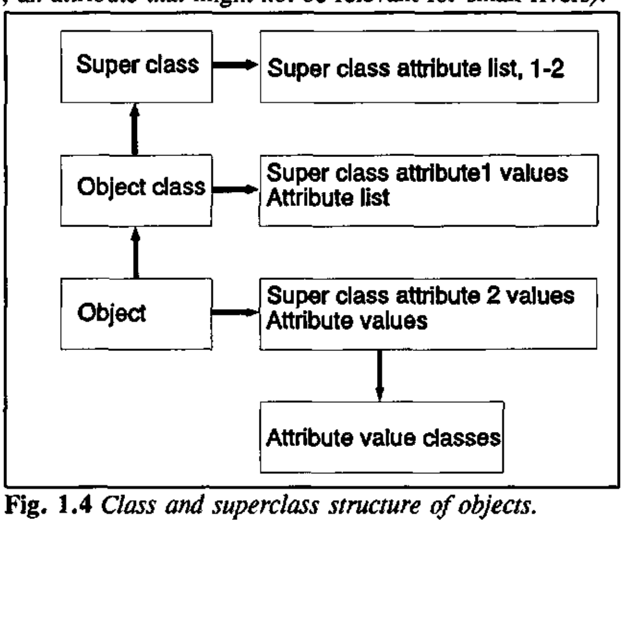


Fig. 1.4 *Class and superclass structure of objects.*

Figure 1.4 also shows the class hierarchical structure. At the highest level are the superclasses with their lists of superclass attributes. Some of these attributes are evaluated at the next level down, the class level, and some are evaluated at the lowest level, the object level. At each level, additional attributes are defined. It is possible to find out whether an object belongs in an object class (or in a superclass) by verifying the presence or absence of attributes and by evaluating attribute values. The classes are defined either by the user or through the analysis of the distribution characteristics of the attribute values. The classes represent clusters within the object space that is defined by the attributes. These clusters are called attribute value classes.

LUZs are used for the inventorying and mapping of land use at sub-regional level. They represent zones with a recognizable land-use pattern. Land-use patterns reveal a farming system or a particular combination of farming systems (e.g. cattle farming or plantation cropping) in a specific sub-region. They represent the object class level. In the classification hierarchy, classes like "agricultural areas" and "residential and industrial areas" are given at superclass level. The specific classification structure depends on the purpose of the land-use inventory.

Aggregation Hierarchies

In aggregation hierarchies, the inheritance lines lead upward. This is unlike classification hierarchies, in which the inheritance lines of attribute structures lead downward, causing the thematic description of terrain objects to become more detailed the farther we go down the hierarchy.

Figure 1.5 shows how composite objects can be built up from elementary objects and how these composite objects can be put together to build up still more complex objects. The composite objects inherit the attribute values from their constituent parts. There is no direct relation to the classification hierarchy. In fact, for every type of composite object, a separate classification structure can be defined (Fig. 1.6).

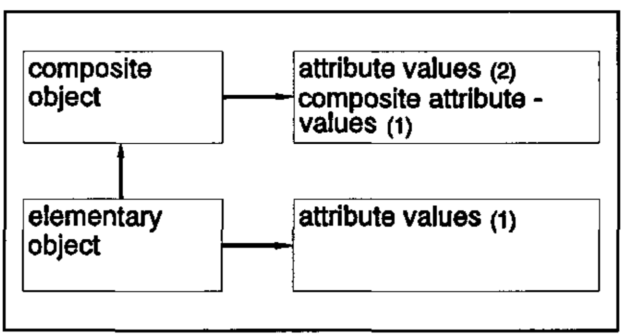


Fig. 1.5 *Structure for composite objects.*

For each level of the aggregation hierarchy, there should be rules for selecting the terrain objects that are to be aggregated to a particular composite object. In a GIS, these rules will be based partly on the topological relations between terrain objects (e.g. connectivity or adjacency) and partly on the thematic relations (e.g. common class values).

Not all of the class values that can occur in a composite object will occur. Even so, contrary to what Molenaar states (1991c), this does not mean that composite objects cannot have a fixed attribute structure. If elementary objects can be aggregated to form composite objects, then their attributes can be aggregated also. For each elementary attribute, there is a corresponding composite attribute, implying that for the fixed attribute structure for the elementary object there is a corresponding fixed attribute structure for the composite objects.

The distribution of the attribute values (or the attribute value classes) within the composite objects can be described with a frequency distribution or with statistical data like mean and standard deviation. The composite attribute values will depend on the specific composition of the composite objects.

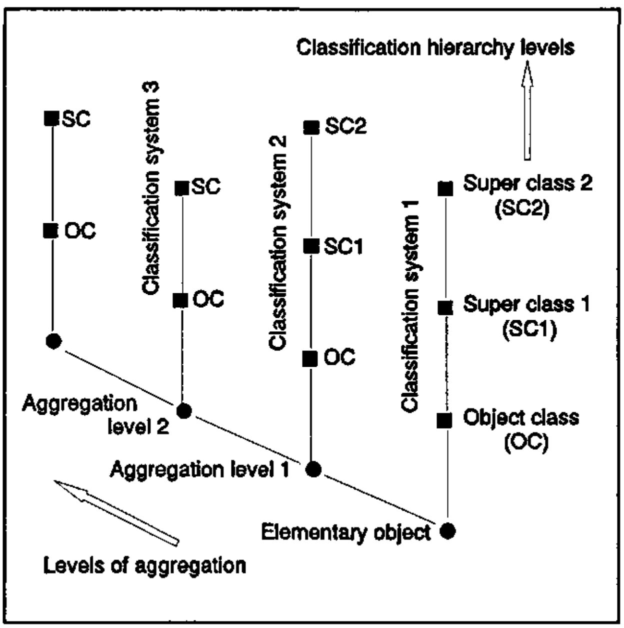


Fig. 1.6 *Relation between aggregation hierarchy and classification hierarchy*

A good example of a composite object is a town district, which can be considered an aggregation of buildings. It can be classified as residential, administrative, or industrial, depending on the function of the buildings, that is, depending on their object class (e.g. houses, offices, shops, factories). But having said this, we should bear in mind that a residential area need not consist exclusively of houses - shops and some offices might be present as well. Nor, by the same token, does an industrial area need to have only factories. The district can be described simply by for example the number of buildings in each object class.

LUZs are composite objects. The elementary objects that combine to form them are fields (or, in more general terms, superficies with a homogeneous land cover). Fields are described through their size and land cover. The composition of the LUZs is then described through mean field size and land-cover composition. (The percentages of land cover are obtained by overlaying the LUZ with a land-cover classification based on satellite imagery.)

As the LUZs delineation is based on recognizable field patterns, so it follows that the shape, size, and orientation of the fields are the actual criteria for aggregation. The spatial pattern itself is an independent (non-aggregated) characteristic of the composite objects, relevant only to the specific aggregation level.

Object Associations

Besides classification hierarchies and aggregation hierarchies, which are very well defined, there is a third, less well defined type of relation between terrain objects: Object associations. The associations between terrain objects are determined by corresponding attribute values (e.g. houses or buildings with the same owner, companies with an office in Amsterdam, and so on). And, unlike the hierarchies, which are characterized by many-to-one relationships (m:1), object associations represent many-to-many relationships (m:n).

Even though they are less strictly defined, associations between terrain objects can be very important for interpreting and evaluating certain phenomena. Object associations are generally found by mounting searches based on attribute values or by analyzing the joint occurrences of different kinds of objects.

Object associations based on ownership apply to LUZs. LUZs with a common owner are sometimes scattered over the region. Often they are adjacent, in which case they represent one farm or one enterprise with various kinds of land use (for which they are mapped as separate LUZs). Information on ownership can be relevant because a change of ownership often implies a change in land use.

OBJECT DYNAMICS

Now that we have a data structure for describing terrain objects, we can describe the dynamic behavior of these objects as well. By dynamic behavior we usually mean:

- Changes in geometric structure;
- Changes in thematic content;
- Changes in aggregation structure.

A change in the geometric structure of an object can mean a change in its position, its shape, or its size. A change in position implies a corresponding change in the topology,

which will then have to be updated. For a discussion of updating geometric data and maintaining consistency, see Molenaar (1991b).
Changes in the thematic content of an object concern:
- Changes in one or more attribute values;
- Changes in the attribute structure (e.g. an agricultural area that is converted to a residential area).

Both kinds of change might cause the object to migrate to another object class.
Changes in the aggregation structure are illustrated in Figure 1.7. They can be:
- The fragmentation of an area into smaller objects;
- The dissolution of small elementary objects into a larger object;
- The changes in the arrangement of the elementary objects that make up an aggregated object.

Changes in the definition of the aggregation hierarchies and in the classification hierarchies are not considered here because they do not concern the dynamic behavior of the objects. They do, however, affect the relations between the objects.

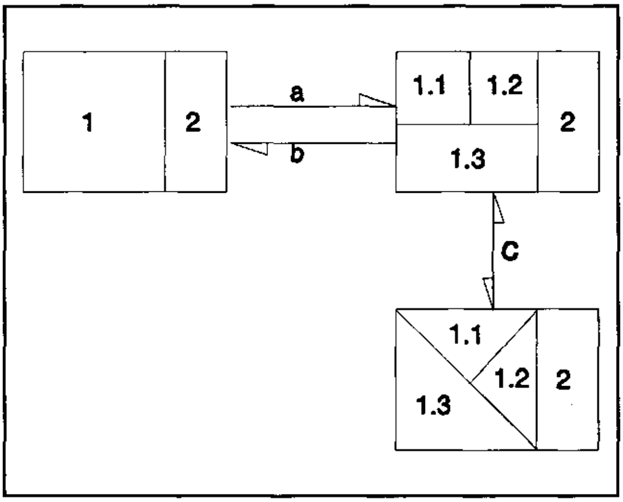


Fig. 1.7 *Changes in aggregation structure. a - fragmentation of objects; b - dissolution of the elementary objects; c - change of arrangement.*

To these three types of change we can add a fourth, one that does not refer only to objects that have already been identified. This is the change caused by the appearance of new objects and the disappearence of existing ones. Describing the appearance and disappearance of objects is difficult when the existing objects are the only frame of reference. In theory, these changes can be described by changes in the aggregation structure of objects at a higher aggregation level (e.g. fragmentation, dissolution). Imagine that a new elementary object appears. Its appearance implies the fragmentation of the composite object to which it belongs. Similarly, the disappearance of an elementary object can be described by dissolution at the composite-object level. But what if the new object is a composite object? It would require defining a higher aggregation level to descibe the associated change, which, in turn, would require redefining the aggregation hierarchies. And this might involve redefining the aggregation structure of the existing objects.

We can describe these changes another way. For example, if a composite object at aggregation level 1 is split, the new objects, which also belong to aggregation level 1, will be associated by their common origin. So, if new objects appear by splitting or merging, only one level of aggregation is involved in the process. We refer to this type of change as change in association structure (Fig. 1.8). If new objects appear by fragmentation or dissolution, two levels of aggregation are involved.

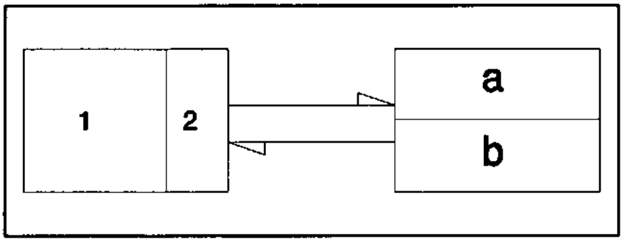


Fig. 1.8 *Change of object association structure.*

Such changes occur when, for example, regional structures develop in a formerly undifferentiated forested area because of colonization. The emerging LUZs can be described by their own aggregation structure and class structure. The changes are then not associated with the changes in the characteristics of a specific LUZ, but rather with the replacementof one LUZ by another or by several different LUZs.

Comparable changes occur when, for example, agriculture on a plantation is changed or diversified. The plantation is split into various areas, each with its own type of land use. This division implies the definition of new LUZs, which are associated because of their common origin.

On the other hand, different LUZs can be merged to form one area with a specific land-use pattern. This occurs when, for example, plots of land belonging to different owners are merged as part of a settlement scheme.

LUZS IN THE ATLANTIC ZONE OF COSTA RICA.

LUZs are used to describe land use at sub-regional level. Their effectiveness for this purpose is determined in part by the degree to which the terrain objects can be differentiated (or classified) on the basis of their composition characteristics. These characteristics stem from the distribution of the attribute values (or attribute value classes) of the elementary objects. Put another way: Can clusters be identified in the data space? And, if so, do the corresponding groups represent meaningful land-use categories? We shall be discussing whether LUZs represent a meaningful aggregation level for the description of land use in Chapter 3.

Land use is subject to change. Structural changes in land use, especially in the Atlantic

Zone of Costa Rica, can occur very rapidly, sometimes within only a few years. This means that a reliable procedure for monitoring land-use change is needed. If the LUZs are to be the carriers of land-use information, then land-use change must be expressed through the object's dynamic behavior. But, before we can investigate the object dynamics, we shall have to design a data model of sub-regional land use that will define objects, their attribute structure and their classification and aggregation structures.

Monitoring land-use change requires identifying the same objects at different points in time, an exercise that could be quite frustrating if all the object characteristics tended to change. Therefore, to monitor change, we must find a constant characteristic. The LUZ characteristic most likely to remain constant is the location of the boundaries (the LUZ geometric characteristic). Stationary geometric characteristics make it possible to evaluate land-use change in the LUZs by verifying changes in their thematic attributes values and their aggregation structure.

In Chapter 2, we show that stationarity of the boundaries usually occurs only in certain LUZ classes. In some of these classes (e.g. agricultural areas), we observe that changes in land use can be modeled through the aggregation structure or the class structure of the object. In other classes (e.g areas of colonization), change in land use is expressed primarily by change in geometry. (Think of a forest front retreating in the face of agricultural encroachment.) In areas of colonization, therefore, we need other methods to describe land-use change than those used for agricultural areas.

In Chapter 4, we evaluate the changes that occur in the thematic attributes of LUZs in relation to land-use change. The land cover composition being an important characteristice in this respect. Because LUZs are composite objects, change in their thematic attributes means change in their composition. Remote sensing is a good tool for monitoring object dynamics (Molenaar and Janssen, 1991), and so land-cover data from satellite imagery was used to evaluate the changes in the thematic characteristics of the LUZs.

CHAPTER 2

LAND USE ZONES THAT DEFINE PERMANENT SPATIAL UNITS FOR MAPPING LAND USE

INTRODUCTION

In the previous chapter we saw that the geometry of a terrain object must be stationary if the changes in that object's thematic characteristics and aggregation structure are to be evaluated effectively. Thematic characteristics depend in part on the object's position, size, and form, and they will be altered by any changes in these factors. If we wish to base an inventory of change in land use on a change in thematic characteristics, we must exclude the changes that have occurred because of a change in object geometry. For example, let us assume that a Land-Use Zone (LUZ) has stationary boundaries. Under these conditions, a change in land use within the LUZ will manifest itself principally through:

- A change in land cover, which at LUZ level means a change in land-cover composition;
- A change in parcelling, which appears as a change in the LUZs aggregation structure.

In this chapter, we shall look at the changes that can occur in the geometry and, to a lesser extent, in the aggregation structure of an object. Of course, not all LUZs will have stationary boundaries. They will probably have changing geometric characteristics, especially where forests and agricultural areas meet. (Think, for example, of a forest front retreating in the face of agricultural penetration.) Indeed, in that specific context, a change in land use will manifest itself as a change in LUZ geometry.

Forests are generally delineated as large, undifferentiated areas, in which spatial structures will develop in the wake of deforestation and because of differentiation in land use. Differentiation refers to splitting of LUZ (change in association structure) or to change in the aggregation structure as a result of fragmentation. Splitting of existing LUZs implies the definition of new LUZs, resulting in a different LUZ structure. It should not be confused with change in the LUZ's geometry.

In agricultural areas, we expect land-use change to occur within the existing LUZ because the spatial sub-regional structure is already crystallized. The object dynamics will then be expressed as change in the thematic attribute values (e.g. changes in the land-cover composition because of changes in the relative importance of crops) and in the aggregation structure (fragmentation and dissolution). But, we also expect some change in LUZ structure due to splitting and merging. (Think of consolidation of property.) In these cases, an inventory of change in land use cannot be based on change in thematic attribute values.

Thus stationarity of LUZ boundaries is a prerequisite to monitoring land-use change by observing change in thematic attribute values. It would be useful if we could infer boundary stationarity from boundary type instead of having to verify it for each individual LUZ. The boundary type is denoted not only by the contents of the polygons on either side, but also by the object class it represents. Property boundaries generally show up as rectilinear stationary features. Similarly, boundaries corresponding to large rivers and physical structures like escarpments, can represent permanent LUZ limits. Unfortunately, these correlations cannot always be defined.

We shall discuss the relation between boundary type and stationarity. But our main interest is in establishing whether LUZs represent stationary spatial structures. Hereafter, the terms "stable" and "permanent" will be synonymous with "stationary". The term "LUZ structure" will mean the spatial constellation defined by the whole set of LUZs.

PROCEDURE AND METHODS

The Procedure

The stability of the LUZ boundaries was examined by comparing two sets of aerial photo (AP) interpretations: one dating from the years 1948-1952 and the other from 1984. The comparison was done according to the steps listed below:
- Interpret the APs and delineate the LUZ;
- Digitize the AP interpretations;
- Perform numerical restitution and transform the geometric data;
- Classify the LUZ;
- Compare the LUZ maps and evaluate change.

The APs from 1948 and 1952 were black and white, with an approximate scale of 1:40,000. The 1984 APs were colour infrared photos and their scale was approximately 1:80 000. (AP interpretation is described in Chapter 2 of Part 2.)

Block Adjustment

The AP interpretations were digitized. The point positions are given in the camera coordinate system. To compare the LUZs from both periods, we must transform these interpretations into a common coordinate system. The method used here was numerical restitution. For each AP, we must determine the transformation matrix. The reference points for calculating the transformation matrix were obtained by block adjustment using independent models (Albertz and Kreiling, 1989). This method entails making independent stereo models and then linking them to each other to form a block, which is then referenced to the terrain.

To calculate the terrain coordinates for each point in the AP, we need to know that point's elevation. Accordingly, to obtain this information for the 1984 AP interpretations, we made a Digital Elevation Model (DEM) by digitizing the contour lines and interpolating between them. To transform the 1948-1952 AP interpretations, however, we used an average elevation of the area corresponding to each photo, as the topography of the area was fairly flat. This minor difference in approach causes a slight deviation in the positioning of the LUZs, especially in the higher areas.

Land-Use Classes

The 1948-1952 and the 1984 LUZs were classified. As there were no remotely-sensed spectral data from 1948-1952, neither were there detailed land-cover data from this period, so the classification of 1948-1952 LUZs was based completely on photo characteristics. Consequently, only the major classes of land use could be mapped. They are:
- Forests and areas of natural and semi-natural vegetation;
- Areas of agricultural penetration;
- Agricultural lands;
- Plantations;
- Settlement areas;

- Wetlands;
- Bodies of water;
- Other.

Hereafter, these will be referred to as "LUZ classes".

Evaluating Change in LUZ Geometry

The 1948-1952 and the 1984 AP interpretations were overlaid and compared. The various LUZ were identified and three categories of change were defined:

- Category 1, which contains 1948-1952 LUZs with no boundaries in common with 1984 LUZs;
- Category 2, which contains 1948-1952 LUZs with some boundaries in common with 1984 LUZs;
- Category 3, which contains 1948-1952 LUZs that correspond either to a 1984 LUZ (1:1), or various adjacent 1948-1952 LUZs that, taken together, correspond to a 1984 LUZ (M:1).

To decide whether the LUZ boundaries corresponded, a maximum distance was defined. This value was based not only on the geometric accuracy that is achieved with the geometric transformation of data, but also on the accuracy of delineating and digitizing of the LUZ boundaries. For the 1948-1952 and the 1984 transformations, there was a difference of about 30 m between the measured and the computed reference points. The margin of error for delineating LUZ boundaries and digitizing the data was assumed to be about 0.3 mm. On the AP scale of 1:80 000, this corresponds to 24 m. The relative error, when we compare the position of a point in one map to its position in the other, is obtained by adding the error in the individual sources. It will range from 100 to 150 m, which corresponds to 2 and 3 mm on the scale of 1:50 000 that was used for the comparison of the LUZ boundary information. So the lines on both maps can be regarded as corresponding boundaries if they do not differ by more than 2 to 3 mm. Applying these rules gives a first impression of the stationarity of the LUZ boundaries.

Quite often, the boundaries of the 1948-1952 LUZs do not correspond with those of the 1984 LUZs. These differences do not necessarily mean a change in LUZ geometry, however. The scale of the 1948-1952 APs is different from that of the 1984 APs, making it difficult to judge whether difference in the delineation of the LUZs is owing to actual change in LUZ geometry, to splitting and merging, or to the difference in the level of aggregation. This judgement is further complicated by the large time span between the two periods of observation. It makes it difficult relating the 1984 LUZ to the 1948-1952 LUZ (determining whether the 1984 and the 1948-1952 LUZ are in fact the same). To deal with this problem, we identified three situations in which difference in the delineation of the LUZs is not taken to indicate a change in LUZ geometry. These are:

- If the LUZs of 1948-1952 correspond to a part of the 1984 LUZ. (Change is then considered the result either of change in aggregation structure, i.e. dissolution, or the result of change in association structure, i.e that the 1984 LUZ corresponds to another aggregation level).
- If a 1948-1952 LUZ corresponds to several 1984 LUZs. (Change is then considered the result either of change in aggregation structure, i.e. fragmentation, or, more likely, the result of change in association structure, i.e. forming of new

LUZs).

- If the shape, size, and topology of the 1948-1952 LUZs and the 1984 LUZs agree, but the differences in the position of their boundaries exceed the threshold value. (Change is then considered the result of misinterpretation of the AP.)

But even with these guidelines, it can be difficult to decide if there has been change in LUZ geometry. There is some freedom in the interpretation of change in LUZ geometry. If we count only the 1948-1952 LUZs that are clearly intersected by the 1984 LUZs as evidence that the LUZ structure has changed, then the rate of geometric change will be relatively low. If we count all boundary changes greater than the specified threshold value, then the rate of geometric change will be relatively high. We can make a more or less conservative estimation of the area of LUZs with non-stationary boundaries.

Aside from investigating the possible causes for the difference in LUZ delineation between both periods in order to make statements on LUZ boundary stability, also the relation between change in LUZ geometry and LUZ class should be investigated. For each category of change the distribution of the LUZ classes is determined. Taking account of the LUZ classes as well as the different causes for change in LUZ delineation, conclusions on the stability of the LUZ boundaries are drawn.

Also investigated is the relation between boundary stability and type of the LUZ boundary. This was done by inventorying the boundary types and shapes per LUZ. The inventory was conducted at LUZ level, instead of looking to each boundary individually, because of the many uncertainties in establishing change in the position of the boundaries (due to the uncertainty in identifying a 1984 LUZ boundary as being the same as the 1948-1952 LUZ boundary). Thus, for the LUZs in each category of change, the class value of the adjacent LUZs was inventoried. The LUZs were categorized as having either at least one adjacent LUZ belonging to agricultural land for or not neighbouring to the agricultural land. The shape of the boundaries of each LUZ was inventoried and categorized as follows:

- I Only irregular boundaries;
- F At least part of the LUZ boundaries correspond to rivers, escarpments or other terrain features;
- S Straight linear boundaries for at least a part of the total LUZ boundary, being interpreted as property limits.

RESULTS AND DISCUSSION

Incidence of All LUZs in the Change Categories

Table 2.1 shows the incidence of LUZs in each of the three categories of change. The figures express area percentages. The total area is the sum of the areas of all the LUZs of 1948-1952, minus the areas of extended forest. Investigating change in LUZ geometry is not relevant for these forest areas, because no LUZs exist.

We see that not more than 40 per cent of the total area the LUZ structure in 1948-1952 corresponds to the LUZ structure in 1984. It indicates that we cannot assume stationarity of LUZ boundaries in general. The data are obtained by looking to difference in LUZ delineation without taking account of the various causes for the difference in delineation. As we shall read in the Methods section, these data alone do not express stationarity of the LUZs geometric characteristics.

Table 2.1. *Incidence of LUZs in the three change categories (expressed in % of the total area)*

15.0%	Category 1 : No correspondence
45.7%	Category 2 : Partial correspondence
39.3%	Category 3 : Correspondence

Incidence of LUZ Classes in the Change Categories

The incidence of the various LUZ classes in each of the three change categories was inventoried. The results are shown in Table 2.2.

Table 2.2 *Incidence of LUZ classes in the three change categories (in %)*

(LUZs of 1948-1952)	Change category 1 (change)	Change category 2 (some change)	Change category 3 (no change)
Class 1 Natural and semi-natural vegetation	47.6%	35.4%	32.9%
Class 2 Area of agricultural penetration	35.7%	26.2%	17.6%
Classes 3 and 4 Agricultural land and plantations	16.7%	38.4%	49.4%

The majority of the 1948-1952 LUZs in category 1 (change) belong to classes 1 and 2, confirming the assumption that LUZs pertaining to these classes may not be assumed to have stationary boundaries. Nevertheless, there is still a strikingly high incidence (32.9%) of class 1 LUZs in category 3 (no change). And, to a lesser extent, this is true for class 2 LUZs (17.6 %). This indicates that also in the forest areas and areas of agricultural penetration stationary structures occur, despite change in land use. The stationary structures can be attributed in part to former banana plantations. The plantations were abandoned between 1920 and 1940, after which a dense secondary vegetation established.

The majority of LUZs in category 3 belong to class 3 (agricultural land) and class 4 (plantations), which, for the sake of convenience, have been grouped together here.

Changes in LUZ Geometry

To draw conclusions on stationarity of the LUZ boundaries from differences in the delineation of LUZs as obtained by AP interpretation, the various explanatory factors should be taken into account. As mentioned in the Methods section, the difference cannot be attributed exactly to one of these factors, leaving some room for interpretation. This is reflected in the minimum and maximum figures denoting area with change in the LUZ geometry. (This can be LUZs with change in all or in part of its boundaries.) Maximum and minimum areas of change are given in Table 2.3. The percentage of maximum change in

LUZ geometry was obtained by accepting all questionable cases as representing change. The percentage of minimum change was obtained by accepting all questionable cases as evidence of stationary boundaries. The area of maximum change was 4033 hectares, that of minimum change was 1618 hectares. The percentages of change were obtained by dividing the area of change by the total area.

Stationarity of the LUZs boundaries may not be assumed for LUZs belonging to the area of agricultural penetration and forest areas, but it often occurs. Because of this reason, calculations are made for both, including as well as excluding these areas. This is reflected in the two ways used to calculate the total area.

The first was to add together the area of all LUZs except for those with extended forest and those corresponding to isolated spots of areas of agricultural penetration (or areas of secondary vegetation) within the forest area. The total area calculated in this way includes the majority of the areas of agricultural penetration. The corresponding percentage of the area of change will underestimate the actual the area of change. Although many of the LUZs that were areas of agricultural penetration in 1948-1952 did not correspond to the 1984 LUZs, only a limited number was taken to actually represent change in LUZ geometry, because of reasons given in the Methods section. However, these LUZs are accounted for in the total area. The total area was 35 955.6 hectares.

The second way of calculating the total area excluded the areas of agricultural penetration and the forest areas, except for those LUZs with stationary boundaries and those for which change in LUZ boundaries was concluded. The total area amounted to 18 613 hectares when departing from the maximum area of change of 4033 hectares (because a larger part of the area of agricultural penetration was taken to represent LUZs with stationary boundaries, which were, therefore, included in the calculation of the total area). A total area of 16 181 hectares resulted when departing from 1618 hectares as total area of change. The percentages in this second way of calculation probably overestimate the actual percentage of the area of change. The figures are presented in Table 2.3.

The figures indicate that not more than about 20 percent of the area exhibits non-stationary geometric characteristics. In other words, a cautious estimation gives an 80 % of the area representing a stationary LUZ structure. This confirms the assumption that once a LUZ structure has established it represents a stationary spatial structure.

Table 2.3 *The total area of change in LUZ structure, calculated by different methods (in %)*

	Maximum area of change (4033 Ha.)	Minimum area of change (1618 Ha.)
I. Total area excluding extended forest areas (35956 Ha.)	11.2%	4.5%
II. Total area excluding area of AP and forest area (18613 and 16181 Ha. respectively)[1]	21.7%	10.0%

[1] Total area depends on whether maximum or minimum area of change is taken. For explanation see text.

Incidence of LUZ Boundary Types in LUZ Change Categories

Boundary stationarity is often related to boundary type. For each LUZ in each of the category of change, the classes of the adjacent LUZs were investigated. The results of the investigation were categorized according to whether only LUZ classes 1 and 2 were neighbours, or to whether there was at least one LUZ of class 3 or 4 as a neighbour. The LUZs were then grouped according to their LUZ class and category of change (Table 2.4).

Table 2.4 *Incidence of LUZs bordered by at least one LUZ belonging to agricultural land or plantation per category of change and LUZ class (in %).*

	Change category 1 (change)	Change category 2 (some change)	Change category 3 (no change)
Class 1 Natural and semi-natural vegetation	30% (N=20)	39% (N=15)	67% (N=7)
Class 2 Area of agricultural penetration	13% (N=23)	35% (N=17)	75% (N=25)
Classes 3 and 4 Agricultural land and plantations	57% (N=9)	92% (N=8)	87% (N=15)

We can see that, of the class 1 LUZs in category 1 (change), 30 per cent have at least one class 3 or class 4 LUZ as a neighbour. (Put another way: 70 per cent of the LUZ belonging to forest land and showing non-stationary LUZ boundaries, is bordered either by more forest land or by areas of agricultural penetration.) In category 3 (stationary boundaries), 67 per cent of the class 1 LUZs are bordered by at least one LUZ pertaining to agricultural land. We can see similar increases if we compare the change categories for class 2 LUZs and for class 3 or class 4 LUZs. This strongly suggests that the stationarity of the boundary is related to the type of the boundary. This relation is especially interesting for class 1 and class 2 LUZs. If proven, it could provide criteria for predicting stationarity of LUZ boundaries, which would be helpful for those situations in which stationarity of LUZ boundaries cannot be verified by long-term measurements.

Boundaries are also characterized by their shape, namely irregular (type I), straight (type S), or curved, as when corresponding to a river (type F). S-type boundaries generally represent property limits. Their occurrence in each of the different change categories was investigated. The results are presented in Table 2.5.

Table 2.5 *Incidence of LUZ with I-type, F-type or S-type boundaries types in the three change categories (in %).*

	Change category 1 (N=35)	Change category 2 (N=40)	Change category 3 (N=32)
Only type I boundaries	60%	18%	3%
Type F boundaries occur	40%	70%	(No data)
Type S boundaries occur	6%	43%	72%

Type I boundaries are not expected to occur in LUZs with, at least partially, stable boundaries. If we assume that type S boundaries are stable boundaries, then a partial change in LUZ boundaries implies the occurrence of type S boundaries. The data in Table 2.5 seem indeed to suggest that S-type boundaries tend to be stationary. The percentage of stationary

LUZs having S-type boundaries is higher than for LUZs having partly stationary or non-stationary boundaries. However, S-type boundaries are not a requisite for stationary boundaries, nor does it assure stationarity of the LUZ boundaries.
The reverse trend is visible for type I boundaries. The percentage of LUZ having only type I boundaries clearly decreases from change category 1 (change) to 3 (stationarity). Stationary LUZs seldom have only type I boundaries. Type F boundaries are observed for LUZs in all categories of change.

ILLUSTRATING CHANGE IN LUZ STRUCTURE

Sample Area 1

Figure 2.1 illustrates stationary LUZ boundaries in the GRS study area. We see that many of the structures observed in 1984 were already established in 1948. Look, for example, at LUZs 58, 59, 61, 142, 62, 63, 66, 140, 60, 64, 65, 155, 67, 68, 143 and 139. Their boundaries in 1984 (continuous lines) correspond either partly or completely to the boundaries observed in 1948 (broken lines). Most of the stationary boundaries are rectilinear lines corresponding to property boundaries. LUZs 59 and 69 show stationary boundaries that are formed by rivers.

Most of the LUZs are located alongside the railway, perpendicular to the tracks. The colonization of the Atlantic Zone started at the end of the last century with the construction of the railway, and in 1948 the train was still the primary means of transport. The claims on large parts of the land in this area (more than 500 ha) and the colonization pattern are strongly related to rail infrastructure. The structure of 1948 was still evident in 1984 and in 1990, the last year for which a Landsat-TM was obtained (not shown).

But we do see a few changes in LUZ delineations. In 1948, LUZs 63, 66, and 140 were defined as one area, corresponding probably to a rubber plantation. In 1984 three LUZs were identified, as consequence of differentiation in land use (splitting) In 1990, it shows as one area again, this time as a macadamia nut plantation. Through this process of splitting and merging the association between the LUZ becomes clear. This is a good example of how information on the LUZ structures of 1948-1952 can improve the 1984 AP interpretation. The correspondence to a single enterprise or management unit could be deducted from the very regular form the three LUZs together represent. The boundaries between the LUZs, therefore, are to be regarded as temporary limits.

Let us now look at LUZs 64 and 65. The straight outer boundaries, clearly recognisable in 1948, indicate that both LUZs are one farm, and so the boundary between them should have been marked as only temporary. Two LUZs were defined because of difference in cover characteristics. Two separate LUZs were still defined for the 1984 situation, based on small differences in the land cover and the lay-out of the fields, which were related to differences in terrain characteristics. The land cover and land use in both units did not differ any more in 1990.

Still another example of association between LUZ is the case of LUZs 155 and 67. Both belong to the same enterprise. The property boundaries are recognisable in the 1984 AP because of the marked and straight boundaries. The area is currently being used to grow ornamental crops. The differentiation of two LUZs in 1984 was based on the difference in land cover, which had already changed by 1986. The property limits are also recognisable

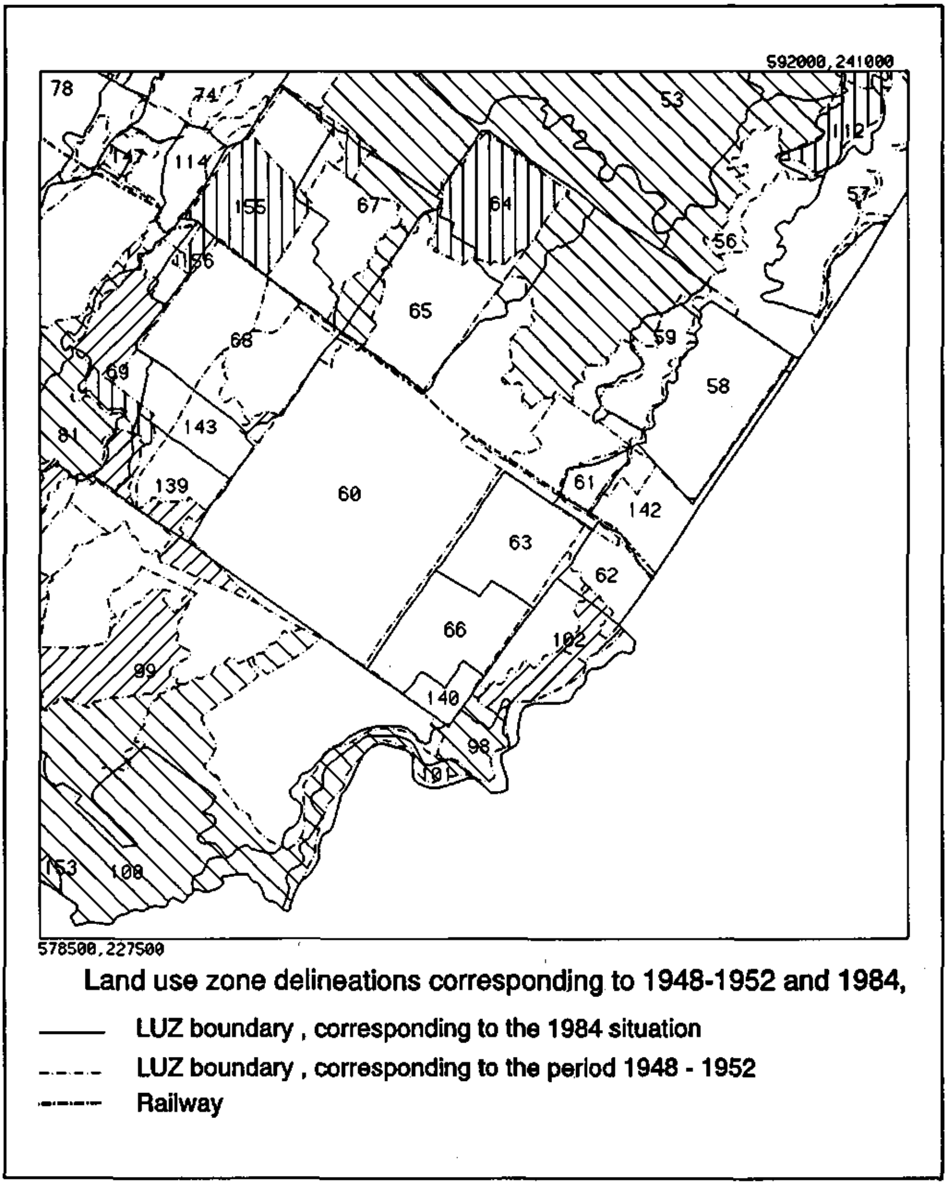


Figure 2.1. Stationary LUZ structure in agricultural areas (for explanation of shade symbols see Fig 2.2).

in the 1948 photo, while the LUZ structure within this area differs considerably. Some of the LUZ boundaries were not considered permanent features because they represented a transition from forest to (partly) cultivated land. The change in LUZ structure could then be attributed to the different scale of the 1948 photos or, indeed, to a process of dissolution and subsequent fragmentation. Whatever the explanation, the important thing is that the changes were confined to the area defined by the straight outer boundaries of the LUZs visible in both the 1948 and the 1984 aerial photos. Almost the same story can be told about LUZs 68, 139, and 143.

Let us look at LUZ 60. This area consists of many small farms instead of a large enterprise, and its boundaries have proved to be fairly stable over the years. Except for one part, which was under forest cover in 1948, no changes have occurred. LUZ 60 probably belonged to a banana plantation at the beginning of this century. Later it was abandoned and the land distributed to small farmers. The companies provided the farmers with cacao, and the cacao fields, though often neglected, are still very characteristic of the zone.

The LUZs farthest from the railway were classified as forest or as areas of agricultural penetration in 1948. These areas were not differentiated further, and their boundaries are not assumed to be permanent features. The area corresponding to LUZ 99 shows some differentiation in 1948, but this was temporary. It later became a settlement scheme, implying that a redistribution of the land had occurred under the auspices of the Institute for Agrarian Development (IDA). Generally, the IDA buys larger farms for distribution among small farmers, a practise that preserves the LUZ structure.

Sample Area 2

Sample area 2 is presented here because it differs from the earlier example area in that the larger part was still covered by forest or overgrown by dense secondary vegetation in 1952 (Fig.2.2). But even in the non-agricultural LUZs, we can see clear rectilinear features, probably traces of the property boundaries of former plantations.

Let us look at LUZs 30 and 28. Their straight broken lines represent stable spatial features where they correspond to the continuous lines of the 1984 AP interpretation, suggesting one large property. And indeed, LUZs 30, 28, and 29 still belong to the same owner today. The broken lines within the area spanned by LUZs 28 and 30 represent only small changes in photo texture, which are not significant as they relate to slight changes in the density of the same vegetation.

We see a resemblance between the interpretations of 1952 and 1984 in the delineation of LUZ 22. This LUZ was classified as one banana plantation in 1984. But in 1952, only the north-east part could be clearly identified. It was distinguished from the south-west part by the difference in land use, the boundary between the two parts being defined by river. The historical data would suggest that the land of LUZ 22 is divided among different owners. And indeed, the two different parts are the property of two different owners, although both parts are banana plantations. This is another instance where information on earlier LUZ structure would have helped to interpret the 1984 APs correctly.

The other LUZs are bounded by irregular linear features that represent a transition from forested to non-forested areas. They do not represent permanent features, and the spatial structure of 1952 and 1984 consequently does not correspond.

CONCLUSIONS

The dynamic behaviour of LUZ as regards to their geometry differs from one LUZ class to the next.

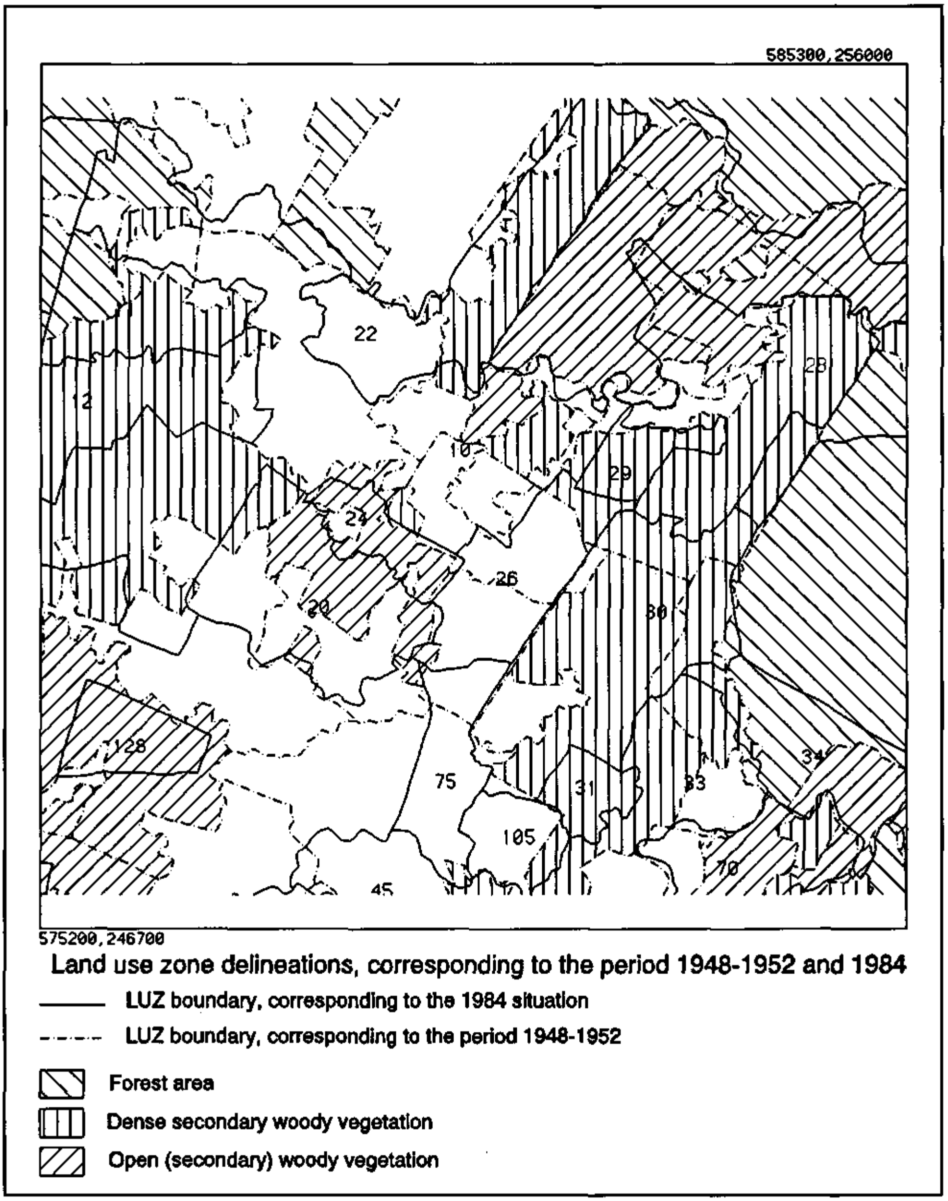


Figure 2.2. *Stationary structures in non-agricultural areas.*

1. The spatial structure of the LUZs representing agricultural areas is stationary. This stationarity occurs despite the short history of and the rapid changes in land use in the region. This means that change in land use can be evaluated by investigating change in the thematic content and the aggregation structure of the LUZs.

2. The geometric structure of LUZs representing areas of agricultural penetration and forests may not expected to be stationary. Indeed, change in spatial structure can be considered characteristic of these areas. Even so, stationary and partly stationary structures can still be identified.

3. There is a relation between the type and the stationarity of LUZ boundaries. LUZ boundaries are defined here by their shape and by the LUZ class of the LUZs on its either side. Straight boundaries corresponding to property limits are generally stationary. Irregular boundaries tend to be non-stationary. Boundaries denoting a transition from one agricultural area to another and from an agricultural area to an area of agricultural penetration tend to be more stable. Boundaries denoting a transition between areas of agricultural penetration and between those areas and forest tend to be less stable. Boundary type could be a useful tool in monitoring change in areas of agricultural penetration and of forest, indicating which boundaries might be assumed stationary.

4. Information on the former LUZ structure can help to determine the current LUZ structure from aerial photos, especially when boundary type is being considered. Association between LUZs can be determined by comparing LUZ structures from different periods.

CHAPTER 3

LAND USE ZONE, LAND COVER COMPOSITION AND LAND USE PATTERN

DEFINING LAND USE PATTERNS IN RELATION TO LAND COVER COMPOSITION

1. INTRODUCTION

LUZs are delineated on the basis of spatial patterns, as observed on aerial photos. The pattern elements refer to agricultural fields. The aerial photo interpretation is described in Chapter 2, Part 2. Subsequently field size distribution (Chapter 4, Part 2) and land cover composition (Chapter 5, Part 2) of the LUZ is determined; classes are defined that express the correspondence or difference in these LUZ characteristics; and the LUZs are assigned class labels. From the classification result we want to infer information on land use. It requires the interpretation of the land cover composition and field size in term of land use characteristics. It is the subject of this chapter. With respect to the land use classification the question is the following:

- First, do the LUZs provide a relevant geographical basis for the description of land use at sub-regional level? Can the LUZs be discriminated on the basis of their land use characteristics?
- Second, if the first question is answered confirmatively, can we relate land use characteristics to the land cover composition and field size characteristics of the LUZs, so that we can map land use on the basis of land cover and field size characteristics.

The term data classes will be reserved to denote the field size classes and the composite land cover classes (CLCCs). That is, to denote classes that refer to the imagery derived LUZ characteristics. The term information categories refers to the classes describing the land use of the LUZs (i.e. LUPs). The translation of data classes to information categories will be referred to as mapping. The rules for translation (or interpretation) are named mapping rules.

The translation of field size classes into farm size classes is the subject of Chapter 4 of Part 2. In this chapter we are concerned with the interpretation of CLCCs in terms of land use patterns. The objective is to define rules for this interpretation. The definition of the appropriate mapping rules requires insight in the land use or agricultural activities in the region and insight in the object image characteristics. The relation between (composite) land cover class and land use (pattern) is ambiguous. The correct interpretation will depend on the context. The context, in fact, limits the range of the mapping of the data classes to the information categories. That is, the context defines a subset of the information categories to which the data classes can be mapped. To use context information we must be able to denote the context and to specify the associated domain.

Land use refers to the functionality of the land, c.q. the purpose for which the land is used and the way in which it is used. This is much in agreement with Clawson and Stewart (1965) who define land use as man's activities on land which are directly related to the land. Land cover refers to what is found on the land surface. See Burley (Rhind and Hudson, 1980), who defines land cover as 'vegetation and artificial construction covering the land surface'.

For the description of land use different concepts have been developed. The FAO (1984) defines the land utilization type (LUT), which "refers to a crop, crop combination or cropping system with a specified technical and socio-economic setting". The definition can be applied at different levels of detail. "A single crop can be regarded as a LUT", but it can also be appropriate to regard "farming system or cropping system as the definition of land utilization type". As such, the definition does not take into account the specific aggregation level. The classification and aggregation hierarchies are not very well defined in land evaluation (LE). Multiple and compound LUTs are defined, however, not in relation to levels of aggregation.

In this thesis the following conventions and definitions are adopted:

1. The LUT is used to describe land use at field level. It describes the cropping system with a specified technical and socio-economic setting.
2. At farm level different activities can be combined. The farm system describes the combination of activities (referring to production of crops or livestock) at farm level, with associated management characteristics. Farming system is used to indicate the classes of similar farm systems (Fresco and Westphal, 1988).
3. At sub-regional level the land use pattern (LUP) is used to describe the association of LUTs or farming systems. When the LUZ corresponds only to a part of a farm or enterprise, the LUP will refer to one single LUT. Sometimes, the LUP will refer to only one farming system.

For example, the LUP of a LUZ that corresponds to a single plantation will be described by one LUT, namely as 'large scale banana production with an advanced management level'. For LUZs that comprise many farms the occurrence of only one LUT or one farming system is very unlikely. Instead, various farming systems will often occur. However, the combination of farming systems will not be arbitrary. The LUZ approach assumes the existence of specific land use patterns. This assumption is verified in this chapter.

2. PROCEDURES AND METHODS

The definition of LUPs requires information on land use characteristics of LUZs. This information was obtained through a farm survey. The survey was carried out especially to the LUZs that comprise many farms, because these are the most difficult to characterize in terms of LUP. The procedure was as follows:

- Interviews, carried out in the study area;
- Definition of farming systems and classification of the farms visited;
- Evaluation of the distribution of farming systems for selected LUZs and definition of land use patterns.

The farms were not selected fully at random, in so far that no list was made of all farmers in the area from which a random selection could have been taken. Such a list could only be obtained at the expense of extreme efforts and costs. Therefore, farms were selected at locations spread across the zones. Generally a few neighbouring farmers were interviewed. Cross checks were done by asking neighbours about their surrounding farms or by inquiring at the pulperias (local stores where groceries and drinks are sold) about the farms in the neighbourhood.

Definition of farming systems

For the purpose of this study a rather general classification of the farm systems was carried out (see Fresco and Westphal, 1988, on classification of farm systems).
The following criteria were used :

- The main farm activity, being the cultivation of crops, animal husbandry or a combination of both. It will be referred to as farm proposition. The proportion of the farm area dedicated to the different activities is considered. Mixed farming is concluded only when more than 10 % of the farm area is dedicated to each activity.

- Management level, determined by the use or possession of machinery (especially tractors and accessories for field preparation), investment in buildings and the use of hired labour.
- Other criteria like off-farm income enjoyed by the owner or family members and whether the owner lives on the farm.

For the description of the farming systems see the appendix.

The low variation in management levels, as reported by Rhind and Hudson (1980) for many tropical countries, also applies to the agriculture in the Atlantic Zone of Costa Rica. Either large commercial (often international) enterprises are encountered, with a high level of management (use of advanced techniques, high level of investment in machinery and buildings and much use of hired labour, e.g. the banana plantations) or farms with 'traditional' agricultural practices are found. The latter implies that labour is provided by family members, with sometimes the hiring of temporal workers for specific activities; The preparation of the fields and other activities are mostly done manually with the use of simple tools (sometimes a tractor is hired for ploughing of the fields); Little investment in machinery and buildings, but with common use of herbicides, pesticides and fertilizers.

Intermediate forms are practically absent. Only a few LUTs with intermediate management level are found. The most important is the cultivation of ornamental crops: sprinkler irrigation is commonly found; nets are applied to protect the plants against sun irradiance; also, the land is used rather intensively, requiring high labour inputs.

The scale of production is an additional criteria adopted to differentiate between farms with a corresponding LUTs, but with varying levels of input. It allows to differentiate between the larger and smaller cattle farms that do not differ principally in management level. Also distinction could be made e.g. between farms with only a few hectares and with up to tens of hectares dedicated to the production of maize, with similar cultivation practices (e.g. inputs per hectare may not differ).

Definition of LUPs and rules for mapping

For each LUZ the composition in terms of farming systems can be assessed. Defining LUPs involves the definition of specific classes of farming system composition. The requisite for the LUPs is that these can be linked to CLCC and the field size class, because we want to use these data classes for the mapping of LUPs. But also that the classes are distinct. Therefore, the distribution of farming systems is listed for each combination of CLCC and field size class. The data are aggregated for LUZs with corresponding CLCC or farm size class. On the basis of these data differences between the LUZs is investigated and the LUPs are defined.

As stated in the introduction, the relation between the land cover class composition and LUP is not unambiguous. The specification of the mapping rules for the interpretation of the (composite) cover classes requires insight in the land use characteristics in the region and how these come to expression in the satellite imagery and the aerial photos. Insight in the regional land use characteristics is obtained by means of the earlier mentioned farm survey and by consulting the relevant literature. The literature (Waaijenberg, 1990; Wielemaker, 1990; Oñoro, 1990) describes the land use in different regions in the Atlantic Zone. The description is based on a survey carried out in 1987.

Especially of interest is the possible relation between farm size and LUT, because we dispose of information on farm size at LUZ level (Chapter 4, Part 2). The information might serve to support the interpretation of the composite land cover. The relation between farm size and land use is investigated by analyzing farm data. However, the LUZs represent aggregated units. Whether difference in land use in relation to difference in farm size is expressed also in difference in LUP in relation to farm size class must be verified.

Mapping of land cover composition involves the interpretation of the individual land cover classes (LCCs) in terms of land use. The per pixel classification of land cover resulted in the identification of general LCCs, like 'wooded area' and 'secondary vegetation'. The interpretation of the general cover classes in terms of specific cover classes or LUTs (e.g. 'cacao' or 'grassland with fruit trees') is dependent on the context. The LUZ may provide the relevant context. The context can be denoted by the general LUZ class (or general LUP) which can be derived from general characteristics, like the presence of agricultural fields. See Chapter 6, Part 2 for the hierarchical structure of the LUZ classification.

3. RULES FOR MAPPING LAND COVER CLASS TO ACTUAL LAND COVER TYPE.

For the land cover classification 17 land cover classes were defined. The description of the land cover classification can be found in Chapter 3, Part 2. The interpretation of these classes is not always uniform. The term cover type is used to denote specific land cover categories that can be distinguished in the field, but that can not be recognised spectrally, which is the case for the land cover classes. For the following three important land cover classes the interpretation is discussed:

- Bare Soil and Built-up area (BB)
- Wooded Area (WA)
- Pasture and grasslands (Pas)

Bare Soil

Maize fields are recognised as 'bare soil', given the recording date of satellite image (february 6th) and the period for field preparation for the cultivation of maize (between december and march). However, 'bare soil' may also correspond to other crops. Beans are sown in the same period, but mostly on very small plots because the produce is generally for home consumption only. Because of their size the plots will, in general, not be recognized. Root and tuber crops, such as Yucca (*Manihot esculenta*), are cultivated throughout the year. Thus, 'bare soil' might in cases also correspond to these crops. Decision as to which crop 'bare soil' will correspond depends on the context.

Maize is generally found on small and medium sized farms. When larger enterprises are concerned bare soil should be interpreted differently, e.g corresponding to root and tuber crops (Yucca, Chamol [*Colocasia esculenta* var. *antiquorum*] and Tiquisque [*Xanthosoma sagittifolium*]) or for the planting of ornamental crops (among others Caña India [*Dracaena massangeana*], *Marginata spp.*, *Helioconia spp.* and *Philodendron spp.*), or to clearings for other purposes. Plantings of ornamental crops and seedlings, whereby nets are used for their protection against the sun irradiance are classified as bare soil as well.

This means that when the spatial pattern indicates the presence of smaller farms, bare soil areas are assumed to correspond to maize plots. When larger commercial farms are concerned no direct conclusion on cultivated crop can be drawn.

Wooded Area

'Wooded area' (WA) may correspond to a variety of land cover types and/or crops. Any densely wooded area, like river banks, orchards, tree plantations, cacao plots or homestead gardens, are classified as such. In many cases distinction between these cover types can be made on grounds of spatial characteristics. When a larger homogeneous area is concerned (> 100 ha.) generally a macadamia (*Macadamia integrifilio*) plantation is found. In general fruit orchards do not extend to such sizes in the Atlantic Zone. An other possibility in this case might be reforestation or tree plantation (mostly of laurel [*Cordia alliodora*] or Eucalipto [*Eucalyptus spp.*]), but generally these show less homogeneous characteristics. When 'wooded area' is found in smaller units it can correspond to:

- Cacao plots (*Theobroma cacao*);
- Orchards of different fruit trees (e.g orange [*Citrus spp.*] or guanabana [*Annona muricata*]);
- Pastures with fruit trees or;
- Homesteads in which a large variety of trees can be found (coco [*Coconut nucifera*], citrus [*Citrus spp.*], Guavo [*Inga spp.*], Guayabo [*Psidium guajava*], to name some common species).

The units are too small to be mapped separately or to be recognised as one of the specific cover types mentioned before.
'Wooded area' can also correspond to river banks and waste lands with high densities of trees. They have a deviant spatial pattern through which they can be recognised. In the colonization areas larger parts, often poorly drained, are found with high density of trees. These are areas that have been deforested, but where valuable trees have been preserved. Or they may correspond to areas with a dense woody secondary vegetation. On other areas with unfavourable terrain conditions (steep slopes, shallow soils, rock outcrop, etc.) a woody cover is often encountered.

When a LUZ consists only of 'wooded area', the correct interpretation can generally be given on the basis of the spatial pattern characteristics, terrain features or other. In case the LUZ has a composite character the data on the cover composition provides information for correct interpretation of the cover class. For example:

- If a high percentage of forest in combination with secondary grassland (indicating area of agricultural penetration) is observed, the WA will correspond to stands of trees (preserved while cutting the forest) and secondary woody vegetation.
- If the field pattern indicates small farms, and a high percentage of WA (more than 40 %, see Chapter 6, Part 2) is observed in combination with small percentages for grassland and bare soil, then the WA will partly correspond to cacao stands. A high threshold value for WA percentage is required to decide on cacao, because a part of the area classified as WA will be explained by wooded river banks, homesteads and other.

Pasture and grasslands

Two classes of grasslands were defined for the land cover classification. One class corresponds to cultivated pasture with a homogeneous grass cover, characterized by little variation in height of the grass vegetation and no or only limited presence of shrubs or trees.

The pasture types are 'ratana' pasture (*Ischaemum ciliare*), mixed pasture of 'ratana' and 'natural' (*Axonopus compressus*), 'estrella' pasture (*Cynodon nlemfluensis*) or pasture with brachiaria species (eg. *Brachiaria ruziensis*). The specific cover types can not be recognised spectrally but rules for the interpretation of this cover class can be given taking account of the farm size (see section 5 hereafter).

The other class of grasslands corresponds to non-cultivated or neglected grasslands. It is characterized by mostly a strong variation in height of the grass vegetation and the grasslands are often - but not necessarily - heavily invaded by shrubs. This class can refer to spontaneous vegetation found in and around lagoons or in poorly drained depressions, or they can be the consequence of negligence and poor maintenance of the pasture. Species encountered are 'Alemán' (*Echinochloa polystachya*), 'Guinea' (*Panicum maximum*), *Paspalum virgatum*, 'King-grass' (*Pennisetum purpureum*) and 'Cola de Venado' (*Andropogon bicornis*) among other species.

The (semi-)natural grasslands which correspond to the lagoons and other in periods inundated areas, are recognised by their spatial pattern. This can be the absence of faces (absence of a field pattern and no or little variation in vegetative cover) or the gradual transitions when variation in vegetative cover is present. The poorly drained bottom lands correspond to 'cut off valleys'. They are related to a specific landscape in which the bottomlands are clearly recognised by the lobbed spatial pattern. In the other cases the 'non-cultivated grassland' indicates neglected grasslands and it provides, as such, information on management.

4. GENERAL LAND USE PATTERN ASSOCIATED WITH THE COMPOSITE LAND COVER CLASS

The translation of the cover class to a specific land use is mostly rather straight forward in cases where the LUZ consists of only one cover class. Examples are the LUZs corresponding to plantations. When banana is the dominant LCC the zone corresponds to a banana plantation. In all cases this implies the production of banana on a large scale with application of advanced techniques and high investments. Plantations with the larger parts classified as 'wooded area' refer to Macadamia plantations. Also in these cases there is little differences in management. Such cases are not further discussed.

In Table 3.1 the number and percentage of farms are listed according to their main activity for each of the following general CLCCs:

1. With cultivated pasture as dominant and wooded area as sub-dominant LCC.
2. With a combination of wooded area, bare soil and pasture.
3. With wooded area as dominant LCC.
4. With the cover composition dominated by forest (area of agricultural penetration).

The data presented in this table as well as in the tables that follow hereafter are all originate from the farm survey.

The area with a land cover dominated by grassland indeed shows a clear dominance of farms only dedicated to livestock production. In areas where different cover classes occur, without clear dominance of either of these, we observe also the highest frequency of mixed farms (with pasture for grazing and crops), although farms solely dedicated to animal husbandry or to crop production are also observed.

Table 3.1 *Distribution of farming system per major CLCC.*

		Farming system[2]				
General CLCC[1]	N[4]	Crop production	Livestock	Mixed farming	Wood/ Livestock[5]	Wood/Mixed
1. Grass	38	1 (3%)	33 (87%)	4 (10%)	-	-
2. Mixed	56	10 (19%)	19 (35%)	25 (45%)	-	-
3. WA	9	5[3] (56%)	1 (11%)	3 (33%)	-	-
4. AP	11	1 (9%)	1 (9%)	0	7 (64%)	2 (18%)

[1]) These are the general CLCCs as described in the forgoing text. The numbers correspond to the numbers indicated in the text.

[2]) The definition of farming systems is based on the main farm activities, being crop production, livestock production, wood extraction or a combination hereof. For a description of the classes see the appendix.

[3]) At four of the five farms visited the crop concerned was cacao. It will be mapped as 'WA'.

[4]) Number of observations. This is the number of farms visited.

[5]) Corresponding to farms with often a considerable part of the farm area under forest cover. Logging is often an important activity. Livestock can not be considered the main activity or the main source of income.

The third general CLCC is distinguished from the previous classes by the high WA percentage (more than 40 % of the area is classified as WA). The large share of WA in the CLCC is unlikely explained only by the presence of wooded river banks, trees scattered over the pastures and some orchards. The high percentage implies a large portion of the area dedicated to the cultivation of tree crops. This indeed proved to be the case. Nine farms were visited. Of these 5 farms were classified as dedicated to the production of only tree crops (cacao, coffee and fruit trees), 3 farmers mentioned tree crops and pastures and one farm had only cattle. Thus, the high percentages of 'wooded area', under the condition that a field pattern is clearly observed, indeed indicates the importance of tree crops (especially cacao) in the land use pattern.

The results show a clear difference in percentage of farming systems occurring for the general CLCCs. It indicates the existence of different LUPs and that these LUPs may be related to the CLCC.

The area of agricultural penetration (AP) defines a completely different context for the interpretation of the land cover classes and composition of the units. The fourth class represents these areas. Activities are less directed to the commercial use of the land for crop or livestock production. Agriculture and cattle breeding seem to be of little importance (Waaijenberg, in: Wielemaker, 1990). Waaijenberg mentions deforestation and land speculation as important activities. Generally few crops are found, and the produce is for home consumption. The limited availability of labour and bad transport conditions make the agricultural activity little profitable. The fact that owners do often not live on the farms indicates the importance of land speculation in the areas. An other phenomena which is illustrative for the deviating situation is the presence of 'precarismo' (squatters).

The data on the 11 farms visited show that a high percentage of forest cover for the area in total, coincides with high percentages of forest cover for the individual farms. Nine farms had a part of the farm area covered with forest, varying form 9 to 69 percent. Often a considerable area was covered with tacotales (secondary woody vegetation). Generally the forest, upon deforestation, is converted to grassland. The land often does not allow for another type of land use, given the physical (often very wet conditions) and socio-economic conditions. 10 of the 11 farms had grassland. The land use of 7 farms was classified as a combination of forest and pasture. Three farmers mentioned the cultivation of crops like maize, rice, yucca or plantain for home consumption and were classified accordingly.

5. LAND USE PATTERNS ASSOCIATED WITH GRASSLAND AREAS

Farm size based rules for the interpretation of land cover.

In the forgoing section the correlation between the major CLCC and general land use pattern is discussed. More specific statements concerning the LUP can be given, when the composite land cover is considered in relation with field size. Field size is related to farm size and farm size is related to land use and management level. Waaijenberg (1990) reports that in the Atlantic Zone of Costa Rica the proportion of the farm area for crop production tends to increase with decreasing size of the farm. Nevertheless, livestock production often remains part of the farm activity irrespective of the size of the farm. But, on small scale it is less a commercial activity and of less importance for income generation. The herd produces milk and calves for home consumption and for the market. The calves are sold when in need for money. It serves as a banking account (Waaijenberg, 1990). This seem to be confirmed by the data presented in Table 3.2. Farms that are dedicated to the breeding, development and fattening of cattle are, in general, larger than farms that are (partly) dedicated to milk production. The smaller farms use the milk for home consumption.
No indication is obtained on differences in animal load in relation to farm proposition or to farm size (see Table 3.2). For the conversion of different categories of animals to livestock units (corresponding to 450 k.) data from Costa Rica were used (CATIE, 1986).

Table 3.2 *Farm proposition and average farm size regarding farms dedicated to livestock production.*

Proposition	Average farm size	Animal load	Number of observations
I. Cattle breeding	124 Ha.	1.20 AU/HA.	23
II. Development and fattening	256 Ha.	1.13 AU/Ha.	7
III. Fattening	200 Ha.	1.11 AU/Ha.	1
IV. Milk production for home consumption and feeding	22 Ha.	1.22 AU/Ha.	26
V. Milk production for sale and feeding	47 Ha.	0.85 AU/Ha.	6
VI. Breeding, development and fattening	308 Ha.	1.63 AU/Ha.	7

Smaller farms tend to devote less attention to the improvement of grasslands. Generally 'ratana' and 'natural', that are considered species of lower quality, are found at the smaller farms. The average farm size of farms having 'estrella' is larger (Table 3.3). The difference is significant at a level of 0.1. The difference in farm size between farms having estrella pastures or other improved pastures is not significant (significance level is 0.488). Also, pastures with fruit trees in it are less often found with increasing farm size, while at the same time the presence of timber tends to increase (Table 3.4).

Table 3.3 *Average farm size in relation to grassland type.*

Composition of the pasture	Average Farm Size	Number of observations
Ratana/Natural	49 Ha.	58
Estrella or Estrella/Ratana	167 Ha.	22
Other improved pasture	367 Ha.	9

The larger farms tend to be focused to one or a few specific activities. Activities are generally directed to breeding or fattening of the cattle. Often improved grass species (Estrella, Guinea and Brachiaria species) are encountered. Fruit trees are not found in the pastures, whereas trees for timber are generally present (see Table 3.2, 3.3 and 3.4). Estrella can be found on smaller farms and ratana/natural pastures are also encountered on larger farms, but a trend as discussed is clearly present.

Table 3.4 *Presence of fruit trees or timber in relation to farm size.*

Farm size class	No. of farms with pasture	No. of farms with fruit trees	No. of farms with timber
1. 0 - 29 Ha.	51	33 (65%)	30 (59%)
2. 10 - 60 Ha.	51	33 (65%)	41 (80%)
3. 30 - 260 Ha.	44	21 (48%)	41 (87%)
4. 85 - 1000 Ha.	24	5 (21%)	19 (79%)
5. > 400 Ha.	7	0 (0%)	7 (100%)

LUPs associated with LUZs with predominance of grassland cover.

This section presents data on the distribution of farms in relation to the CLCC in combination with farm size class. Whereas in the former section trends are presented on the basis of data of the individual farms, in this section trends are analyzed at LUZ level. Table 3.5 present the distribution of farming systems for combinations of CLCC and farm size class, only with respect to areas with pasture as dominant land cover. The data indicate:

- First, a clear dominance of farms dedicated to livestock production for all LUZs with dominance of grassland cover. Dominance of cattle farms for areas with dominance of grassland cover was concluded before. Important, in this context, is that the dominance shows for all combinations of CLCC and farm size class involved.
- Secondly, differences in land use pattern exist between LUZs. They seem to be related to the farm size class.

The following observations are made with respect to the area of study:

1. High percentages for cultivated pasture (> 40 %) and low percentages for Wooded Area (< 30 %; class "A" in Table 3.5) only occur in LUZs with large to very large farms (farm size class 4 or 5). This is conform findings presented in the forgoing section. The farms are only dedicated to livestock production. Large herds are kept and the pastures often consist of improved grass species. The presence of WA, that may reach up to 30 %, do not have a specific meaning. Generally 20 to 30 % of the area

is classified as WA, corresponding to wooded river courses, trees along roads and other densely wooded spots not directly relevant with regard to the interpretation of land use.

2. LUZs with very small farms (class 1) and predominant grassland cover do not occur. This seems to confirm findings of Waaijenberg who states that proportion of the farm area dedicated to crop production increases with decreasing farm size. Very small farms dedicated exclusively to livestock or milk production are rare. Consequently, LUZs with very small farms and clear dominance of pasture in the cover composition is not found.
3. Areas with small and medium sized farms (class 2 and 3) tend to have a reduced percentage of cultivated pasture (<35 %) and an increased percentage of WA (25% <WA< 45%), indicated by CLCC "B". This phenomenon is related to the occurrence of farms that are not dedicated exclusively to animal production and with the common occurrence of fruit trees and timber trees in the pastures.
Further interpretation of the CLCC in terms of land use depends on average farm size. LUZs with CLCC "B" and small farms are associated with presence or orchards and pastures with trees[1]. The occurrence of medium sized farms seems to indicate less importance of milk production for home consumption and a use more specific dedicated to stock breeding. The higher farm size class relates to a lower percentage of farms with fruit trees and difference in farm proposition[2]. For LUZs with larger farms in this category, the land use is characterized by grazing and the increased importance of timber production (reforestation). The lower percentage for cultivated grassland for this category indicates the presence of pastures of poor quality[3].
4. The other CLCCs pertaining to the general class of area with dominance of grassland show non-cultivated pasture as dominant grassland type (CLCC "WAPAS3"), or have a part of the area classified as forest ("WAFORPAS1"). Large farms dedicated to livestock production occur, with sometimes a few hectares dedicated to arable cropping. The larger percentage of non cultivated pasture, in case of the WAPAS3, is caused by

---

[1] **Small farms.** In two LUZs with CLCC "B" and FS class "2" thirteen farms were visited. Of the 13 farms 10 were dedicated to livestock production and 3 had both crops (maize, beans and plantain) and cattle (table 3.6). The increased percentage for WA is the reflection of the common presence of trees disperse over the farm area or the presence of orchards. Of the 10 farmers having pasture, 7 mentioned the presence of fruit trees (limon, naranja, aguacate [avocado], coco and guavo) and 8 mentioned having timber (Laurel and Cedro) disperse over the farm area.

[2] **Medium sized farms.** 10 farmers distributed over 4 LUZs were interviewed; 7 farms were located in one LUZ. Of these 7 farms 6 were dedicated to livestock production and one to the production of perennial crops (orchard of fruit trees) with a small part of the farm for maize production. Of the six farms with cattle, 5 had livestock production as the main activity, with in some cases small parts dedicated to the production of cacao or of maize and yucca and one farmer was dedicated to both production of maize and to livestock production for market sale. The data of the three farms located in the other LUZs confirmed the pattern. All the 10 interviewed mentioned timber disperse over the area a minority mentioned also the presence of fruit trees (table 3.7). For the group in total we see a decrease, from 69 to 32 per cent, in farms which mention the presence of fruit trees and an increase, from 67 to 89 %, of the farms which mention the presence of timber wood.

[3] **Larger farms.** Data on six farms were obtained. All were dedicated to livestock production with in some cases very small proportions of the farm area dedicated to the production of maize, plantain or cacao. The relatively high percentage for WA in this class in this case is explained by the presence of trees for timber wood in the fields. In one case 10 hectares was reforested (with Surá [*Terminalia chiriquensis*], Eucalipto and Cedro). Three other farms mention a few hectares under forest cover and sometimes a considerable percentage of tacotal. A further decrease in presence of fruit trees is observed. Four of the interviewed mention the presence of considerable parts of the farm area with 'pará aleman', a grass species which is only found on poorly drained soils.

Table 3.5. *Farming system distribution in relation to CLCC for the area with pasture as dominant land use.*

CLCC[1]	Farm size[2]	No. of LUZs[5]	N[3]	Crop production	Livestock	Mixed farming	Management category
A. PAS1, PASBB1.	4, 5	10	5	0	5 (100%)	0	1,2
B. WAPAS1, PASWA2, MPASWABB	2	4	13	0	10 (77%)	3 (23%)	3
	3	4	10	1 (10%)[4]	8 (80%)	1 (10%)	3
	4, 5	8	6	0	6 (100%)	0	2
C. WAPAS3.	4, 5	3	1	0	1	0	-
C. WAFOR-PAS1.	4, 5	2	2	0	2	0	-
MWAPASSV, MFORWAPAS	4, 5	2	1	0	1	0	-

[1]) The entries in this column denote CLCCs. The names denote the most important cover classes in the composition, in order of importance. PAS stands for pasture; WA for wooded area; BB for bare soil or built-up area; For is forest. The prefix M stands for 'mixed', indicating the absence of a dominant cover class. CLCCs with similar cover composition are grouped. These groups have the following characteristics:
A. Cultivated pasture > 40 % and Wooded Area < 30 %;
B. Cult. pasture < 35 % , WA between 25 and 45 %, Non cultivated pasture < 25 % and BB < 20 %.;
C. With non-cultivated pasture as dominant grassland type or with part of the area classified as forest.
[2]) Farm size class (see Table 3.4).
[3]) Number of observations for each LUZ category.
[4]) Refers to tree crops.
[5]) Denotes the number of LUZs with the combination of CLCC and farm size class as denoted.

very poorly drained often drowned valleys in a hilly landscape.
Another land cover composition results when part of the area is classified as forest. The two farms in the corresponding LUZs for which data were obtained, indeed mention a considerable part (30 and 50 hectares) under forest cover or as reforested area. The occurrence of remnant parts of forest can possibly be explained by poor soil and terrain conditions. (see Part 3 of this thesis).

5. The last category (Table 3.5) indicates LUZs with forest and secondary vegetation as important land cover classes, aside from grassland. These LUZs represent the areas of agricultural penetration. The presence of secondary vegetation suggest that the much of the areas are not used - or only very extensively - for grazing. This can indeed be ascertained when visiting the recently deforested areas.

In table 3.6 the number of farms according to the farm proposition and presence of trees in pastures is listed for combinations of CLCC and farm size class. Only farms with pasture and livestock are considered, but the farms may belong to LUZs with a dominant land cover class other than grassland. Table 3.6 intends to show that the correlation between farm proposition and farm size, as shows from the individual farm data, is also reflected in the LUP of the LUZs, although these represent aggregated units. Totals are presented for LUZs with corresponding farm size class. The totals indicate a decrease in percentage of farms that produce milk, a decrease in percentage of farms mentioning fruit trees and an increase in percentage of farms mentioning timber, with an increase in farm size class of the LUZs. Important is that the patterns are more or less consistent for the LUZs with the same farm size class, irrespective of the CLCC. The size of the populations however is too small to confirm this statistically.

Table 3.6. *Farm proposition and presence of trees disperse over the farm area, in relation to farm size class, regarding all farms with pasture and livestock.*

Composite Land Cover Class[4]	Farm Size Class	Number of farms according to farm proposition[1]				Number of farms mentioning trees in the pastures[2]		
		N[5]	IV	V	I/II/III/VI	N[5]	FRUIT	TIMBER
MBBWAPAS1	1, 2	11	4	1	6	15	8	10
WAPAS1 MWAPASBB	1, 2	5	3	0	2	11	8	8
BBWA1 MBBWAPAS2	2	9	7	1	1	9	8	4
MWAFORBB		8	4	2	2	7	2	5
WAPAS1 MPASWABB	2	5	2	0	3	10	7	8
WA2	1, 2	3	0	0	3	4	3	3
TOTALS		41	20 (49%)	4 (10%)	17 (41%)	49[3]	34 (69%)	33 (67%)
MPASBBWA	3	3	1	0	2	4	2	3
MBBWAPAS	2, 3	4	0	1	3	5	2	4
WAPAS1 MPASWABB	2, 3	6	0	2	4	10	2	10
TOTALS		13	1 (8%)	3 (23%)	9 (69%)	19	6 (32%)	17 (89%)
PAS1	4, 5	1	0	0	1	1	0	1
PASWA2	4, 5	6	0	0	6	6	2	6
WAPAS3	4, 5	2	0	0	2	2	0	2
TOTALS		9	0	0	9 (100%)	9	2 (22%)	9 (100%)

[1] The proposition of activities regarding livestock were only determined for those farms for which data on the composition of the herd was obtained. For the explanation of the roman numbers see table 3.3.
[2] The numbers only refer to farms which reported grasslands as part of the farm area.
[3] Totals exclude farms being part of the area of agricultural penetration.
[4] The CLCCs is given to indicate that results remain the same in spite of changing cover composition. For the explanation the names see the notes for table 3.5.
[5] 'N' indicates the number of observations. That is, the number of farms with pasture and livestock, pertaining to the LUZs with a combination of CLCC and farm size as listed, and for which data on farm proposition respectively presence of trees was obtained.

We see furthermore that small and medium sized farms with livestock are often part of land use zones with a land cover partly consisting of bare soil (indicated by BB in the name of the CLCC). That is, small farms with livestock are often part of LUZs where beside grazing also other use of the land is made, whereas the larger farms with livestock tend to be part of land use zones indicating a land use only for grazing.

In summary of chapter 5, it can be stated that the CLCC in combination with farm size class relates to the land use pattern. This is illustrated for example by high percentages of cultivated grasslands (CLCC "PAS1" and "PASBB1") always corresponding to the presence of large cattle farms. The interpretation of the land cover class 'wooded area' depends on the farm size class of the LUZ. Larger farms tend to possess more timber wood and less fruit trees in the pastures. Activities like reforestation are more feasible for larger farms. These tendencies are reflected in the LUPs of the LUZs.
Deviant land cover class composition (e.g. including secondary vegetation) sometimes seems to be related less to land use characteristics than to soil and terrain characteristics. In the case of the agricultural frontier area they do relate to deviant land use patterns.

6. LAND USE PATTERNS ASSOCIATED WITH A LAND COVER COMBINATION OF WOODED AREA, BARE SOIL AND PASTURE.

General considerations for mapping LUZs with a mixed land cover.

To interpret CLCCs that are a combination of wooded area (WA), bare soil (BB) and pasture (Pas), the farm size class should be taken into account. Smaller farms tend combine crop production and animal husbandry, whereas the larger farms tend to focus on one or a few specific activities. The scale and purpose of the activities differ as well. For the smaller farms, the cultivation of arable crops is generally the most important activity in economic terms. On the basis of this relation between land use and farm size, one would expect the presence of mixed farms for LUZs that have small farms, rather than the presence of farms belonging to various farming systems. The larger farms (with hundreds of hectares) can often be mapped individually. Therefore, LUZs corresponding to larger farms with a LCC combination of WA, BB and PAS are not expected. These expectations are confirmed by the result of the classification (see table 3.7). LUZs related to larger farms (farm size class 4 or 5) and showing a combination of bare soil, wooded area and pasture do not exist! Only for LUZs with small and medium sized farms such cover combinations are found.

The LUZs can further be differentiated on the basis of specific CLCC and farm size class. Especially the percentage of the area classified as bare soil is of importance. CLCC "A" indicates a land cover composed of limited percentages of 'bare soil' (< 20% of the area). CLCC "B" represents areas with 25 to 45 percent 'bare soil'. This latter group can be differentiated further on the basis of the farm size class. In Table 3.7 data for the individual LUZs and group totals are given, in order to see whether the patterns are consistent for LUZs with the same farm size class and corresponding CLCC.

Bare Soil representing less then 20 % of the area: CLCC "A".

One group of LUZs belonging to CLCC "A" is not presented in the table. It represents LUZs with an average field size of less than 2.4 ha. (class 1). These areas are often associated with small villages and their surroundings. Three of the 5 concerning LUZs corresponded to villages known as Rio Jiménez, Los Angeles and Humo. The other two LUZs also had a residential function. The agricultural activity In these LUZs is directed to crop production for home consumption in small fields or garden lots. For one LUZ data was obtained. It indicated the presence of various properties of 1 hectare. The owners were mostly employed at the banana plantation. Generally crops are grown in small quantities (maize, yucca, beans, plantain) and generally tree crops are found (some fruit trees and cacao).

Table 3.7 *Land use patterns associated with a composite land cover of wooded area, bare soil and pasture.*

CLCC[4]	Farm Size Class	LUZ ID.[1]	N[2]	Crop production	Livestock	Mixed Farms	Management class
A. WAPAS1	1,2	94	7	0	1	6	3
MWAPASBB	1,2	99	4	1[3]	0	3	3
WAPASBB	1,2	23	2	1	0	1	3
1. TOTAL			13	2 (15%)	1 (9%)	10 (77%)	
B. BBWA1	1	8	3	1[3]	1	1	3
MBBWAPAS1	1	93	5	0	1	4	3, 4
MBBWAPAS2	1	106	3	0	1	2	3
2. TOTAL			11	1 (9%)	3 (27%)	7 (64%)	
B. MBBWAPAS1	2	10	13	3[3]	8	2	2, 3
3.[5]	2	17	6	3	0	3	2, 3
B. MPASBBWA	3	18, 39	3	0	2	1	2, 3
MBBWAPAS	2,3	36	8	1	5	2	2, 3
4. TOTAL			11	1 (9%)	7 (64%)	3 (27%)	

[1]. Identification of the LUZ for which the presented data is obtained.
[2]. Number of farms visited.
[3]. *Crop production in this case refers to orchard or plantation (fruit tree, coffee or cacao or production of ornamental plants).*
[4]. CLCCs with comparable cover composition are grouped: CLCC "A" represents the following cover composition: 'WA' between 25 and 45%; Cultivated pasture between 10 and 30%; Non cultivated pasture between 10 and 20 % and BB less than 20 %. CLCC "B" represents: 'BB' between 25 and 45 %; WA between 20 and 35 %; Cult. Pas. between 10 and 25 % and non cult. pas. less than 20 %. The numbers indicate combinations of CLCC ('A' or 'B') with farm size class.
[5]. No totals are generated because of the deviating land use pattern for LUZ 10.

The majority of the farms of the LUZs with farm size classes 1 and 2, were dedicated to mixed farming (Table 3.7). Although pastures are mostly found, livestock production is not the main activity[4]. The relatively low 'bare soil' percentage of this LUZ class with farm size class 1 and 2, should not be explained as the result of less area being devoted to annual crop production. But rather as the consequence of less area of the total area for crop production being devoted to the cultivation of maize. This is illustrated in Table 3.8. This table presents data on the area of maize cultivation as percentage of the farm area and of the total area for crop production of the farm. The lower section of the table lists data for LUZs with small

[4] Of the 13 farmers interviewed 10 had arable crops, tree crops and grassland. One farmer was dedicated to livestock production only; One was dedicated to arable cropping only; One farmer was dedicated primarily to the production of cacao and coffee (farming system '4C', see appendix). The farmers involved in arable cropping mentioned the cultivation of cassava, chamol or ñampi in combination with maize mostly in smaller quantities. Only one farmer reported to cultivate maize only.

proportions of 'bare soil' in the cover composition. The area of maize cultivation varies between 17 to 73 percent of the total area destined for arable crop production. This is in contrast to the 90 to 100 per cent of the total arable crop area for the production of maize reported for LUZs with a higher proportion of their area classified as BB (upper two sections of the table).

Table 3.8 *Average plot size, percentages of total arable crop area and percentage of total farm area of the area for cultivation of maize and/or rice in relation to CLCC.*

LUZ No.	Farm Size Class	CLCC	N[1]	Maize (and/or Rice) area		
				Average Plot Size	Avg. Percentage of farm area	Percentage of total crop area
93	1	MBBWAPAS1	4	2.3 Ha.	31 %	90 %
106	1	BBWA1	2	1.5 Ha.	23 %	100 %
10	2	MBBWAPAS1	6	2.8 Ha.	15 %	93 %
17	2	MBBWAPAS1	2	3.0 Ha.	48 %	100 %
36	2/3	MBBWAPAS2	5	10.8 Ha.	21 %	100 %
39	3	MPASBBWA	3	10.7 Ha.	11 %	100 %
94	1/2	WAPAS1	4	1.8 Ha.	20 %	44 %
99	1	MWAPASBB	3	0.7 Ha.	5 %	17 %
23	1/2	WAPASBB	2	8.2 Ha.	39 %	73 %
89	2	WAPAS1	6	2.0 Ha.	9 %	45 %
111	2	PASWA2	1	2.5 Ha.	8 %	50 %

[1] The number of farms for which data on acreage of arable crops were available.

Bare soil covering between 25 and 45 % of the LUZ: CLCC "B".

For the definition of LUPs corresponding to this CLCC further differentiation according to farm size class is required. For LUZs with very small farms the LUP is similar to the forgoing: generally mixed farms are found, with cattle breeding not representing a commercial activity; the relative importance of maize is reflected in the higher percentage of the LUZs classified as bare soil[5].

For the LUZ with small farms one expects mixed farming as the dominant farming system. This is not true for LUZ number 10. Predominant farms dedicated to livestock production were found. This might be the result of a possible error in the aerial photo interpretation. The LUP of LUZ number 17 is agreement with the expected pattern. Mixed farms were found with maize as far most important annual crop. The interviews do indicate a higher degree of specialization in land use. This is reflected for example in the occurrence of farms

[5] Very small farms. Of the 11 farms 7 were mixed farms; three were dedicated to livestock production only and one was dedicated to the production of ornamental crops. 7 of the 11 interviewed reported arable crops. When producing for home consumption mostly maize and yucca were found. When for market selling, maize was the most important crop. 9 of the 11 farms had pasture land with sometimes a very limited amount of cattle (in cases zero or one). 8 of them reported fruit trees, sometimes also timber trees, scattered within the pastures.

dedicated to the production of ornamental crops[6]. For LUZs with even larger farms (the medium sized farms) a lower percentage for 'bare soil' is observed. Less farms are classified as mixed farms, as consequence of a decreasing proportion of the farm area dedicated to arable cropping. However, the total area dedicated for annual crops is not less. Respectively 4,3,8,14 and 6 hectares dedicated to the production of basic grains (maize predominantly, one reported rice) were reported by farmers who had the by far largest part of the farm area covered with grassland. In those cases where crop production was the main activity large areas were reported for the cultivation of maize. This is reflected in the high values for average size of the maize plots (10.8 Ha. respectively 10.7 Ha., Table 3.8). The presence of fruit trees decreases. Activities are found, that are only feasible for larger farms, such as reforestation, attributing to a somewhat increased percentage of WA[7].

7. SUMMARIZING RESULTS AND CONCLUDING REMARKS

1. The land use zone proves to be a useful geographical entity for the description of land use at sub-regional scale. In many cases the corresponding land use pattern refers to only one land utilization type or farming system. This is the case, for example, for banana plantations. However, land use zones composed of many farms and exhibiting a combination of land cover classes a composite land cover demonstrate also a limited number of farming systems. Generally a dominant farming system is observed.

On average, the dominant farming system represents 72 percent of the farms of a LUZ. This figure is based on a total of 109 observations divided over 14 LUZs. The lowest value observed was 64%. Therefore, the dominant farming system is in many cases sufficient to characterize the LUP. It is concluded that the LUZ represents a relevant aggregation level for the inventory of land use and that the land use pattern is a useful concept to describe the land use at sub-regional scale.

2. The composite land cover class in combination with the farm size class relates rather well to the observed land use pattern. Only in one of the 14 LUZs investigated did the LUP not correspond to the pattern that one would be expected on the basis of CLCC in combination

---

[6] Small farms. Of LUZ no. 10 only 3 of the 14 interviewed could be classified as mixed farms. 8 were classified as cattle farms. Observing that the areas classified as bare soil corresponded to two larger, regular shaped areas, it might indicate the presence of a larger commercial farm which was erroneously included in the unit under consideration. However, the assumption that with increasing farm size the activities are more commercially oriented and directed to specific activities is confirmed. When cattle farming is concerned this is reflected in the decreasing number of farms which report fruit trees in the pasture lands. Only 7 of the 12 farms with pasture reported the presence of fruit trees. However, orchards are observed more often. One farm had an orchard of fruit trees, one had a cacao plot and one farm produced ornamental crops.

Of the six farms visited in LUZ no. 17 one farm was purely dedicated to the cultivation of maize and two farm produced flowers and ornamental crops. The tendency to commercialization of the production is further confirmed by farmers which report a few hectares with orchards in stead of fruit trees disperse of the farm area.

[7] Medium sized farms. Of the 11 farms 3 were classified as mixed farms, 7 as cattle farms and 1 farm as dedicated to arable cropping. Only 4 of the 9 farmers with grassland (for one farm no data on presence of trees was obtained) reported fruit trees. Presence of trees disperse in the fields for wood production is frequently reported (7 farmers). Two farmers mentioned reforestation. One farmer reported 300 trees of cedro amargo, in another case 1500 eucalyptus and teak trees were planted and an additional 120 laurel trees and cedro amargo were found.

with the farm size class. This means that in 93% of the cases (referring to LUZs) the LUP is in agreement with the CLCC and FS class. It indicates that the CLCC and farm size class serve to map LUPs.

The error of 7 % might partly be explained by misinterpretation of the aerial photos or might be due to changes in land use between time at which the photos were taken (1984) and the time of recording of the satellite data (1986).

The results are based on the 14 LUZs most complex in terms of land cover class composition and land use pattern. The classification of most other LUZs is less complex. For example, all banana plantations were recognized and mapped correctly by the LUZs. The same score was obtained for the forest areas and other areas of natural and semi-natural vegetation.

The viability of the LUZ approach was tested for the GRS area. The questions whether the LUZs express clearly distinct land use patterns and whether these land use patterns can be mapped on the basis of land cover and field size inventory are answered affirmatively. De idea behind the land use zone approach for mapping of land use is, that the mapping rules are defined on the basis of data obtained for a representative part of the area to be mapped, and that these rules are subsequently applied to map the total area. The validity of the mapping rules (relating CLCC in combination with field size class to land use pattern), should still be tested on an independent data set.

The rules for mapping, investigated in this chapter, are valid only for the region concerned. The mapping rules should be adapted for application to other regions, in Costa Rica or elsewhere.

CHAPTER 4

DETECTING CHANGE IN LAND USE THROUGH CHANGE IN THE LAND-COVER COMPOSITION OF LUZS

INTRODUCTION

In chapter 2 we saw that LUZs represent permanent spatial structures if they represent agricultural land. In this chapter we discuss whether changes in LUZ aggregation structure (field composition) and thematic content (land-cover composition) provide information on land-use change. We shall focus on change in land-cover composition, because changes in aggregation structure generally involve changes in land cover.

Changes in crop or vegetative cover can best be observed with satellite imagery (Molenaar and Janssen, 1991). Accordingly, we used Landsat-TM imagery of 1986 and 1990 to determine the land-cover composition of LUZs. The results apply to the Guacimo-Rio Jiménez-Siquirres study area, an area of about 30 x 30 kilometres, representative of the northern part of the Atlantic Zone (AZ) of Costa Rica. These results were not evaluated statistically because there were no systematically obtained data available for either of the periods corresponding to the Landsat images. Field observations were made from 1988 to 1990. Information on land use obtained in this period were used to evaluate the results. In addition, we present secondary evidence of land-use change to evaluate results.

The recent history of land use in the AZ has been characterized by intensiveness and radical change. This is manifest in the colonization of new areas and an associated high deforestation rate. But also changes in existing agricultural areas occur, caused by change in socio-economic conditions and initiated by agrarian policies. For example, a new policy was announced a few years ago. Called "agriculture of change", it entailed a reduction in the subsidies for basic grains (maize and rice) and the stimulation of alternative crops for export (Waaijenberg, 1990). Other important factors for change are the fluctuations in market and export prices. For example, meat prices dropped (with serious repercussions for the local economy, as much of the AZ is used for meat production) and cassava prices increased. The search for new cash crops has led to the introduction of palm heart (pejibaye), macadamia nuts, and ornamental trees, to name a few.

METHODS AND PROCEDURES

Jensen (1983) mentions two methods of using remote sensing to detect change: image differencing and post-classification comparison. Image differencing uses the difference in radiance values to detect change in images from different dates. The disadvantages of this method are:

- The difficulty in locating the threshold between radiance change and no change;
- The occurrence of radiance change as a result of other phenomena than change in land cover. The method is not specific.

Variations of the image-differencing method use least-square transformation or principal components to reduce adverse affects from differences in atmospheric conditions or sun angles (Fung and LeDrew, 1987). Post-classification comparison is hampered by the often high classification errors (Jensen, 1983). Both methods imply a per pixel comparison of the scenes. As a consequence, spatial misregistration is an additional cause of errors in change detection.

A partial solution to these problems is to reduce the number of change categories. This can be done by looking at change within a restricted area (see Vogelmann, 1988, who uses image-differencing to detect change in forests, or Christensen *et al.*, 1988, who use post-classification to detect change in wetland). It can also be done by generalizing the change categories. At the highest level of generalization, we register only whether change has occurred or not; the nature of change is not specified.

To detect change in the AZ, we used the post-classification method. It does not require us to evaluate differences in atmospheric condition for the two scenes. Any change we detect is immediately expressed in terms of the LCCs. We can counteract the effect of high classification errors and improve classification performance by generalizing the LCCs and incorporating the context information from the LUZs. And we can counteract the effect of spatial misregistration by using data aggregation, i.e. using spatial units (LUZs) whose sizes are large in comparison to the registration error.

We used two Landsat-TM images: one from 6 February 1986 and the other from 9 February 1990. We followed the steps listed below to detect change:

- Classify the land cover;
- Generalize the land-cover classes to a level where they are common to both scenes;
- Aggregate from pixel level to LUZ level;
- Interpret changes in composite land cover in terms of changes in land use.

The procedure is illustrated in Figure 4.1. The steps are explained in the following paragraphs.

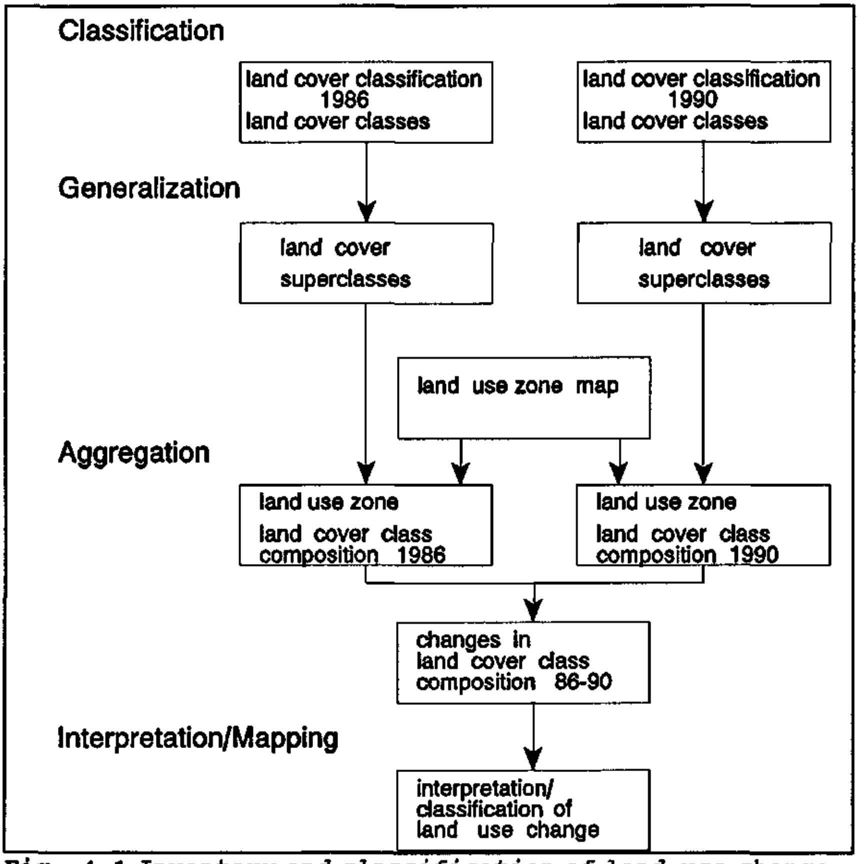


Fig. 4.1 *Inventory and classification of land-use change.*

Independent Classification of Both TM Images and the Need for Generalized Cover Classes

A land-cover classification of both TM images was obtained. The 1990 classification defined the LCCs differently from the 1986 classification. This difference was caused partly by the changes in land cover that had occurred in the intervening period, requiring the definition of new classes. But the main reason was that it was possible to gather field data for the classification of the 1990 image, while this was not possible for the 1986 land cover classification. This made it possible to relate spectral characteristics more precisely to land-cover characteristics, allowing for a more specific definition of the 1990 LCCs. Comparison of the two periods required us to define common land-cover classes. We did this by adopting a higher generalization level in the classification system for both images, at which super-classes are defined common to both systems of classification. It implied recoding of the picture elements of both classifications.

For example, various spectral classes in the 1990 classification were defined as different types of woody vegetation, all corresponding to plantation crops, which allowed us to define a "woody plantation" superclass. And as this superclass could also be defined for the 1986 classification, though the classes (at the lower level of generalization) had a content different from the 1990 classes. We could compare both images at the superclass level. The same principle of generalization was applied to other LCCs. Although the classes defining the lowest hierarchy level were often not the same, it is possible to define common superclasses.

Aggregating From Pixel Level to LUZ Level

The two scenes were compared not on a per pixel basis, but per LUZ. There were two reasons for this:

- The positional accuracy of pixels is too low. Misregistration introduces an artificial change in class values when classifications from two periods are compared. The relative error decreases as the aggregation level increases;
- The pixel level is too detailed for an inventory of land-use change at sub-regional level. Information on land cover change on sub-field level is not relevant when change in area of crop cultivation is required.

So pixel information is aggregated to information about the land-cover composition of the LUZs and then LUZs are compared with respect to their land-cover composition.

LUZs correspond either to a part of a farm with specific land use, or to a whole farm, or to an area with various farms but with corresponding land use. In general, we can say that:

- Specific changes will be restricted to the area defined by the LUZ;
- Changes will occur over the total area of the LUZ.

LUZ boundaries are therefore considered to provide the relevant geographical basis for evaluating land-use change.

Interpreting Change in Land-Cover Composition

To infer information on change in land-use at sub-regional scale from change in land-cover requires us:

- To interpret the land cover classes (LCCs) in terms of land use;

- To interpret the change in the composition of LCCs in terms of change in land-use pattern (LUP).

LCCs usually represent general types of cover like "bare soil" or "wooded area", which have to be translated into land-use classes. The translation depends on the context, which is provided by the LUZ. For example, "wooded area" can be interpreted as cacao when it is found within a LUZ belonging to the class of agricultural area, with small farms, and when a large part of the LUZ is classified as wooded area. (A percentage of 10 to 20 percent, and even 30 percent of "wooded area" can often be explained by the presence of wooded river banks, homesteads, and other wooded areas. Higher percentages are needed to denote the presence of tree crops.)
"Bare soil" can indicate various specific cover types or uses, such as maize-growing area, forest clearings or residential area. The interpretation depends on the form and size of the 'bare soil' area and on the land cover class of the neighbouring areas.

It is important to realize that changes in land-cover composition often take place gradually and that, within a LUZ, the stages of change will often vary. Different LCCs can occur simultaneously as a consequence of the same process of change. For example, the planting of a macadamia plantation is done in phases. It will take some years before the nut trees are fully established and producing. In the meantime, the areas corresponding to the different ages of the trees will be mapped as a wide range of LCCs, varying from degraded grasslands and secondary vegetation to woody plantation. But all the changes observed within this specific context are part of the same process. Therefore, changes in land cover in a specific location cannot be considered independently from changes in the neighbouring areas (information on the context is required for the proper interpretation). Because land use is related to farm size, also land-use change will be related to farm size. For example, because of the large investment required, macadamia cultivation is feasible only on large farms or plantations. Therefore, a strong increase in "wooded area" indicates an increase in macadamia cultivation only if the farm-size class indicates large farms.

Thus, the LUZs provide a context not only for interpreting LCCs, but also for interpreting a change in land cover. This is because we can define only a limited number of possible changes for each LUZ, dependent on the farm-size class and the existing type of land utilization.

We conclude that there has been a change in land cover or land use only if there is a marked difference in the land-cover compositions between the 1986 and the 1990 images. Changes in the percentage of the LCCs are interpreted taking the 1986 LUZ classification into account.

Marked Differences in Land-Cover Composition

The aggregation of pixels into LUZs, implies the aggregation of the per pixel land-cover class to LUZ land-cover percentage. At LUZ level, the LCCs with their percentages of occurrence, together determining the composite land cover (CLC). A threshold value is required to indicate change or no change in the land cover percentage, when LUZs from different periods are compared. We conclude that there has been a change in the land-cover composition of a LUZ when the difference in percentage of any LCC exceeds 15 per cent. A difference of less than 15 percent is not considered significant. Of this 15 percent error margin, 10 percent is explained by inaccuracy of the land cover classification (see Chapter 5, Part 2, for explanation on how this value is obtained), and 5 percent is explained by

geometric inaccuracies.
The land cover composition is determined by overlaying the LUZ map with respectively the 1986 and 1990 land cover classification. Difference in land cover composition might occur as consequence of difference in the geometry between both land cover classifications. This 5 percent is an average value. The difference in land cover percentage as consequence of geometric inaccuracy can vary from 0 to 12 per cent, depending on the size of the LUZ and the land cover percentage. Smaller units and higher LCC percentages mean higher possible errors in establishing change in land cover percentage. The maximum error was calculated assuming a maximum relative geometric error of 60 metres (2 pixels), considering the minimum size of one square kilometer for the LUZ, and departing from a land cover percentage of 40. (The areas of the major LCCs such as bare soil, wooded area, and pasture often cover 40 percent of a LUZ.) A threshold value of 15 per cent seems acceptable, considering that slight changes in the LCC percentage would not really alter its proportionality within a LUZ. Also, 15 per cent corresponds to the width of many composite land cover classes, i.e. the range in percentage of the LCCs defining a composite Land cover class. See Chapter 5, Part 2 for explanation of the reliability in determining land cover composition and for the definition of the composite land cover classes.

LAND USE CHANGE IN THE GUACIMO-RIO JIMENEZ SIQUIRRES AREA FROM 1986 TO 1990.

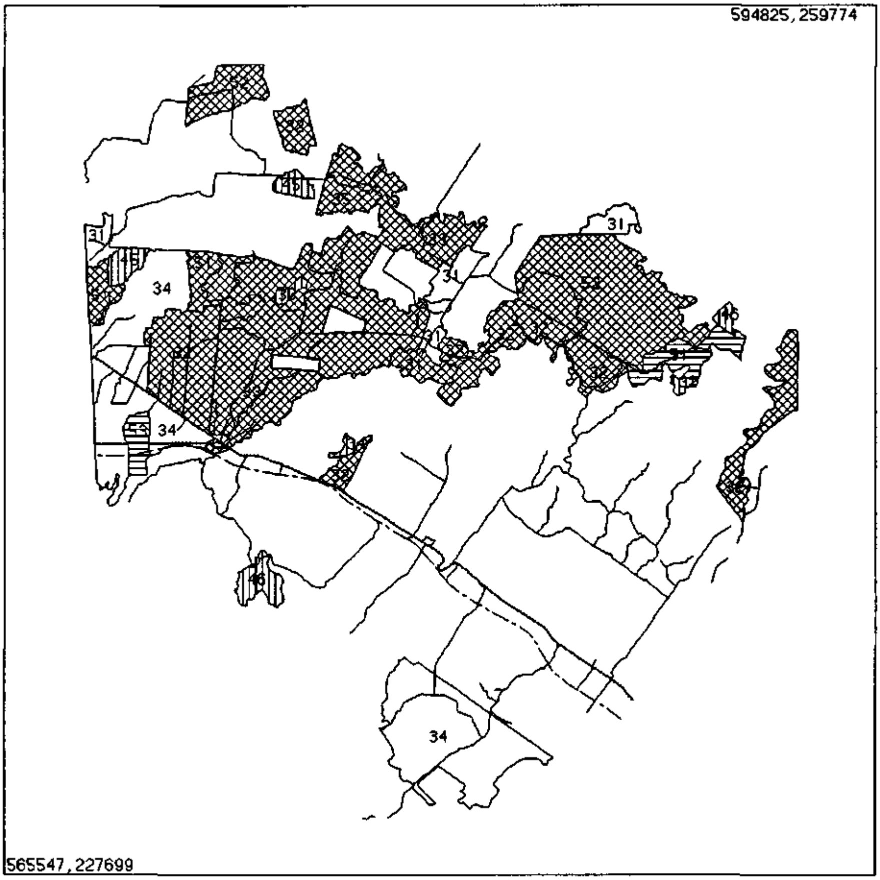
Applying a threshold value of 15 per cent results in 96 of 136 LUZs showing a marked difference in their land-cover composition. Of the original 157 LUZs, 21 were excluded, either because they were either too small or because clouds covered a too large part of the LUZ, preventing the determination of the land cover composition.
On the basis of the changes in the land cover characteristics the following dominant processes of change in land use in the GRS area are identified:

- Decrease in the area for the cultivation of maize;
- Decrease in and degradation of the grassland area;
- Increase in plantation area (banana, macadamia, other);
- Deforestation;

Decrease in the Area for Cultivation of Maize

Change in the area for maize-growing is concluded from change in the area classified as bare soil. At time of scene recording (February 6, 1986 and February 9, 1990) fields are being prepared for the sowing of maize or maize had been sowed shortly before. Therefore, maize areas show as bare soil. However, this interpretation is only valid in the context of LUZs with very small to small farms and with a mixed land cover with bare soil as dominant land cover class. For other areas, changes in the percentage of bare soil should be interpreted differently.

The LUZs with very small farms and a mixed land cover (indicated as class 31 in Figure 4.2) did not show a reduction in the areas of bare soil. Parts of these LUZs are residential, so some of the area classified as bare soil may correspond to buildings, houses, and roads.



LUZ's with small to medium sized farms and mixed land cover.
Marked changes in the 'Bare Soil' area from 1986 to 1990.

- Area with small farms and mixed land cover, marked decrease in the percentage of 'bare soil'. Interpreted as strong decrease in area of maize cultivation from 1986 to 1990.
- Areas with very high percentage of 'bare soil' in 1986 and strong decrease in 'bare soil' area in 1990. 'Bare soil' in 1986 is a result of clearing (e.g. deforestation). Current land use varies.
- LUZ with medium sized farms, mixed land cover dominated by 'wooded area' and 'bare soil'. Strong decrease of 'bare soil'. The 'bare soil' 1986 corresponded only partly to area of maize cultivation. Partly it corresponded to areas being cleared for other purposes.
- No evidence of land-use change.

Fig. 4.2 *LUZ's showing a decrease in the area of maize cultivation between 1986 and 1990.*

No evidence was found of a reduction of the maize-growing area in these LUZs. This is not surprising if we consider that most of the maize grown is for home consumption, and so less sensitive to price fluctuations.

The LUZs with small farms and a mixed land cover (classes 32, 33 and 52) all showed a marked decrease in the amount of bare soil, which was interpreted to mean a decrease of the maize-growing area (Fig. 4.2). In these areas root and tuber crops (cassava, chamol [*Colocasia esculenta var. antiquorum*] and other), palm heart (*Bactris gasipaes*), and ornamental crops, to name a few, have gained in importance. This is expressed primarily as an increase in the area classified as plantation.

There were no data available on the actual decrease in the area dedicated to the production of maize for the GRS study area or for the specific LUZs, nor were there data available on maize production. Nevertheless, the conclusion that the maize area has decreased considerably is supported by the sharp drop in maize prices from 1986 onwards (Table 4.1). The price of yellow maize in December 1989 was 40 per cent lower than the 1985 price. The price of white maize, which is most grown in the AZ, was in decline until 1988, after which it recovered somewhat. But even with this increase, the price in May 1990 was still 28 per cent lower than the September 1985 price.

Table 4.1. *Maize prices per metric ton from 1985 to 1990 (in U.S.$)*

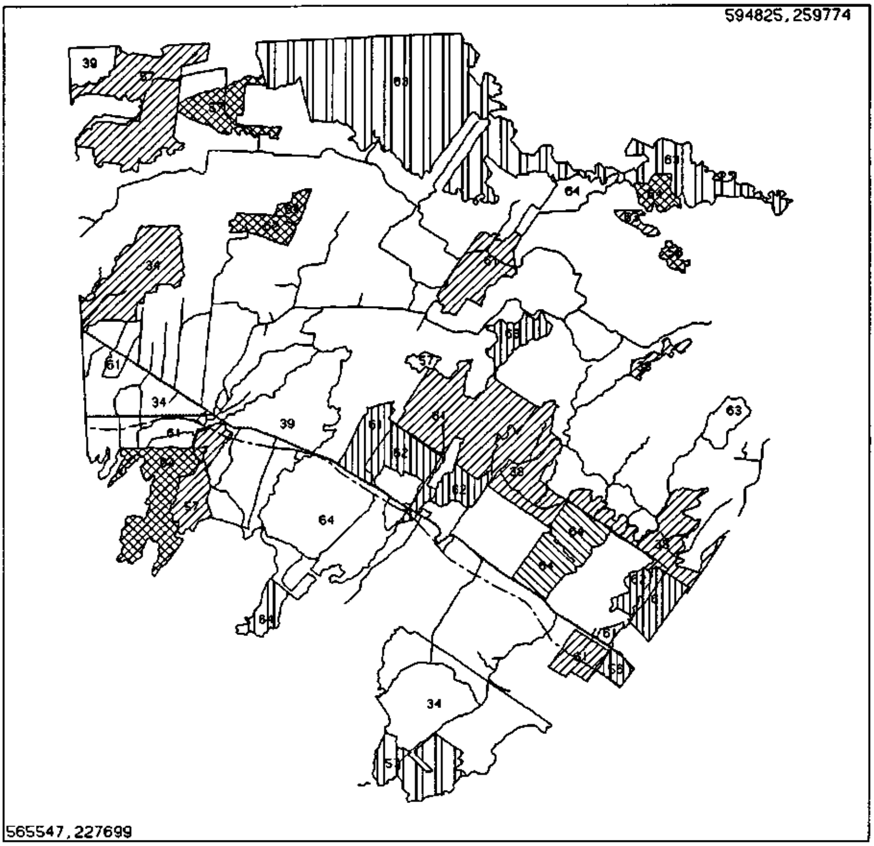
Yellow Maize		White Maize	
Date	Price	Date	Price
Sept. '85	$ 270.90	Sept. '85	$ 270.90
Dec. '86	234.30	Dec. '86	234.30
Dec. '87	201.00	Dec. '87	201.00
Dec. '88	173.00	Dec. '88	173.00
Dec. '89	163.60	Sept. '89	189.39
		May '90	195.43

Source: Department of Economic Studies, CNP. Prices are based on the official exchange rate set by the National Bank of Costa Rica.

Three LUZs with small to very small mixed farms (class value 34, Fig. 4.2) did not show a significant reduction in the bare soil percentage. All three, however, had smaller percentages of bare soil in 1986. Annual crops are cultivated in these areas, but maize is relatively unimportant. For a description of this class I refer to the preceding chapter.

Decrease in and Degradation of Grassland Areas.

The changes in land use with respect to area used for grazing are mapped in Figure 4.3. To evaluate changes with respect to the areas used for grazing one has to take account of the farm size class. For some LUZs the data on the cover composition suggested a complete change in land use (LUZs shaded with a checkered pattern in Figure 4.3). These LUZs corresponded to single farms. Grasslands had made way for uses such as banana plantation, plantations of palm heart, ornamental crops, reforestation, and cassava (the new land use is indicated in Figure 4.4). These changes were confirmed by field observations. The area of change they represented, however, was but a small percentage of the total.



Land-use change in the Guacimo-Rio Jiménez-Siquirres area: 1986-1990.
Change in areas of Livestock Production.

- LUZ's with a very strong decrease in cultivated grassland. The areas were used for grazing in 1986. In 1990 the land was used for other purposes (e.g. banana plantation, reforestation).
- LUZ's with a marked decrease in cultivated grassland. Parts of the area, formerly used for grazing, now used for other purposes (e.g. perennial crops).
- Decrease in cultivated grasslands, but with an increase in uncultivated grasslands or secundary vegetation. Grassland degradation occurring because of low stocking rates or negligence.
- Decrease in uncultivated grasslands, but with an increase in the cultivated grasslands. Upgrading of the grasslands because of improved management.
- LUZ's with no marked changes in land cover composition. No evidence of change in land use.
- LUZ's for which no data are available because of cloud cover.

Fig. 4.3 *LUZ's showing a decrease in the grassland area or grassland degradation between 1986 and 1990.*

It is worth noting that many LUZs showed a decrease in the percentage of (cultivated) grasslands and an increase in the proportion of grasslands classified as secondary grasslands (grasslands invaded by herb- and scrub vegetation, as result of poor grazing and management). These changes could not be verified because of lacking field data for 1986. For a number of LUZs the decrease in area of grasslands is confirmed by observations in the field. The decrease was in one case due to expansion of a neighbouring banana plantation. In another cases, grasslands were converted to area for the cultivation of cassava and to area planted with achiote (tree crop producing dye). The changes in grassland area seem to bear out the story of sharply decreased livestock production that is told by the figures in Table 4.2. Table 4.2 presents figures on the national production of cattle. The decline in the production is assumed to have occurred also in the AZ. The data on land cover change suggest that the decrease in production is reflected partly in the decrease of the area for grazing and partly in the degradation of the grasslands as result of more extensive use of the land. Only in one case, there was a clear improvement in the state of the grasslands. Verification in the field revealed that this area corresponded to a farm whose management had changed around 1986. In none of the LUZs was there an increase in the area of grasslands.

Table 4.2. *Total number of cattle produced for the domestic market and for export.*

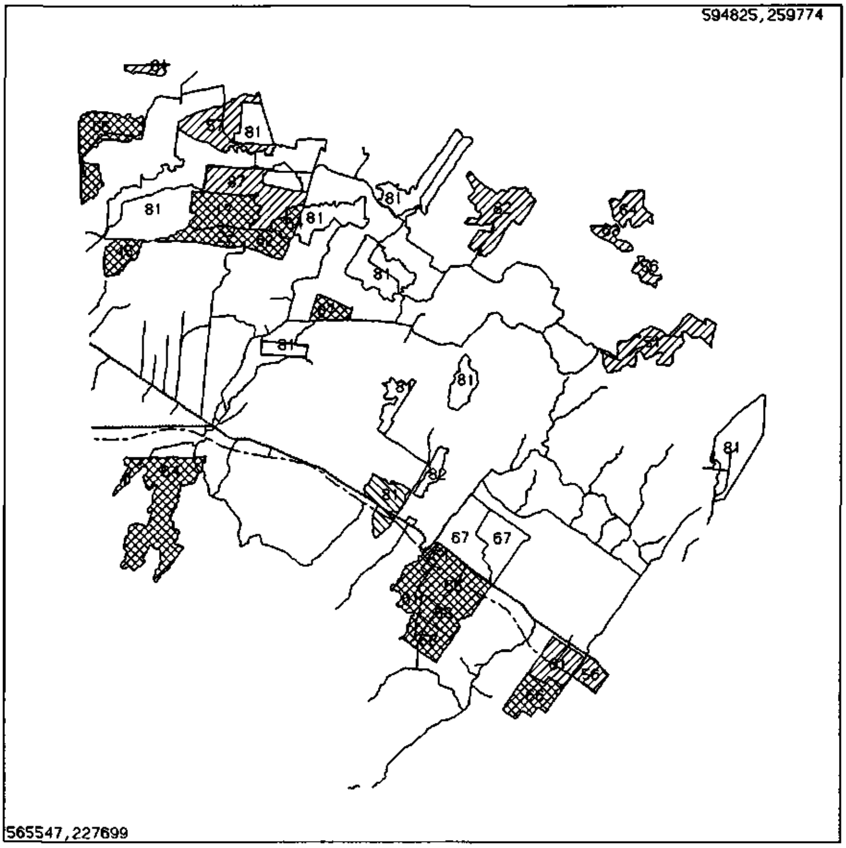
Year	Total domestic consumption	Total export
1986	355.594	206.544
1987	331.858	156.698
1988	283.814	137.268
1989	270.689	117.561
1990	300.562[1]	112.487

[1] The increase in the number of cattle for domestic consumption is primarily the result of an increase in the number of calves born. Source: Departamento de Estudios Económicos, Consejo Nacional de Producción, Departamento Pecuário.

Increase in Plantation area

It is fairly easy to see where new banana plantations have been established and where old plantations have been removed. One LUZ was classified as a new banana plantation even though the greater part of it was "bare soil". This plantation had been established only a short time before the date of the 1990 image, and so large parts of it were either covered with recently-planted material or had no cover at all. We could also see where existing banana plantations had been extended or improved (Class 8, Fig. 4.4). This was the case in two LUZs. One LUZ corresponds to the finca Santa María. The markedly higher percentage of the area classified as banana for this finca agrees with the plantation's improved productivity: from 1494 boxes/ha/year for 1986, which is far below the average yields, to 1935 boxes/ha/year in 1988 (Source: ASBANA, Costa Rica). The changes were the result of improved management.

New plantations, other than banana plantations, are also fairly easy to detect, if they occupy large areas. Figure 4.4. shows an increase particularly in the macadamia plantations



Land-use change from 1986 to 1990. Large commercial enterprises other than cattle ranches and plantations.

1. Marked decrease in 'bare soil' and incases grassland. Marked increase in 'tree crop plantation' and 'secondary vegetation'. Present use is macadamia plantation.
2. Marked decrease in 'bare soil' and increase in 'plantation crops' (not very pronounced). 'Bare soil' in 1986 corresponds to clearing; Former land use unknown; current land use for (semi-)perennial crops.
3. Marked decrease in grassland cover and increase in 'woody plantation'. Former pasture land now being used for reforestation.
4. Decrease in pasture, increase in 'plantation', 'banana' or 'secondary vegetation'. Land use has changed to the cultivation of perennial crops (macadamia, palm heart) or root and tuber crops (cassava, chamol).
5. Medium sized farms. Decrease in 'wooded area', increase in 'plantation crops'. Land use has change to the cultivation of palm heart.
6. Decrease in forest cover, increase in 'plantation crops' and 'banana'. Area partly converted to plantation (banana or other).
7. New banana plantation, established between 1986 and 1990.
8. Banana plantations with an increase in the banana cover, indicating increased density of the vegetative cover or extension of the planted area.
9. Former banana plantation. No longer in existence in 1990.
10. No evidence of change in land use.

Fig. 4.4 *Increase of banana- and macadamia plantations and area of reforestation.*

(first and fourth class of change). Also change as consequence of reforestation is evident (third class). The superclasses, used to compare the LUZ land cover composition, do not provide information on the specific plantation crop (e.g. cassava, palm heart, macadamia nuts, and ornamental crops, eucalyptus, etc.). More detailed information was obtained from the 1990 classification, but also the 1990 LCCs do not indicate a specific crop. Nevertheless, with some additional information (e.g. plantation size and elevation, terrain and soil), we were able to draw some conclusions on land-use change (Figure 4.4). In areas with small to medium sized farms, detecting an increase in the area of specific crops is possible only through more detailed study of the LUZ concerned, or through direct observations in the field.

Deforestation

In many cases, the percentage of forest area had decreased considerably, but this was often compensated by an increase in the percentage of secondary vegetation and woody plantation. Different training samples have been used to determine the spectral signatures of forest cover, secondary vegetation and woody plantation for the classification of the two different periods. The content of the land cover classes of both classifications will, therefore, not be the same and will probably overlap. Accordingly, a reduction in forest cover, if so compensated by secondary vegetation or woody plantation, was not considered evidence of change in land cover or in land use. A reduction in forest cover was accepted only if the difference, after subtracting the increase in secondary forest and woody plantation, was more than 10 percent. LUZs with no difference after subtraction were classified as areas with no evidence for deforestation. Data on the most interesting area (the colonization area) could not be obtained because of a too high cloud cover.

A few LUZs corresponding to isolated spot of (remnant) forest areas had been completely deforested and banana plantations were established on these sites. Some LUZs, partly with forest area and partly agricultural area, showed a decrease in forest cover. Others showed no such decrease.

CONCLUSIONS

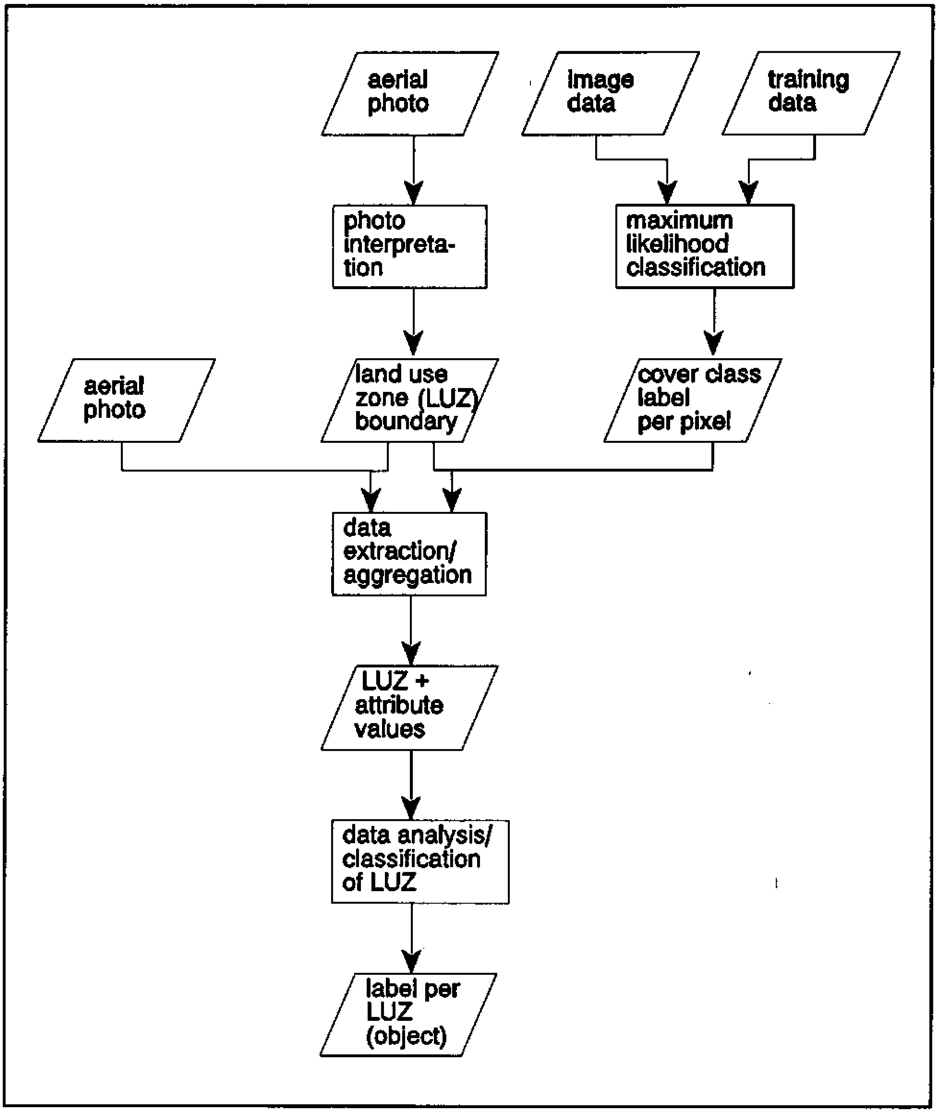
Land use change can be detected adequately by determining change in land-cover class composition of the LUZs. Evaluating change in LCC composition will in many cases provide better results than detecting change in land use than evaluating land cover on a per pixel basis only. This is illustrated, for example, by the detection of the new macadamia plantations, which show a range in land cover classes due to the difference in age of the trees in the different parts of the plantation. Without the context provided by the LUZ the content of the change is difficult to establish.

Areas with change in land use, known from field observation, all appeared as areas of change using the procedure described. In some LUZs, we observed a partial change in land cover. These cases require more detailed study to determine where exactly the change has occurred and what the corresponding change in land use is. The same satellite imagery can be used for these studies at a more detailed level.

PART II.

USING REMOTE SENSING AND AERIAL PHOTOGRAPHY IN AN OBJECT ORIENTED APPROACH TO THE INVENTORY OF LAND COVER AND LAND USE.

1. INTRODUCTORY CHAPTER



1. INTRODUCTION

Part One of this dissertation describes the land use and land use change in terms of land use zones and their characteristics. This second part devotes attention to methodological aspects. Consecutive chapters reflect the information theoretical analyses of the inventory process. These chapters specify the methods applied in the land cover and land use inventory in the Atlantic Zone of Costa Rica.

Our aim is to develop tools for the inventory and description of land use at sub-regional scale. Aside from the farm survey few tools are currently available for the inventory of land use. Aerial photo's and, more recently satellite imagery have been used for the inventory of land cover in particular.

From the beginning of this century, aerial photos have been used for the inventory of land cover, land use and agricultural geography (Schneider, 1974; Bomberger and Dill, 1960). However, since the advent of the earth observation satellite for natural resource studies (the launching of the ERTS satellite in 1972), the use of satellite data has become more important and is now widely applied.

Both techniques have advantages and drawbacks. These are related to their spectral and spatial resolution. The low spectral resolution of the aerial photos limits their use for the discrimination of land cover. More discriminative power is obtained with the second generation Landsat satellites, for instance. This is due to the higher spectral resolution (measurements in seven spectral bands). However, the Landsat imagery shows a relatively low spatial resolution (30 by 30 meters).

On the other hand, the high spatial resolution of the aerial photographs makes them especially suitable for the discrimination of spatial features, such as texture and structure. There is an additional advantage to the use of satellite imagery: the data is presented in digital form. This allows a more quantitative approach and formal description of data structures and processes. Yet, the improved scanning techniques also make digital processing of aerial photographs possible.

2. APPROACHES IN THE USE OF SATELLITE IMAGERY FOR LAND COVER AND LAND USE INVENTORY

Satellite imagery has been widely applied for the mapping of land cover and land use. Different techniques are available for the classification of the spectral data (Ahmad, 1986). Nevertheless, the generally level of accuracy does not meet the requirements. Accuracies range between 40 and 90 percent (Ioka, 1986). Overall performance generally lies in the range of 70 to 75 percent (Kenk, 1988; Janssen, 1990). In chapter Three and the Appendix, we consider ways to improve the classification performance. These possibilities include improved procedures, especially stepwise refinement of the training statistics and incorporating procedures for checking of results. Or they may require change of classification strategy, which involves applying spectral indices for the characterization of cover types.

Other techniques to improve classification performance use ancillary data (see Hutchinson, 1982; Kenk *et al.*, 1988). These techniques entail:

- The use of ancillary data prior to the classification. This may be done through stratification or segmentation of the image (Hutchinson, 1982; Cross and Mason, 1985);

- The modification of the classifier by using a priori probabilities. These are based either on the estimated composition of the known object classes or on the known association between object classes and ancillary data (Swain, 1978; Strahler, 1980; Hutchinson, 1982). For example, Janssen (1990) uses information on past land cover to classify current land cover. Gong and Howarth (1990) add structural information to the feature space upon which the classification is executed;
- Post-classification approaches (or sorting) use ancillary data after the multi-spectral classification to improve the results. The use of a majority filter is a simple example (Gurney, 1983). Digital elevation and forest cover have also been used as ancillary data (Kenk, 1988). And elevation, slope and aspect (Wilkinson and Mégier, 1990) have been used as criteria in a decision model for better classification performance.

Figure 1.1 displays the flow charts, corresponding to the three approaches mentioned above. Of those strategies in particular the latter has proven to be successful. All three approaches are pixel-based classifications.

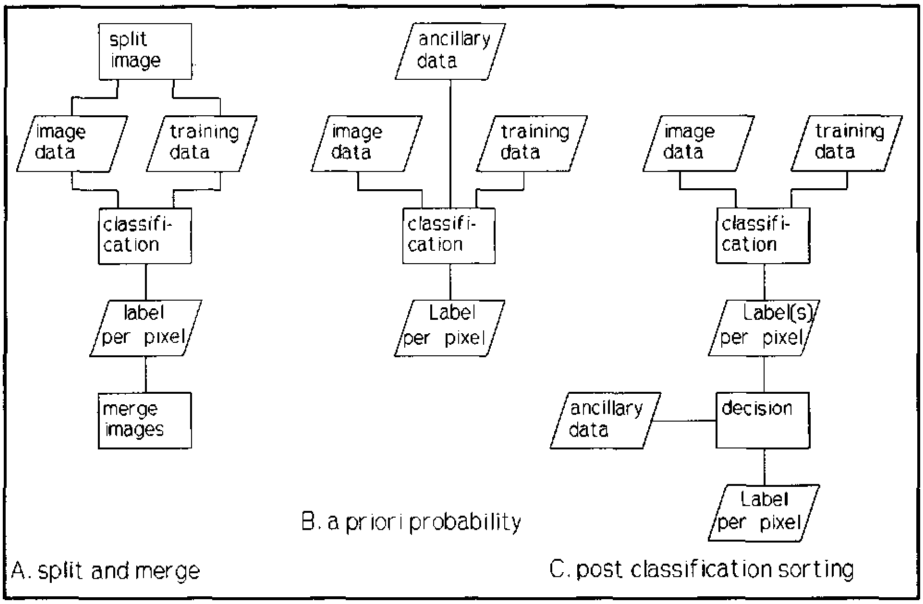


Fig. 1.1 *Schematic representation of the classification approaches.*

3. OBJECT-ORIENTED APPROACH

More recently the object-oriented approach to land cover and land use classification has attracted more attention. This has been stimulated by developments in the field of GIS and expert systems. Janssen *et al.* (1990) and Janssen and Van Amsterdam (1991) correct misclassifications by providing a spatial context. This is the geometry of the object, being defined as an area in which only one land cover type is expected (i.e. agricultural fields, in their case).

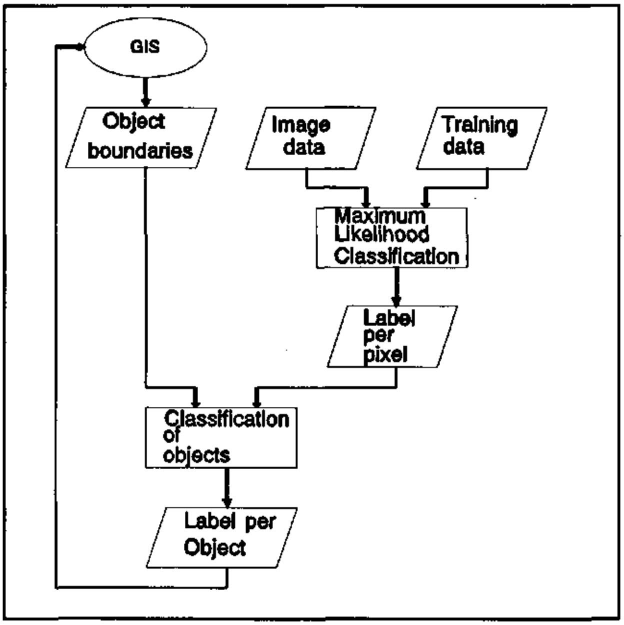


Fig. 1.2 *Flowchart for the object classification approach.*

The corresponding flowchart for this object classification approach is given in Figure 1.2 (Janssen and Van Amsterdam, 1991). In these cases, the per pixel classification is used to obtain information on the objects concerned. Such an approach, for example, has been used in obtaining statistical information on land cover and land use per region or strata. See Bunschoten (1989) for the application of remote sensing at the Central Bureau for Statistics, see Husson (1989) for a related project concerning the agricultural statistics for the European Community. Refer to Hall-Könyves (1990) for crop area estimation in harvest regions or, for example, the use of satellite imagery to obtain statistical data on crop condition per crop-reporting districts (Manore and Brown, 1990). In most of these cases, information on the districts or region is obtained from topographic maps.

The land cover and land use inventory of the Atlantic Zone of Costa Rica adopts an object oriented approach. This choice was based on the consideration that the pixel level did not correspond to the level of detail of the required land use information. Therefore, spatial objects were defined at sub-regional scale: the land use zones. The land use zones approach also incorporates aspects of the three classification approaches.
The goals for the object definition were twofold:

- The object should provide a context (spatial as well as thematic) for the interpretation of land cover and land use. This could include scene stratification for the reduction of the variation in land cover. In this manner cover types that cannot be distinguished spectrally can be separated (Hutchinson, 1982). Or the objects should provide ancillary data (as for example on spatial characteristics) for the classification of land cover and land use. (See Ton *et al.*, 1991, who uses spatial knowledge for the segmentation of Landsat images.)

- The objects themselves should be entities meaningful to the presentation of land use information.

4. DATA ACQUISITION AND OBJECT CLASSIFICATION

Three phases can be distinguished in the information process:

1. Definition phase: definition and description of the terrain objects and their characteristics;
2. Data acquisition phase: how to obtain data on object characteristics;
3. Data processing and information extraction phase: the classification of terrain objects and mapping.

The various aspects of this process will be briefly introduced in the following sections, The descriptions will refer to the land use inventory as carried out in the Atlantic zone of Costa Rica.

4.1. Object definition and identification

The first step in the information process is the definition of the relevant object types. Sometimes the object types are clearly defined and individual objects can easily be identified. In other situations, only the context might be clearly defined, with the object types only conceptually defined. The definition and identification of the objects then forms an integral part of the information process. Soil or land use inventory might serve as an example of a situation in which the objects are not clearly defined in advance. In that event, the identification of the objects is an explicit task in the inventory process. The spatial objects then might be identified through the interpretation of aerial photos (e.g. delineation of soil mapping units). In Costa Rica, aerial photo interpretation served to delineate (i.e. identify) Land Use Zones. These are the objects of concern for the mapping of land use at sub-regional scale. Satellite imagery was used for the recognition and mapping of land cover (by means of land cover classification).

Object recognition requires the image characteristics of the objects to be known. By way of those characteristics the object can be recognized in the image (satellite or aerial photo). The sequence is as follows:

1. Definition of the objects;
2. Determination of the image characteristics of objects;
3. Recognition of the objects.

Objects are defined in terms of their characteristics and the object relationships. Objects are characterized by (Molenaar, 1991a, 1991b):

- Geometrical structure;
- Thematic content;
- Dynamic behavior.

Relationships between objects are defined, apart from the topological relationships, through:

- Classification hierarchies;
- Aggregation hierarchies;
- Associations between objects.

The data structure for the description of terrain objects is presented in Chapter One, Part One. The remaining chapters of Part One deal with the description of the Land Use Zones.

The subject of this second part is the description of the information process. The next phase in the information process concerns the data acquisition to determine the object characteristics and object relationships.

4.2. Data acquisition for object characterization

The second phase refers to data acquisition. It defines which characteristics can be measured and how this is done.
To obtain data on object characteristics, different sources are generally available. Data might be obtained through direct measurements in the field, by the use of aerial photos or satellite imagery or from different sources of textual and tabular data.

The use of aerial photos or satellite imagery may serve two purposes:

1. Identification and recognition of objects, which was the subject of the former section, and;
2. Data extraction for the inventory of the objects' characteristics.

The second objective requires the object to be identified in the image. Then the image characteristics of the objects can be defined through data extraction. The sequence is:

1. Identification of objects in the image;
2. Data extraction;
3. Description of object (image) characteristics.

In both processes, knowledge is required on how object characteristics are expressed in the imagery used. In the first process (object recognition) knowledge is necessary to determine the relevant image characteristics for recognition. In the second process it is needed to infer the relevant information from the measured image characteristics.
An example is the land cover classification. The spectral characteristics of a land cover class can be determined. On that basis, the land cover class can be recognized and mapped. In the other situation an object is identified in the image. By analysis of the spectral characteristics conclusions may be drawn regarding its land cover characteristics.

Landsat-TM provides land cover data, while traditional aerial photo interpretation techniques are better suited to provide information on land use activities (Lindgren, 1985). Land use, which indicates the functionality of the land and how it is used, cannot be deduced directly from land cover data. (See also Bunschoten (1989), concerning the use of remote sensing as a tool to obtain data agricultural statistics).
Land use expresses itself in land cover as well as in the spatial structure, as determined by the size, shape and orientation of the fields. Therefore, the inventory included both land cover and spatial characteristics. In the present study, satellite imagery data (Landsat-TM) were used for the inventory of land cover. For the inventory of spatial elements aerial photos were used.

The data extraction process, as described above, often involves data aggregation. The

Land Use Zones define the object boundaries mentioned in Figure 1.2. All the picture elements belonging to the object are then used to describe the object.
The land use zones represent composite objects, being composed of many pixels. The land cover characteristics of these zones are described in terms of composite land cover. The data is obtained by aggregation of the pixels and their corresponding land cover class.

$$\textbf{pixel+land cover class} \xrightarrow{\textbf{aggregation}} \textbf{land use zone+composite land cover}$$

If the land use zones exhibit a characteristic composition of land cover, then they define a context for the per pixel land cover classification. Accordingly only specific land cover classes are expected within the context defined by the land use zone.

$$\textbf{LUZ+characteristic composition} \xrightarrow{\textbf{context}} \textbf{per pixel land cover classification}$$

This introduces a feedback mechanism for the classification of land cover. This feed back mechanism is used in the present study to improve results. It explains the dual purpose of defining land use zones (Section 3 of this chapter). The land use zones serve as a context for the land cover classification. They also comprise an entity for mapping of land use at the sub-regional scale.

The use of satellite imagery and aerial photos provides specific characteristics of the land use zones (e.g. land cover composition and field size). These characteristics might be used to deduce information on land use. However, they do not provide data for the description of the land use itself. To obtain information on the existing types of land utilization and land use patterns, additional data sources are required. In the case of Costa Rica farm surveys and literature sources were used.

4.3 Data analysis and processing

Data analysis and processing refers to the classification of objects. The classification entails:

- Class definition; and
- Assignment of a class label by means of a decision (or classification) rule.

With respect to the class definition, it is useful to distinguish between data classes and information categories. They have a different function. Furthermore the rules for class label assignment are specific for both types of classes. Two phases are distinguished in the classification process:

1. Data analysis, data class definition and assignment of a data class label to objects of concern; and
2. Definition of information categories and mapping of data classes to information categories.

Data classes refer to the attribute values of objects, for which they are also termed attribute value classes (see Figure 1.3). The classification can be expressed mathematically as:

$$'x \in S'$$

The spatial object ('x') belongs to ('$\in$', is assigned) class 'S'. This formula is taken from Klir and Folger (1988), who use it for the modeling of uncertainty.
The object description ('x') is discussed in a previous section. Class definition corresponds to a process of defining clusters of observations in an n-dimensional feature space, whereby 'n' refers to the number of features considered relevant to the classification. Spectral classes might serve as an example of data classes.

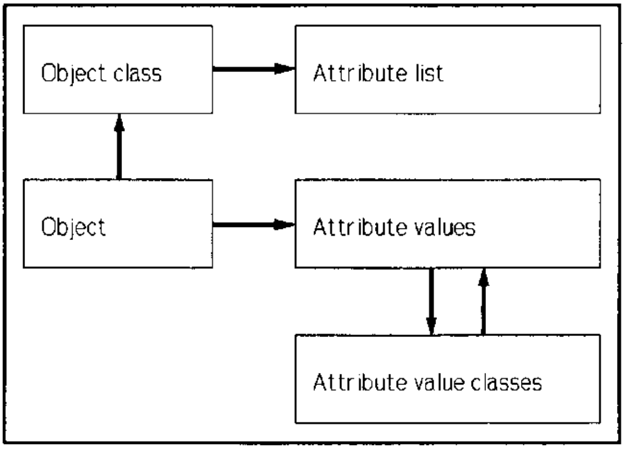


Fig. 1.3 *Class structure of objects and attributes.*

As regards classification, two approaches exist: supervised and unsupervised. In the supervised approach, the data classes are defined on the basis of the statistics of a population specified by the user. In the unsupervised approach the user specifies homogeneity criteria and a cluster algorithm to define the clusters representing data classes.

The advantage of the first approach is that it ensures that the data classes represent relevant information. Though with a possible trade-off in classification accuracy. The advantage of the unsupervised approach is that it ensures certain minimum accuracy requirements for the assignment of data class labels. The discriminative power of the features used in the classification process can be accounted for in the specification of parameters which control the cluster procedure. This is done by specifying minimum requirements for size of the classes and distance between groups (i.e. data classes). The minimum required size and distance are determined on the basis of the reliability of the input data (i.e. the values of the relevant object 'x' attributes). The relevance of classes in terms of their informational value is determined afterwards.

In the case of composite objects, the object characteristics can be described in statistical terms (e.g. mean, standard deviation) because one object refers to many observations. This facilitates a statistical evaluation of the differences between the objects and the use of probabilistic criteria in the definition of the data classes (see Lee *et al.*, 1987).

The terms 'supervised' and 'unsupervised' approach are used in the field of remote sensing. Here the unsupervised approach will be referred to as the data-driven approach. This approach was adopted for the classification of the land use zones.

The classification rules specify the criteria and methods for class label assignment (for example the 'maximum likelihood classification rule'). Data classes, often referring to the specific techniques and methods applied, do not have informational value. (This is already discussed above.) It requires the translation of the data classes to information categories. This process will be referred to as the mapping of data classes onto information categories. The term 'mapping' is used in a mathematical sense, referring to the function of assigning a unique element of a set 'B' to an element of a set 'A', represented as (Lipschutz, 1968):

$$f\colon A \dashrightarrow B$$

In the context of the present study, set 'A' refers to the set of data classes and set 'B' refers to the set of information categories. Aronoff (1984) used the term 'labelling' for this process.

The information categories have direct relevance with respect to the output of information. They represent classes with the appropriate level of aggregation and generalization and with an information content directly relevant for mapping. A land cover class might illustrate the concept of an information category. (The reader is referred to Lee *et al.* (1987) for further discussion on data classes and information categories.)

The decision rules for mapping often take the shape of conditional assignment of class labels to objects. It is especially in this mapping procedure that context-dependent information is used.

6. LAND USE CLASSIFICATION IN THE ATLANTIC ZONE OF COSTA RICA

The next chapters are introduced by a short summary of the procedure used. The complete process of land use inventory and mapping in the Atlantic Zone of Costa Rica is illustrated in figure 1.4. It is in fact a further elaboration of the flowchart for object classification presented in Figure 1.2.

6.1 Photo interpretation for the delineation of land use zones.

The first steps involve the definition of the object boundaries. This is done by means of aerial photo interpretation. The intendion is not to provide information on the thematic content of the objects, as that is considered a separate phase in the inventory process.
Given the sub-regional level of the land use inventory, photo features should be selected that reflect the composite nature of the objects to be defined. The spatial pattern, which refers to the arrangement of image elements, is such a feature. The spatial pattern served as the main key for the interpretation of the aerial photo's. The photo interpretation for delineation of land use zones is the subject of Chapter 2.

6.2 Land cover classification

The land cover composition is one of the land use zone characteristics. Its determination requires the inventory and mapping of land cover, for which satellite imagery are used. Given the information theoretical context of this research, attention is devoted to procedural and strategic aspects rather than technical ones. Standard procedures (where possible quantitative) are needed to implement the inventory process in a system environment. The land cover data obtained pertains to a specific point in time. However, both land cover and land use change in the course of time. Also in this respect uniform procedures are important. It ensures consistent classification of scenes recorded on different dates. In this manner the results will be comparable.

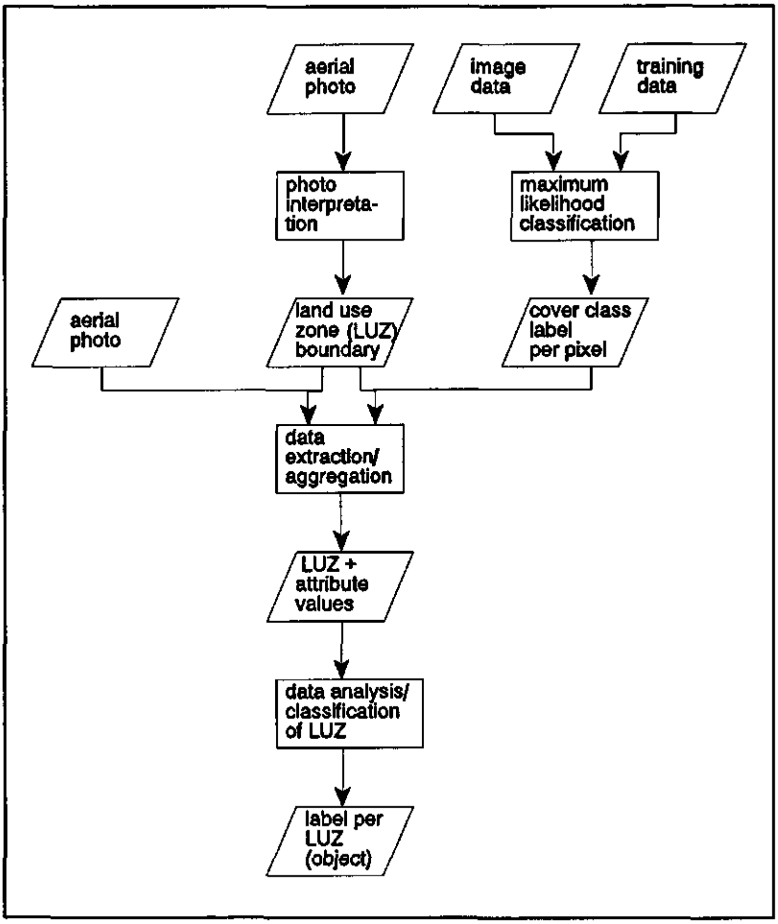


Fig. 1.4 *The land use classification procedure.*

Chapter 3 treats the land cover classification. Attention is devoted especially to the definition of the land cover classes and their associated spectral classes. A procedure is given which ensures the set of spectral classes to be exhaustive, spectral separable and to represent meaningful land cover categories.

Appendix A investigates an alternative strategy to the land cover classification. It makes use of spectral indices to spectral characterize the land cover classes. The indices are related to measurable land cover characteristics and provide, as such, better insight in the spectral complexity of the scene.

Appendix B reviews the a-priori selection of the appropriate spectral feature combination. The purposes of this combination is to obtain colour composites with high informational value as an aid the selection of sites for ground survey.

6.3 Data analysis and definition of (composite) data classes

The land use zones are characterized by land cover composition and by characteristics derived from aerial photo (like the size of fields, referred to as faces in the interpretation process). The data is obtained through the overlay procedure depicted in Figure 1.5. Subsequently attribute value classes are defined, taking account of the distribution of attribute values per composite object. The attribute values refer to the elementary objects. In defining the classes for classification of the LUZ's, homogeneity requirements need to be fulfilled, as in the definition of the spectral classes for the per pixel classification.

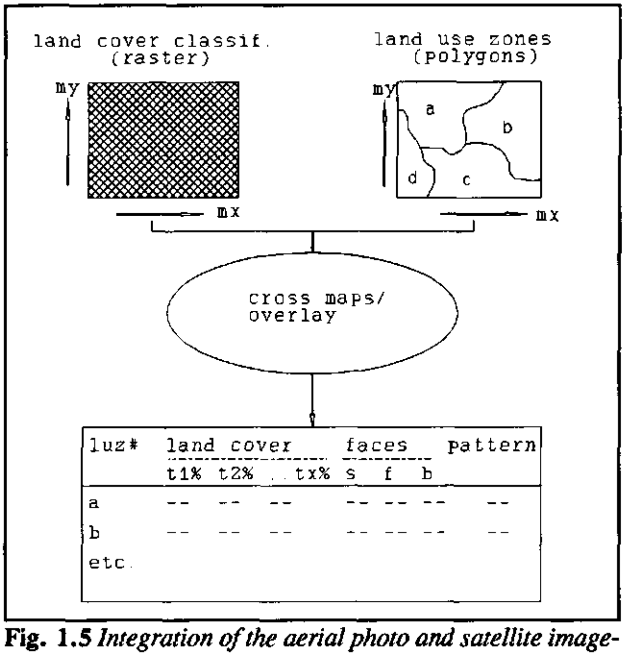


Fig. 1.5 *Integration of the aerial photo and satellite imagery derived products to obtain data on the LUZ.*

Chapter four describes this process for the situation in which the observations concern a continuous variable, namely field size. Data classes (field size classes) are defined, on the basis of the results of variance analysis and these are subsequently mapped to farm size classes by way of correlation.

Chapter 5 refers to a situation where the observations concern a variable with a nominal value, i.e. the land cover class. For each land use zone, a distribution of land cover classes is obtained. In this case, clustering techniques were used to define the data classes, i.e. the composite land cover classes. Two factors influence the uncertainty or possible error in determining the land cover composition. These are the accuracy with which a pixel is assigned a land cover class label (thematic accuracy) and the accuracy with which a pixel is assigned to a specific land use zone (geometric accuracy). Both are evaluated and accounted for in the definition of the composite land cover classes.

Once the data classes are defined, the land use zones are attributed a data class label on the basis of its attribute value.

6.4 Classification system and assignment of class labels to land use zones

The classification of the land use zones represents a multi-source classification problem. This has been subject of many studies on the integration of remote sensing and GIS (Foody, 1988; Goodenough, 1987; Lee et al. 1987; Srinivasan and Richards, 1990; Wu *et al.*, 1988).

Chapter 6 describes the classification rules used to assign a land use class label to each zone. This classification involves mapping of the objects with their associated attribute values and data class values to the information categories. The information categories describe land use in terms of land use patterns.

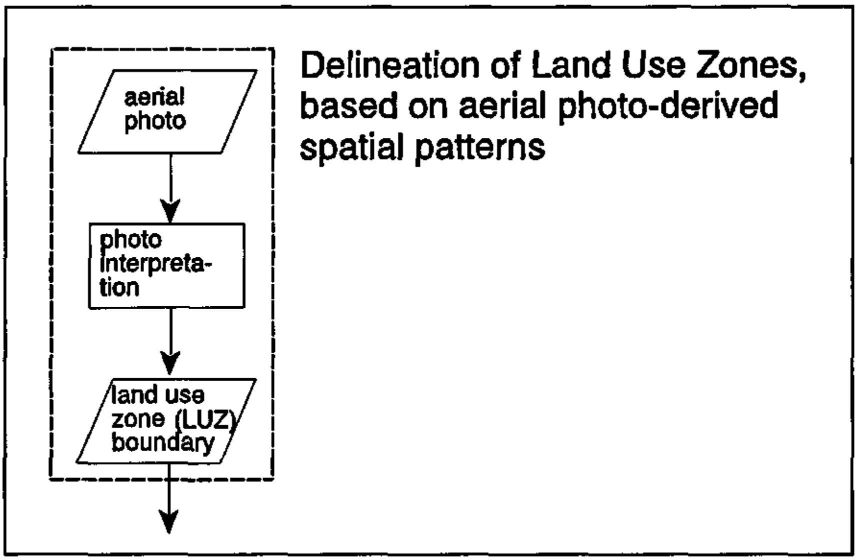
The information categories are hierarchically ordered in classes and superclasses. This generates a classification system similar to, for example, the USGS land cover and land use classification system (Anderson, 1976) or the CORINE land cover classification system (CORINE, 1989). The hierarchical ordering of categories is governed by a decision tree, representing a hierarchy of the classification rules. The decision tree evaluates attributes at different levels. The rules take the shape of the conditional assignment of class labels to objects. These rules are context dependent.

Examples of decision trees can be found in many taxonomic systems. For instance, in the 'soil taxonomy' classification system, specific soil attributes are evaluated at each level in the classification hierarchy. This may range from occurrence of diagnostic horizons at order level to specific physical and chemical soil properties in soil series (USDA, 1975).

In remote sensing many authors have recognized the power of hierarchical classifiers or decision tree classifiers (see Swain and Hauska, 1977; Ton *et al.*, 1991). Especially in the classification of multi-source remote sensing and geographical data, the decision tree classifier is very efficient (Srinivasan and Richards, 1990; Wilkinson and Mégier, 1990).

CHAPTER 2

AERIAL PHOTO INTERPRETATION FOR THE DELINEATION OF LAND USE ZONES



1. INTRODUCTION

With respect to the land use inventory in the Atlantic Zone of Costa Rica, the interpretation of aerial photos (color infrared, scale 1:80000, 1984) aims at the delineation of land use zones. The land use zone (LUZ) is defined as a geographic area that exhibits a characteristic land use pattern (LUP). The LUP may comprise a range of agricultural uses and management. The idea of regionalization of land use based on aerial photo characteristics is not new. Examples can be found in the disciplines of land use and agricultural geography. Kelly used aerial photographs to delineate 'utilization strata' and to select sampling areas as part of an annual agricultural survey programme (see Nunnally, 1974; Thaman, 1974). Another example is given by Murphy *et al.* (see Myers, 1983). They define 'agrophysical units' as geographic areas having definable and comparable agronomic and physical parameters which reflect a range in agricultural use and management.
A stratification making explicit use of the photo characteristics is presented by MacPhail (see Peplies, 1976). He defines 'photomorphic units' as a complex of a variety of tones. These are the result of vegetation, crop type, soil moisture condition, field pattern, and settlement morphology. The photomorphic units are determined on textual and tonal properties.

As mentioned in the introductory chapter, the interpretation process is difficult to structure and formalize, because it integrates various objectives. A solution is sought in the identification of the different tasks and their speparate execution. In the next section, the process of photo interpretation process is described in general terms, subsequently the implementation of the aerial photo interpretation for the delineation of land use zones is explained.

2. THE INTERPRETATION PROCEDURE

Aerial photo interpretation aims at the recognition of spatially defined objects (in this case LUZs). Ideally the process should consist of the following steps:

- Object definition, whereby the objects are defined which are to be recognized;
- Definition of the object characteristics;
- Determination of the image characteristics of the objects and specification of interpretation keys;
- Object recognition or interpretation.

These steps are schematically represented in Figure 2.1A. These are typical for the satellite imagery-based land cover classification, discussed in the next chapter. The land cover types are defined on the basis of field checks. Their associated spectral characteristics are determined. These are an expression of general characteristics such as the presence of green vegetation, bare soil, or water bodies. At the same time, they express specific characteristics of vegetation cover types, including biomass, height and geometry of the vegetation. The land cover types are recognized upon the per pixel classification of the satellite image. -

The aerial photo interpretation is a related process in which the order of steps is normally different. The process is executed in the order shown in Figure 2.1B. The 'objects' are mostly not known in the beginning. Instead a standard set of interpretation keys are used, depending on the purpose of the interpretation. From the many image features (tone, texture, etc.) the interpreter has the task to distill mutually discernable spatial 'objects'. This presumes knowledge on the possible occurrence of object classes given a certain context.

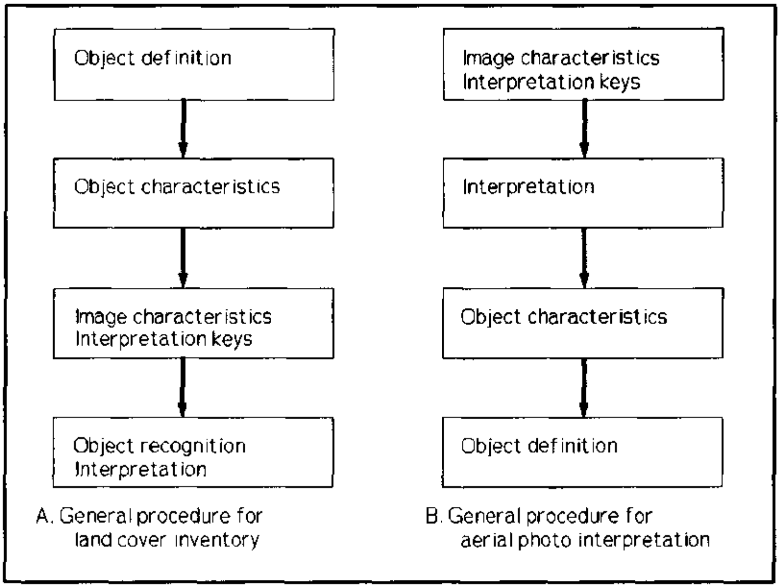


Fig 2.1 *The photo interpreation process.*

It also requires knowledge on the image representation of the possible 'objects'. Some knowledge of the region, often obtained through field trips, is required to tell which image features are to be given special attention.

Photo interpretation is an informal process. The knowledge referred to above, is often informally incorporated. The delineation of the interpretation unit is an interactive associative process in which object definition and recognition are intertwined. It is an iterative process of disaggregation in which a dynamic adaption of the classification system takes place. The following dangers are involved:

1. Object definition may be confused with object classification.
2. The use of specific information may be overlooked.
3. Confusion of aggregation levels, since the photo scale corresponds to another level of aggregation than the individual observation in the field.

For example, in aerial photo interpretation for soil inventory, first the interpretation is carried out and then data is gathered in the field. At that point ideas are formed about the possible soil classes. The photo interpretation is subsequently adapted to the findings in the field. As a consequence of the photo interpretation sequence, attention is focused on the final step. That generally consists of classification of the objects based on the results of the field survey. Since the 'objects' are not yet defined, a mingling of the object definition and object classification is the result. That is, 'objects' are often defined in terms of their classification. The object definition and classification process are not recognized as independent processes.

This is confirmed by the observation that the interpretation process itself is not usually described. Therefore, the relation between the image characteristics and the objects are not made explicit. Also some confusion exists with respect to terminology. In soil survey the mapping unit is often referred to as a class, as it occurs in the legend of the map. Yet it can also refer to the single spatial unit on the map (i.e. the object with its specific characteristics).

The present study takes a different approach. Aerial photo interpretation is described as an independent process. We indicate which image characteristics are used to delineate the units and their relation to the objects identified. In this manner, we are obliged to take account of the scale dependencies, as well as how object identification depends on the type of material used. The description of the interpretation process serves three purposes:

- It allows for an evaluation of the interpretation results and describes the methodology for future interpretations;
- It identifies specific information present in the aerial photos that can be used for further processing and classification of the data;
- It describes the regional context, since the object definition, the object characteristics, as well as their image characteristics (the interpretation keys) are context dependent.

The interdependence of thematic and geometric aspects of objects makes it difficult to structure the process of interpretation and classification. With respect to the land use inventory in the Atlantic Zone of Costa Rica, this problem was tackled by treating the photo interpretation and classification as distinct processes. In other words, the spatial units are defined through photo interpretation. Thematic characterization and classification of the units is done afterwards using other techniques. The outcome is then used to evaluate the results of aerial photo interpretation.

3. LAND USE ZONES AND THEIR PHOTO CHARACTERISTICS

The uses of the land depend upon land ownership (Lindgren, 1985). Parcels would therefore represent useful spatial entities for the description of land use. Yet, this cadastral information was not available for the area under study. However, for larger farms (a few hundred hectares or more), the property boundaries can often be detected by their rectilinear form. Thus, some information on land ownership is provided. With respect to the recognition of smaller farms it is assumed that these are spatialy grouped, allowing us to delineate regions with farms of a certain size. This is very clear in the case of settlement schemes, which exhibiting a uniform farm size distribution.

Agricultural fields have been used effectively as objects for the classification of land use on the basis of spectral information (Janssen and Van Amsterdam, 1991). However, it would provide too much detail for the land use mapping of the Atlantic Zone on sub-regional scale. Furthermore, fields could often not be discerned on the photos, due to quality and scale of the aerial photos (1:80,000). This does not preclude the use of field type information for the interpretation of aerial photos (e.g. in the sense of field pattern).

A field pattern is associated with the type of land utilization. For instance, the cultivation of ornamental crops corresponds to small fields, often arranged in a regular pattern. A grassland area is visualized as surfaces corresponding in tone and texture, delimited by clear linear features that are dark in tone. The lineaments represent living fences as boundaries of the fields. Therefore, the spatial pattern, as determined by the spatial arrangement of farms, fields and crops within fields, reflects the land ownership pattern as well as land utilization. Photo archeology recognizes spatial pattern as an important characteristic and key to the interpretation of aerial photos. Examples are known of ancient patterns of land division that have been preserved for centuries (like the Roman and Japanese patterns that could still be recognized; see Schneider, 1974). Other examples are known where spatial patterns are analyzed, however mostly in relation to terrain features. Schneider (1974) describes field patterns ('flurvormen') in relation to mountainous areas.

Spatial characteristics also reflects the history of land use. For instance, the limits of a former banana plantations are often still visible. Also former settlement schemes are still recognized by their spatial pattern. Areas with a shorter history in agricultural use may be recognized by the absence of living fences.

Historical data can be highly relevant in studying current land use and evaluating development options. Vrana (1989) states that the inability to access historical states of land use may restrict analysis to a snapshot view. He favors the use of historical data as an explicit component of land information systems.

The single-date satellite scene represents a snapshot view. The interpretation of aerial photos on the basis of spatial pattern might provide relevant geographical units for incorporating historical information (see Chapter 2 and 4 of Part 1).

4. PHOTO INTERPRETATION KEYS

4.1 Spatial patterns

A spatial pattern can be defined as a function of photo pattern elements. A pattern element can be defined as surface artefacts that are homogeneous in tone and texture having a certain size and shape. They appear as 'faces' on the aerial photograph. They are delimited by line features that may correspond to elements like fences, roads or tracks or they may be recognised through change in tone or texture, corresponding to change in land cover.

The pattern elements represent too small an area to serve as a basis for mapping. Given the sub-regional scale of the aerial photos (1:80,000) the units of interpretation will represent aggregated areas for which the (spatial) pattern provides the pre-eminent key for the interpretation.

The spatial pattern observed on the aerial photos is considered to correspond to the arrangement of terrain objects, such as agricultural fields. In the manual of photographic interpretation (A.S.P., 1960), pattern is defined as the spatial arrangement of farms, fields, and crops within the fields. The pattern defined as a function of terrain objects only partly corresponds with the pattern defined as a function of photo objects.

In Costa Rica, large tracts of uncultivated and partly cultivated land exist (agricultural frontier areas). Patterns can also be described for these areas. The pattern elements in the latter case consist of forested and deforested (mostly grassland) areas, which appear as patches within the forest area (see Fig. 2.2) or elongated areas (corridors along rivers and streams). The rivers provided access to virgin lands. In the deforested patches of land, no field boundaries could be detected, nor do the limits of the areas correspond to property boundaries. In natural vegetated areas patterns can also be distinguished (as defined by spatial photo elements that are homogeneous in tone and texture). However, in this case the boundaries are often curved and represent gradual transitions, reflecting vegetation patterns determined by the soil and terrain conditions.

Patterns indicate man-made landscapes, when sharply defined, mostly straight linear features are encountered. Non-linear features and gradual boundaries indicate areas with natural vegetation. Within the man-made landscape, a distinction between farms, fields, or crops within fields can often not be made. This is because the boundaries of these object types have a similar appearance on the aerial photo. However, when larger areas are delimited by long rectilinear features, they are interpreted as property boundaries.

4.2 The land use zone boundaries

Another important criterion does not concern the content of the zone but refers to the boundary of the land use zone. Property limits tend to be straight lines that often do not correspond to terrain features. The larger properties are easily recognized by these linear features; they often extend over hundreds of meters, sometimes even kilometers. In these cases, the boundaries of the zone are determined by these straight lines even though within the area sometimes large differences in tone and pattern exist.

Property boundaries generally represent stationary features (see Chapter 2 of Part 1). The structure indicated on the old cadastral maps at the beginning of colonization of the Atlantic Zone early in this century is still reflected in the present spatial pattern, as observed on the aerial photographs.

The recognition of different patterns within a unit of landownership will results in the delineation of a number of LUZs, that are then associated by common ownership. It is important to register this type of association; shifts in land use are often related to change in ownership. Information on this type of object association can often not be derived from a single aerial photo. Analysis of historical material (aerial photos) in particular may reveal the association of land use zones and farm properties. The situation is similar to that of the parcels in multiple use, as described by Vrana (1989) in reference to urban use. He uses historical information to indicate change with respect to urban use of the parcels. Also, land use zone boundaries often correspond to physical barriers, such as rivers, escarpments, etc. They represent stable features. On the other hand, some boundaries correspond to changes in cover (e.g. from forest to grassland) along a curved line but not to any physical change. Such boundaries tend to be less stable in character. The type of boundary is therefore an important key to interpretation.

Our investigation revealed that the aerial photos perform poorly in regard to the recognition of vegetation types and crops. Only major cover types like forest, plantation crops and grass- and arable cropland could be recognized with some accuracy. Therefore, we did not attempt to identify vegetation and crop type by means of aerial photo interpretation. Land cover was determined later, by interpreting satellite imagery. An exception was made for the identification of forest cover, because the presence of forest cover is an important characteristic of particular spatial patterns and herewith for the delineation and definition of LUZs.

Summarizing, the criteria for the interpretation of aerial photos to delineate LUZs land use zones are presented:

Concerning the land use zones:
- Type of boundary of the unit;
- Presence of faces;
- Pattern
- Infrastructural characteristics with respect to the accessibility of the LUZs.

Concerning the faces, the following criteria were considered:
- Boundary type of the faces and level of discernment;
- Form of the faces;

- Tone and texture;
- Presence of trees;
- Size of the faces.

The photo characteristics are listed for each land use zone, in order to provide criteria for its classification. See Chapter Six for the relevant features.

5. SPATIAL PATTERNS IN THE ATLANTIC ZONE OF COSTA RICA.

A set of characteristic patterns for a region can useually be defined and subsequently used for the identification of a LUZ. To describe all patterns occurring in the Atlantic Zone would not be very useful. However, some characteristic patterns are presented for illustration.

The colonization pattern is characterized by the presence of forest and deforested areas. In this case large patches of deforested land being spread within forest area.

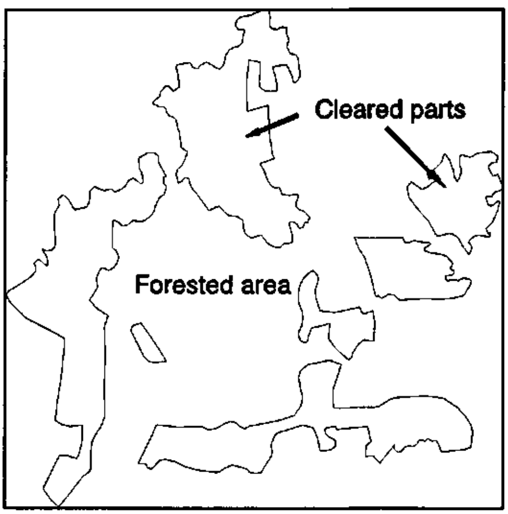
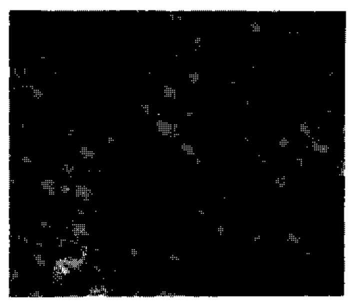


Fig. 2.2. *A colonization pattern;*



The deforested patches appear to be more or less randomly distributed. No larger rivers are present that could have determined the settlement pattern. The deforested patches can be considered the elementary objects. Grassland is found within the deforested area. No field pattern can be recognized within the deforested patches on the aerial photos. In case a field pattern can be discerned, it should be described and the entire unit should then be described in terms of an association. It is therefore important to denote whether the pattern elements are interpreted as agricultural fields or as other artefacts. The presence of agricultural fields is one the important criteria for the classification of the LUZs.

Another example is the **banana plantation pattern**. A banana plantation is characterized by a very regular layout of the production sectors. The sectors themselves are normally difficult to recognize, but the drainage canals and roads give a clear indication of the pattern.

Faces can not be recognised. Nevertheless, the layout of the drainage system and roads suggest the presence of larger units in a regular arrangement. The variation in tone and tex-- ture within and between the sectors is very small, indicating a uniform land cover. The example of Figure 2.3. represents a land use zone of which the boundaries largely corres- pond with river courses (i.e. the curved boundaries of the plantation).

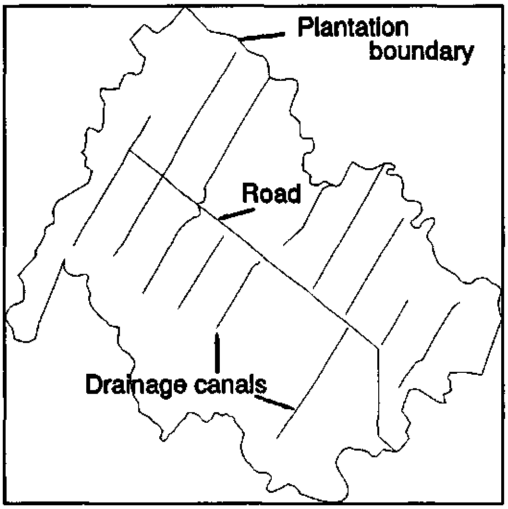
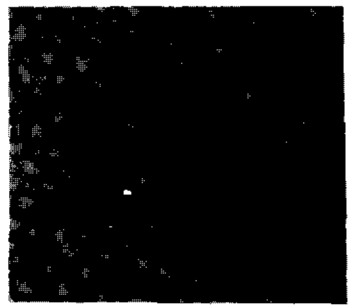
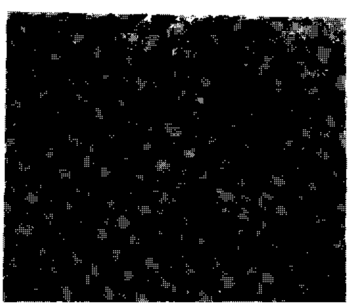


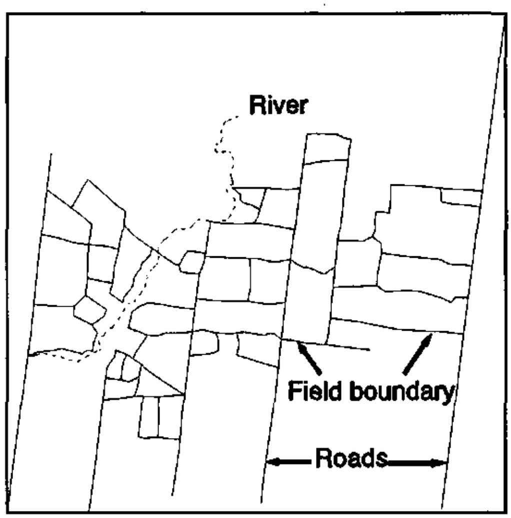
Figure 2.3 *Banana plantation.*



A third example is found in the **state-controlled settlement pattern**. All settlements schemes show a uniform distribution of the size of the faces, corresponding to agricultural fields (see Figure 2.4). The fields show a clear orientation and arrangement. Sometimes an extremely regular layout of the parcels is found. In these cases, a very systematic layout of the infrastructure is common. This depends on the topography, among other things.

Figure 2.4 *A settlement pattern*

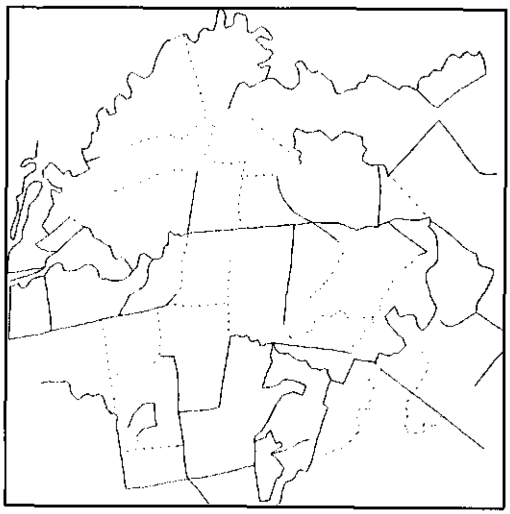




The area depicted in Figure 2.4 originally belonged to a banana. This may be recognized by the road pattern. The roads, parallel and at a fixed distance from each other, traverse the area. The river running through the area (broken line) disturbs the regular layout of the fields. Faces corresponding to agricultural fields are easily recognized. The boundary of the faces corresponds to living fences and and to change in cover (i.e. evident changes in tone and texture). The variation in tone and texture also indicates presence of various types of land utilization.

A relatively **young agricultural pattern** is characteristic for areas developed in the last 20 years. The more recently colonized areas are often characterized by the absence of living fences ('cercas vivas'), which hamper the recognition of the separate fields or parcels. Land use can be very uniform, which implies that parcels cannot be distinguished by changes in tone or texture. On the photo this results in some linear features that are clearly present but do not join to form faces. For many line features it is uncertain whether they actual represent boundaries (the broken lines). Many irregularly shaped features are present (areas delimited by drainage and rivers). Obviously, the degree of distinction of the faces is an important criterion.

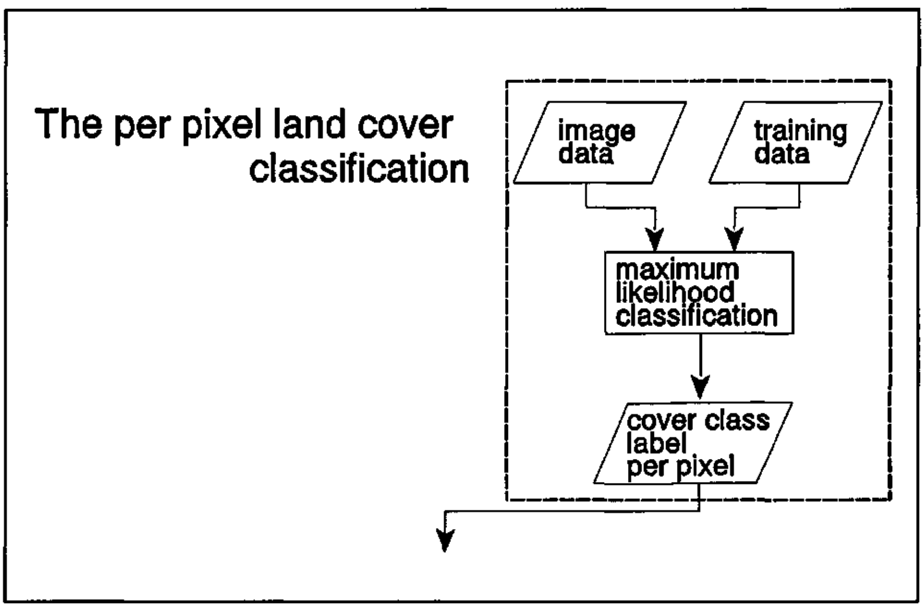
Figure 2.5 *Pattern of more recently colonized area used predominantly for grazing.*



Some small forested areas are present. Densely wooded areas are also found as are areas with scattered trees. Trees scattered over the fields reveals something about the management and intensity of grazing. Grasslands with many trees dispersed over the area are often found in the more recently opened up areas. The trees indicate that the area is not being grazed or, if so, only very extensively. Trees often represent future revenues, for which they are preserved.

CHAPTER 3

THE PER PIXEL LAND COVER CLASSIFICATION PROCEDURE



1. INTRODUCTION

This chapter describes a structured approach to land cover inventory using remotely sensed data. The focus is on the definition of the training statistics, which form the basis for good classification results. The procedure provides an efficient way to obtain adequate training statistics for classification of land cover. Only land cover is concerned, since this can be directly sensed by the sensor. Land cover includes both (semi-)natural vegetation and agricultural crops. Functional aspects of the land are only addressed in this chapter in as far these pertain to mapping the cover of the land. By adequate training statistics, we mean a set of classes fulfilling the requirements formulated by Swain and Davis (1978):

- The list of classes must be exhaustive;
- The classes must have informational value;
- The classes must be spectrally separable.

The literature devotes much attention to specific techniques and methods for classification. Comparatively little attention has been given to the selection and definition of the training statistics (Chuvieco and Congalton, 1988) as well as to the output phase. Both are important aspects of the classification process.

Inadequacy and inefficiency of traditional approaches, have been pointed out in the literature. The criticism of Buchheim and Lillesand (1989), for instance, refers to both locating and delineating supervised training fields with statistical properties appropriate for maximum likelihood classification. They attribute the inefficiency and inadequacy to the spectral complexity of second-generation multispectral data. The reasoning behind the merger and deletion decision concerning the spectral classes is said to involve more art than science. This need not be a problem when the classification product stands alone and when the analyst is adept at the art. But the fact that results are dependent on the personal decisions of the analyst can be a problem. This is the case when comparable classifications are to be executed for other regions or for different points in time.

The definition of a set of classes fulfilling the requirements mentioned above is a process in which object definition (definition of the land cover classes), spectral characterization and the recognition of the object are interwoven. The task requires insight in the complex field situation and in the spectral characteristics of the scene (as well as in factors explaining the spectral response). In case of Costa Rica the effort to meet these requirements is further hampered by:

- The heterogeneity of the land cover in the Atlantic Zone and the gradual transitions between the cover types;
- The discrepancy between scheduled and actual recording date of the scene (due to the high cloud cover);
- Rapid changes in land cover in the area.

A structure is need to steer the inventory process. At the same time it should document the activities and decision making process.

2. GENERAL CHARACTERISTICS OF THE AREA

The inventory of land cover was carried out in the Atlantic Zone of Costa Rica, a humid tropical region in Central America. Part of the area is used for agricultural production, varying from banana plantations and extensive grazing to subsistence farming. The areas of subsistence farming are strongly fragmented. Consequently, a large part of the pixel population will have a mixed spectral signature (mixed pixels) as result of boundary effects. Much of the area is covered by forest, which includes large swamps and peat bogs. Deforestation and colonization of new areas are dominant processes. The agricultural sector is characterized by pronounced changes in the relative shares of the different crops.

The topography is rather flat except for some old volcanic remnants in the coastal plain and the central volcanic mountain range, part of which falls within the Landsat-TM scene. The area to be mapped covered the complete Landsat-TM scene (path/row: 15/53). The recording date is February 6, 1986. Field work was carried out two and three years later in 1988 and 1989.

3. PHASES IN THE CLASSIFICATION PROCESS

The procedure presented is based on the supervised classification approach. It contains the sequential steps illustrated in the flow chart below (Fig. 3.1). The steps form part of the three phases of the digital classification process which are distinguished in the literature: training; classification; and output and evaluation. Part of the sequence is iterated several times until the set of training classes fulfils specified requirements. The steps are the following:

1. The selection of a feature combination for a color composite assumed to have a high information content (the feature combination that provides best discrimination between the cover classes). Also a principal component transform is carried out. These products are taken to the field to select locations for observation.
2. Fieldwork is carried out in one or a few areas considered representative for the region as a whole. In the field, the vegetation characteristics and terrain features are registered (see Cihlar *et al.*, 1987; Pouget and Mulders, 1988; these sources describe land cover and land surfaces for correlation with remote sensing data). Preliminary land cover classes are defined.
3. Definition of preliminary training statistics through selection of training fields. The training fields are selected in agreement with the spectral and spatial resolution of the imagery. That is, the only classes allowed have a limited range in spectral values and correspond to areas that are homogeneous in terrain and cover characteristics. For example, areas with the same land cover but varying slope and aspect (which causes deviating spectral signatures) are assigned as separate training fields. Training fields are selected for a number of sites with corresponding land cover. This is done in order to characterize the spectral dispersion and to prevent incorrect spectral definition of the land cover types due to possible change in land cover. (This might have occurred because the satellite data acquisition date and the dates of field data gathering do not correspond).

   In the second pass an unsupervised training approach is adopted. The homogeneity criteria ruling the clustering process are selected such that the size of the resulting spectral clusters corresponds to the smaller spectral classes as defined previously.
4. In the first cycle, the training statistics are tested using conventional methods: visual analysis of the feature space plots and calculation of errors of omission and commis

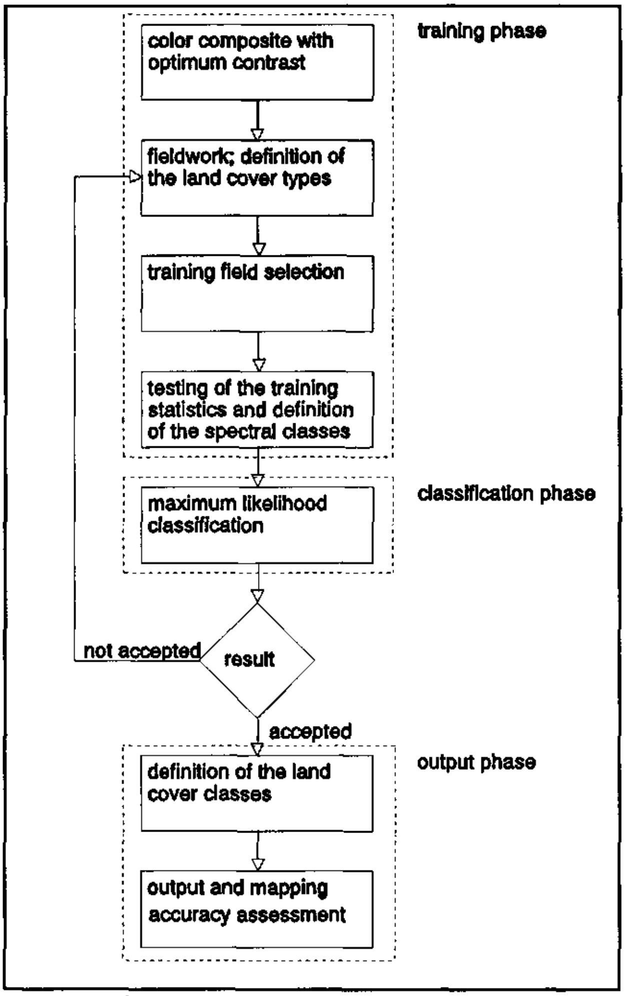


Fig. 3.1. *Flow diagram illustrating the subsequent steps in the classification process*

sion from the contingency matrix (see Richards, 1986). Classes with errors of less than 10 percent are accepted; others are evaluated for spectral overlap.
The spectral classes that together best represent the spectral domain of the total set of overlapping spectral classes are selected. This requires redefinition of land cover classes if the rejected spectral classes rejected pertain to land cover types not yet represented by the remaining spectral classes.
In the second pass of the cycle, the supervised training classes are evaluated for overlap with the training statistics obtained from the unsupervised approach.

5. A first classification is carried out with the maximum likelihood classifier (MLC). The MLC enables the specification of a threshold value based on likelihood values. The MLC uses the likelihood that a pixel belongs to one of the spectral classes as

criterion for the assignment of the pixels. The threshold value, specified by the user, determines the confidence region. Pixels with spectral values that locate them outside the region of confidence are left unclassified. A confidence level of 0.9 is taken.

6. The classification result is first checked against the percentage of unclassified pixels. The fieldwork, a repetition of step 2, is then directed to those areas corresponding to the unclassified pixels. Also the preliminary classification result is tested in the field. The final pass of the cycle only involves testing of the classification.
7. The spectral classes are mapped to the land cover classes. In defining the final land cover classes care should be taken that they do incorporate the land cover types associated with the earlier rejected spectral classes. If the content of the final cover classes does not cover all relevant land cover types the land cover classes have to be redefined. When spectral clusters form part of the final training statistics, the corresponding cover class is determined in the field.
   The information categories for final mapping are defined, taking into account the information requirements and the reliability of the field data (i.e. the reliability of cover class assessment). The information categories represent the appropriate level of generalization.
8. Finally the image is referenced in the output phase, if not before. Color coding and smoothing of the image is carried out before producing the land cover map. Finally the map's accuracy is assessed.

4. METHODS APPLIED

4.1 A priori feature selection for color composites.

Ranking methods are sometimes used for a priori selection of the spectral feature combinations (Lillesand and Kiefer, 1987; Mulders *et al.*, 1991). These rank the combinations by the total variance of the combinations and the correlation between the features. The methods consider the variance and correlation independent of each other. Therefore these methods do not provide an exact measure of the information contained in the feature combination.

This study uses a measure based on the variance of difference. In this case the variance of difference expresses the variance of the difference between two spectral features. The measure then sums the variance of difference of all the possible pairs of features that occur within the set of features defining the feature combination. The measure subsequently devides the sum by the number of possible feature pairs. The variance of difference measure is given by:

$$\frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N} [VAR(i) + VAR(j) - 2COV(i,j)}{0.5N(N-1)}$$

whereby i=1...N and j=i+1...N and j>i, N is number of spectral features in the feature combination.

A priori feature selection aims to provide the feature combination with highest information content in terms of independent variance. This does not mean that the feature combination will also have highest informational value. That is because the factors explaining the spectral variation might not be relevant to the mapping task. Information content, being defined as the probability of correct classification, can be estimated by the average pairwise transformed divergence (APTD, Singh, 1984; Swain and Davis, 1987). The use of the variance of difference measure as a means to accomplish feature selection is evaluated in Appendix B.

4.2 Spectral clusters to test supervised training statistics

In land cover inventory, classification strategies have been grouped in two categories: the supervised and the unsupervised classification approach (see Ahmad, 1986, for an overview). In the supervised approach, the spectral classes are defined on the based of the training fields selected by the user. Statistics of the classes are determined afterwards. The spectral classes statistically characterize the information categories. The unsupervised approach defines spectral clusters based on criteria for spectral homogeneity. The pixel population from which the clusters are derived is obtained by taking a (semi-)random sample from the spectral scene. The clustering identifies spectrally homogeneous groups in the image.

There are two reasons to evaluate the supervised training statistics against the unsupervised outcomes:

- First, to test whether the supervised classes cover the total spectral range (i.e whether the list of classes is exhaustive) and to correct for possible overestimation of contrast between spectral classes as a consequence of spatial (auto-)correlation effects (Campbell, 1981);
- Secondly, to critically review the size and spectral homogeneity of the spectral classes.

The procedure is illustrated in Figure 3.2. The spectral clustering is executed per image segment corresponding to the major cover categories (e.g. 'forest', 'grasslands', 'built-up area and bare soil') to accommodate the evaluation of the spectral classes. This gives some indication of the cover type to which the spectral clusters correspond. The segments are obtained through masking of the image. These masks are derived from the preliminary classification result. The parameters for the clustering were the number of pixels sampled; radius from population means, outside which new spectral classes are defined; minimum allowed distance between population means; and maximum allowed number of spectral classes. These are chosen such that the dispersion of the spectral data of the clusters corresponds to the dispersion of the values of the supervised classes.

Testing of the supervised training classes is done by submitting the total set of supervised spectral classes and the spectral clusters to a hierarchical cluster analysis. Squared euclidian distance between means is used as the criterion for grouping.

The result of the hierarchical clustering indicates class pairs or groups of classes with small differences in class means. The spectral divergence between classes is further evaluated by means of visual analysis of the feature space plots. (The feature space plots visualize the spectral overlap between the spectral classes in a two-feature space.) The classes that together best cover the total range in spectral values and show least spectral overlap are selected. The classes are thus evaluated on the basis of the distance between class means and size of the spectral classes.

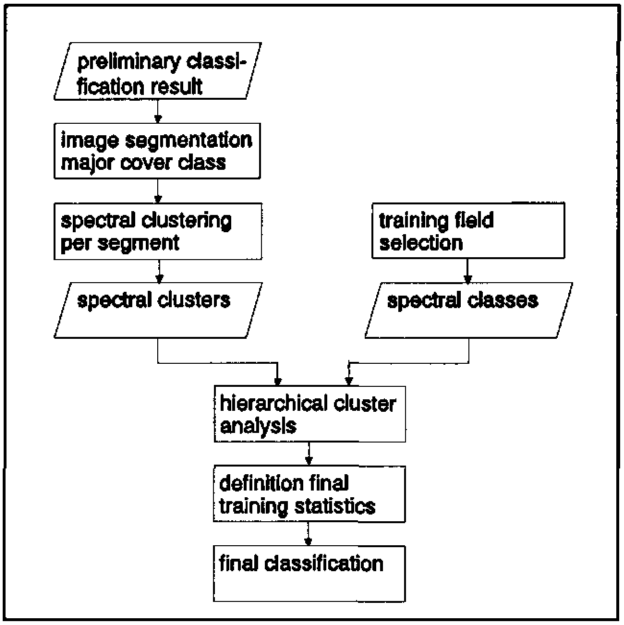


Fig 3.2 *Flow diagram illustrating the procedure for the integration and evaluation of the supervised and unsupervised training statistics.*

Chuvieco and Congalton (1988) also combine the training statistics generated by both the supervised and unsupervised approaches. These statistics are subsequently subjected to a hierarchical cluster analysis. Their focus however is slightly different. They use the groupings to interpret the meaning of the unsupervised classes and to improve the spectral definition of the information categories. They base the merger decisions on the findings of the hierarchical cluster analyses. No merging of classes is proposed by the present procedure, since this might easily lead to reduced classification accuracies.

4.3 Mapping of spectral classes to land cover classes

For classification output, the spectral classes need to be mapped to the information categories (the final cover classes). The process of expressing the classes of one classification system in terms of classes of a second system has also been called 'labeling' (Aronoff, 1984, who speaks of the assignment of 'resource classes' to 'image classes'). The process is shown schematically in Figure 3.3.

The correspondence between the spectral classes and the cover classes is examined. For mapping, the final cover classes are defined. These should reflect the appropriate level of generalization, giving expression to the internal variability of groups with similar spectral characteristics, as well as describing the difference between groups. Aside from the purpose of mapping, the appropriate level of generalization is determined by the accuracy and reliability of the measured land cover charactersitics.

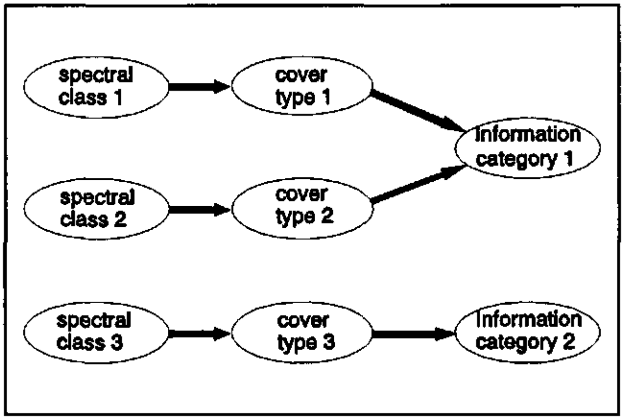


Fig. 3.3. *Mapping of the spectral classes to the cover types and informational categories.*

In the supervised approach, the correspondence of the spectral classes to the cover type is known. This relation is obvious because the spectral classes are obtained through sampling of the image for sites of known cover class. However, the date of field data recording and the image data recording often do not correspond. For that reason, uncertainty might exist concerning the cover characteristics at the time of recording. The uncertainty should be accounted for in the definition of the final cover classes. Reliability of the mapping results is increased by applying more general cover classes. These are obtained through a grouping of classes on the basis of corresponding characteristics (which includes the definition of wider class boundaries).

The spectral classes, when defined in agreement with the spectral and spatial resolution of the imagery, will often provide information on land cover that is too detailed for general purpose land cover mapping. Also in this respect the definition of more general land cover classes is called for.

The mapping is done by simple recoding of picture elements. That is, the class label for the initial cover class is replaced by the corresponding information category label). Only many-to-one relationships between the spectral classes and the final land cover classes are allowed.

Grouping of classes by mapping is preferred to merging of spectral classes. Merging of the spectral classes, when it is based on the logical grouping of the land cover classes, can easily create spectral classes with a large spread of spectral values. This will lead to a considerably lower precision of the classification. On the other hand, grouping of the spectral classes on the basis of the outcome of the hierarchical cluster analysis can easily result in heterogeneous information categories. The advantage of the proposed approach is that the original spectral classes are retained. Moreover, final output can be adapted through a simple recoding procedure to fulfil other information requirements (i.e. modification of the classification system).

5. RESULTS

5.1 Feature selection

On the basis of the 'variance of difference' measure, a combination of Landsat-TM bands 3-4-5 is selected for making the color composite. The combination was supposed to possess the highest information content. However, in our case it did not represent the most relevant combination for mapping of land cover in the area. Afterwards we evaluated the informational value of band combinations by comparing their 'average pairwise transformed divergence' (APTD, Swain *et al.*, 1971). This evaluation revealed that the band combination 2-4-5 provided the highest discriminative power between the land cover classes. However, the 3-4-5 band combination is shown to represent the next-best solution. In the Table 3.1, the 'variance of difference' and the APTD values are listed for nine feature combinations. Higher values indicate greater information content, respectively more discriminative power. (The APTD reaches a limit at 2000, representing complete separation of spectral classes.) See further Appendix 2 for a description of the variance of difference measure for a priori feature selection.

5.2 Preliminary classification results

For the first (i.e. preliminary) maximum likelihood classification, the application of a threshold value (confidence level of 0.90) resulted in an estimated 21 percent unclassified area in the image. However, the percentage varied for the different regions. This is probably due to the geographical distribution of problematic cases (e.g. areas with a fragmented and diverse land cover versus areas with uniform land cover, such as extensive grassland areas).

Table 3.1 *Values for the variance of difference measure and average pairwise transformed divergence for eight feature combinations.*

Feature combination	Var. of Diff.	APTD
3-4-5	493.7	1949
1-4-5	485.5	1936
2-4-5	476.6	1955
4-5-7	475.5	1920
3-4-7	403.0	1939
1-4-7	389.8	1922
2-4-7	375.1	1948
2-3-4	351.9	1922
3-5-7	163.0	1885

In many cases unclassified pixels were those with a mixed spectral signature. They corresponded to roads, rivers and riverbeds, residential areas in combination with homestead gardens, etc. In general the unclassified areas were small in extent. In some cases larger areas were left unclassified, representing cover classes not yet included in the training statistics. Examples include a plantation of plantain and one of a specific banana variety ('gran nain'). The classes were added to the set of cover classes.

5.3 The final set of training statistics

The first cycle yielded a set of 66 supervised classes. With the unsupervised method, 30 spectral clusters were added to the set of training classes, making a total of 96 classes. The hierarchical cluster analyses and the evaluation of the spectral classes (by means of the feature space plots) for redundant spectral information reduced the final set to 52 training classes. (See further Huising and Mulders, 1992.) The 52 classes were evaluated for errors of commission and errors of omission (through classification of the pixels corresponding to the training samples; see Richards, 1986). The results are presented in Table 3.2.

Table 3.2 *Accuracy and reliability of the training classes.*

Name of signature	Accuracy[1]	Reliability[2]	Name of signature	Accuracy	Reliability
Grasslands			Forest		
Pas01	93%	95%	For01	93%	100%
Pas03	97%	98%	For04/2	88%	91%
Pas04	96%	89%	For04/28	92%	92%
Pas05	100%	93%	For02/28	100%	100%
Pas12	91%	94%	For03/28	100%	100%
Pas06	81%	76%	For03/29	99%	98%
Pas21	100%	96%	For03	94%	45%
Pas07	99%	97%	Swv02	99%	100%
Pas24	95%	99%	Perennial crops		
Pas18	98%	99%	Ban02	93%	985
Pas19	98%	100%	Ban03	99%	98%
Pas20	100%	70%	Ban04	98%	90%
Pas22	100%	97%	Ban05	95%	90%
Pas23	100%	98%	Pla01	94%	100%
Secondary vegetation			Orn01	97%	99%
Sec02	93%	93%	Orn02	97%	94%
Sec06	96%	93%	Orn03	100%	100%
Arb01	97%	95%	Bam01	92%	97%
Sec01	92%	79%	Bam02	86%	70%
Sec05	89%	93%	Pej01	100%	92%
Sec07	89%	94%	Pal01	93%	100%
Sec09	88%	91%	Pal02	100%	100%
Residential area and bare soil			Pip01	100%	100%
Bua01	100%	100%	Pip02	94%	94%
Bso01	97%	100%			
Bso02	98%	98%			
Ped	100%	100%			
Lag01	100%	89%			
Rio	100%	100%			
Nub01	100%	100%			
Som01	98%	100%			

1 Accuracy is defined as 100% minus the error of omission.
2. Reliability is defined as 100% minus the error of commission.

The classification of grasslands gives some idea of the number of classes that might be involved in classification process. Initially, more then 40 training samples were taken, of which 19 were selected for the preliminary classification. In the second cycle, additional training samples brought the total to 24 classes, of which 14 classes remained after testing.

Large distances between means for the spectral classes and the spectral clusters were found only for those belonging to 'bare soil and built-up area'. These distances were due to differences in size of the spectral classes and clusters; the two spectral classes (Bso01 and Bso02) covered the same spectral range as the set of 4 spectral clusters.
The examination of corresponding land cover did not result in meaningful categories. For this reason the original bare soil classes were selected for the final training statistics.

5.4 Definition of the information categories and mapping of the spectral classes

Hierarchical cluster analysis generally resulted in the grouping of related land cover classes. This allowed for merging of the spectral classes. However, in some cases, groups of highly divergent land cover classes resulted. For example, bamboo was grouped with one class of secondary vegetation. Some grassland classes were grouped with certain classes of secondary vegetation. And some perennial crops were also grouped with secondary vegetation. For a description of the hierarchical cluster result, see Huising and Mulders, 1992.

Vice versa, merging of the spectral classes of corresponding the land cover classes would have reduced the accuracy of the classification. For example, considere the four information categories defined for the 1:100000 scale land cover map: 'Grassland 1', Grassland2', 'Grassland3' and 'Secondary vegetation'. In this case, by merging the spectral classes corresponding to the same information category two of the original secondary vegetation classes (corresponding to 'Sec1' and 'Sec9') would have been classified as 'Grassland3'. The grassland cover type 'Pas6' (pertaining to 'Grassland2': neglected pastures) would be classified as 'Secondary Vegetation'; Pas12 (also corresponding to 'Grassland2') would be classified as 'Grassland3' (wooded grasslands).

Given these negative consequences, no merger of spectral classes took place. Instead, general cover classes were defined and the spectral classes were mapped onto these general cover classes.

Two years lapsed between acquisition of the remote sensing data and the collection of field data. Consequently uncertainty about cover characteristics predominantly concerned grasslands, secondary vegetation and agricultural crops. The grassland cover types (listed in Table 3.2) were grouped according to grass species showing corresponding general plant characteristics. These cover types were also grouped according to management characteristics such as presence of weeds and shrubs in the pastures. The latter criterion, however, might not be considered a permanent characteristic. The general plant characteristics refer to general geometric characteristics of the grass species: general height of the plants; size and form of the leaves and stems; and whether the grasses grow in tussocks.
'Ratana' (*Ischaemum ciliare*) and 'Natural' (*Axonopus compressus*) pastures were grouped at the first level of generali-zation. 'Estrella' (*Cynodon nlemfluensis*) was retained as a separate group. Together these groups represent the information category 'Grassland 1'. This corresponds to pastures with short to medium-high vegetation with smooth vegetation surfaces.
'Guinea' (*Panicum maximum*), 'King grass' (*Pennisetum purpureum*), and pastures dominated by 'Cola de Venado' (*Andropogon bicornis*) are some examples of coarse grasses reaching up to three meters; These were grouped to 'Grassland 2'. Cover types pertaining to this class

might represent different uses. Guinea pasture is an improved pasture, whereas the other grassland types in this group often refer to neglected grasslands characterized by the presence of weeds, herbs, and small shrubs. Pastures containing trees were classified as wooded grasslands (Grassland 3).

Two general classes of secondary vegetation were defined:

1. The 'tacotales', representing very dense covers of shrub vegetation often containing 'platanillo' (*Musa spp.*). The shrubs vary from 1.50 to about 4 meters in height and with a rather smooth vegetation surface.
2. The more woody (somewhat older) secondary vegetation types. These contain shrubs and trees and have a less regular vegetation surface.

The definition of only two general classes, while large differences in vegetation characteristics might exist, reflects the uncertainty about the 'true' vegetation characteristics at time of scene recording.

With respect to the banana plantations a wide variation in spectral values was found. Deviating spectral characteristics might be due to different plant characteristics of the varieties used (height of the plants, size of the leaves). Deviation might also be due to different management characteristics. These could include general condition of the plants and plant density. Also, whether the fruits are transported over road or by cable to the packing station influences spectral response. (In the first case roads traverse the plantation, while in the second case the banana plants form a closed canopy.) Defining various categories of banana plantations would require further investigation of the correlation between spectral characteristics and cover characteristics. Therefore, only one class of banana plantation was defined. A distinction was made between plantations of plantain and the banana plantations.

The situation was more or less the same more or less the same for coconut plantations, palm heart plantations (stands of young pejivalle trees, *Bactris gasipaës*). That is, the two spectral classes corresponding to coconut plantation (Pip01 and Pip02) seemed to correspond to difference in tree density and height of the trees. Because the differentiation was not considered relevant to mapping, the consistency of this relationship was not examined further.

For the major cover class 'bare soil and built-up area', spectral clusters resulted after testing the training statistics. The associated land cover was investigated by means of classifying a test area and evaluating the results. The classification yielded fragmented bare soil areas. Furthermore, the fringes of the bare soil areas were classified as clouds. It was concluded that the spectral clusters had a spectral range that was too narrow to adequately represent the spectral variation associated with bare soil cover types. The original classes were selected for classification. One general class of bare soil and built-up area was defined for final classification. The list of final classes for the land cover map is presented in Table 3.3.

5.5 Final classification results

The final classification was checked in the field. A total of 240 observations were made, both along transects at regular intervals and at a number of randomly selected sites (though

Table 3.3 *Land cover classes for 1:100,000 scale mapping.*

0 Background	9 Banana
1 Forest	10 Bamboo
2 Peat and swamp forest vegetation	11 Ornamental crop
3 Secondary forest	12 Woody plantation crops: coconut, pejivalle, palmito, and other.
4 Wooded area	13 Bare soil and built-up area
5 Secondary vegetation	14 River courses and rocks
6 Grassland	15 River
7 Natural and degraded grassland vegetation	16 Lagoons and shadow of clouds
8 Plantain	17 Clouds

mostly referring to the agricultural areas). The items only registered were correspondence or non-correspondence of the classification result with the actual cover class. The percentage of correspondence varied from 64 percent for ornamental crops to 94 percent for the forest cover class. These percentages are listed in Table 3.4.

Testing was done only for a small sample. We felt that tests conducted three years after scene recording would not provide a very reliable indication of classification accuracy. In fact the grassland classes, which do not represent very stationary cover characteristics, show a low level of accuracy. At 95 percent confidence level, the minimum level of accuracy for the grassland classes is 10 to 11 percent lower than the values listed in Table 3.4. These figures apply to a sample size of about 50.

The low score of 64 percent for ornamental plants is partly explained by the wide variety of plants that exhibit clear differences in geometric characteristics. And in part it is also explained by the size of the plots; the small size prevents the determination of the spectral characteristics. The spectral classes gave a poor picture of the spectral range associated with ornamental crops. Two of the three training fields for ornamental crops referred to Caña India (*Dracaena massangeana*), since these are found on larger plots.

Table 3.4 *Percentage of correct classification of a number of land cover classes.*

Banana	80 %
Bamboo	82 %
Ornamental crop	64 %
Forest	94 %
Secondary vegetation	89 %
Wooded area	84 %
Pasture land	70 %
Secondary grasslands	72 %

Calculated in areal percentages, 87 percent of the image was correctly classified. This high percentage is partly caused by the high degree of accuracy for the classification of forest (94%) and the high percentage of the area covered by forest (38%).

A more detailed classification was tested for the grasslands. On the next-lower level of generalization, seven grassland classes were defined. Testing at 41 locations resulted in 18 cases of corresponding land cover classes (44%). In cases of erroneous classification the actual and classified value generally corresponded to related land cover classes (possibly as result of the object migrating to a related cover class). The tests indicated that the three grassland land cover classes represented the appropriate level of generalization.

6. DISCUSSION

6.1 The use of spectral clusters

The present procedure is based on the supervised classification approach. The use of spectral clusters is intended only for testing of the set of supervised training classes. The spectral clusters are not used for classification purposes because they may not discriminate between the relevant land cover classes. There are two reasons:

- The pixels for determining the spectral clusters do not generally represent homogeneous land cover because pixels are selected at random.
- Spectral clusters are defined according to homogeneity criteria. Therefore they do not reflect the characteristic spectral variability of land cover classes.

If testing reveals shortcomings in the training statistics, then additional supervised classes should be defined as replacements for the unsupervised spectral clusters.

6.2 Land cover and land use

The present procedure refers to land cover mapping only. Conventionally land cover and land use are mapped in associa-tion. However, land cover can be directly observed, as it refers to "the vegetation and artificial constructions covering the land surface" (Burley, 1961). In contrast land use can not be observed directly, as it denotes "man's activities on land which are directly related to the land" (Clawson and Stewart, 1965). Lillesand and Kiefer (1987) pose that ideally land cover and land use should be separated, but that from a practical standpoint it is most effective to mix the two systems. The common association of land cover and land use reflects the interdependent nature of activity and form, as well as the conceptually fuzzy distinction between them. Rhind and Hudson (1980) state that this interdependence is sometimes a critical assumption in land use surveys. This assumption was also made with regard to the land cover and land use survey in the Atlantic Zone of Costa Rica.

The application of land use criteria in defining the classes tends to result in overlapping spectral classes. On the other hand, the definition of cover classes on the basis of characteristics less relevant for the spectral response might cause spectral confusion.

The study in Costa Rica yielded many examples of diviating land use with corresponding land cover (and thus spectral) characteristics, and vice versa:

- The natural grasslands in and near lagoons have spectral characteristics similar to those of the grass vegetation in bottomlands. The latter are used for grazing in the drier periods, whereas the former are not used.
- A macadamia plantation (a tree crop) parts of which are also used for the cultivation

of maize by the labourers, differs spectrally from other macadamia plantations that have a grass vegetation as under-story. These also differ from plantations where the grass has been cut, shortly before recording the scene. There is no substantial difference in land use, while the spectral characteristics differ strongly.
- Cacao plantations (especially the older ones) are difficult to distinguish from other densely wooded areas, like wooded grasslands, wooded river banks, some reforested areas or secondary woody vegetation.
- Secondary shrub vegetation cannot be distinguished from a young planting, for instance of achiote (a tree crop used in producing dye).

The classes to be applied to remote sensor data should therefore refer to land cover characteristics. Land use can be inferred from land cover by considering additional characteristics, such as field pattern. Thre land use classification is considered separately from land cover classification.

The classification procedure would be improved if the definition of cover classes and their spectral expression could be based on known relationships between cover characteristics and spectral response. Such relationships are investigated in the appendix A.

6.3 Classification and aggregation hierarchies

Classification systems are often used in the classification of land cover using satellite imagery. The classification system described a hierarchical structure of materials. They enhance the understanding of a complex field situation. Problems may arise due to the improper definition of classification and aggregation hierarchies. This may happen especially with the second generation satellite imagery,

The USGS land use/land cover classification system (Anderson, 1976) is one example of such a classification system. Another is the CORINE classification system, designed to classify land cover and land use in Europe (CORINE, 1989). Both classification systems treat land cover and land use in association.

The various levels of the USGS classification system are designed to accommodate remote sensor data of varying resolution. Only the first and second level of the classification system are defined. These levels are principally used on a nationwide or statewide basis. Levels three and four, which provide information on a regional level, are intended to be designed by the local user.

Aside from representing different levels of aggregation, the hierarchy levels represent grouping of classes according to logical or functional criteria (e.g. classes 'Cropland and pasture' and 'Orchards, groves, vineyards, nurseries and ornamental horticultural areas'). The confusion between aggregation and generalization makes it very difficult to define the appropriate set of classes. This may be elucidated by considering some examples from Costa Rica. The classes, recognized using remotely sensed data, correspond to different levels of generalization. Also areas classified as homogeneous refer to different levels of aggregation. For example, banana plantations can be easily recognized and mapped because they cover large areas. Even specific varieties of banana crops (Gran nain, Valery, Cavendish) can be recognized. On the other hand a crop like 'Palmito' (young palm trees [*Bactris gasipaes*] harvested for palm heart), cannot be recognized as a specific crop. Palmito is generally grown in smaller units and it is liable to spectral confusion with other cover types. It can only be recognized as the general class of dense shrub vegetation.

The same applies to cacao (*Theobroma cacao*), which can only be recognized as the general class of 'wooded area'. The category of wooded area includes land cover classes like 'wooded river banks', 'densely wooded pastures', 'citrus plantations' and 'macadamia plantations'. With respect to ornamental crops, larger fields of Caña India (*Dracaena massangeana*) can be mapped. Other ornamental crops like *Marginata sp.*, generally grown on smaller plots, cannot be recognized.

Because of these considerations, no a priori defined set of classes was used for the Atlantic Zone land cover classification. Rather the procedure departs from the idea that it should first be investigated which classes can be discerned, given the resolution of the material employed. In a later phase (the mapping phase) final land cover classes and classification hierarchy levels are defined. Data aggregation is a problem in itself and is treated elsewhere.

6.4 Improvements of the classification procedure

Automated procedures could be employed for the feature extraction to further reduce the impact of the human factor (see Buchheim and Lillesand, 1989). However, problems would still be associated with the specification of the relevant parameter values ruling the feature extraction process. Also, the selection of seed pixels would still be a matter of concern. Further investigation should reveal the usefulness of such automated procedures.

In the present procedure, the distance between spectral classes and clusters is evaluated by measuring the distance between means (and also by visual interpretation of the feature space plots). Other distance measures could be used, such as the mahalanobis distance. Moreover, quantitative criteria could be specified for merging or deleting of training classes. One shortcoming of these measures is that they evaluate distances between two populations, whereas the set of classes should be to evaluated for spectral overlap. Visual analysis of the feature space plots is an efficient way to do this.

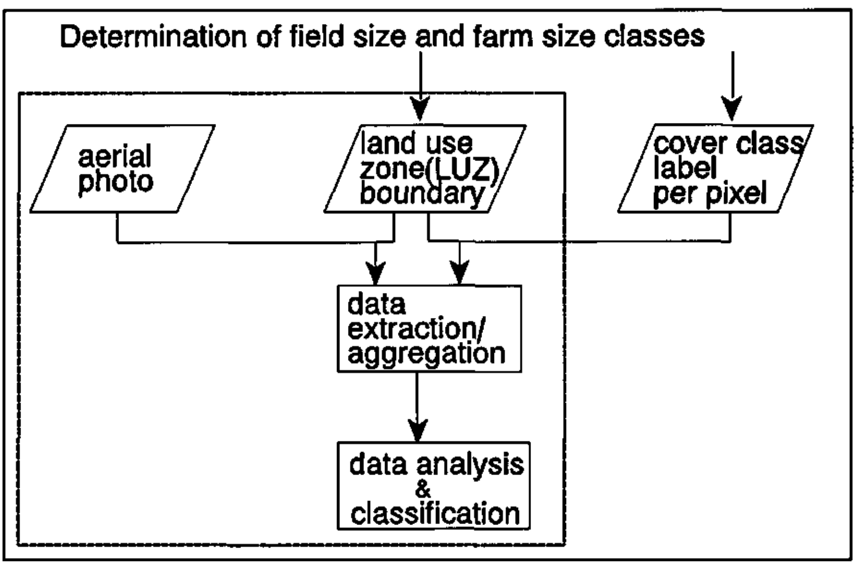
However, on the basis of the analysis of the feature space plots, alternative sets of classes could be defined. Measures like the average pairwise transformed divergence could then be used as quantitative criteria in selecting the best out of alternative sets.

For the mapping (or labeling) of spectral classes an optimized labeling algorithm has been designed by Aronoff (1984). The algorithm finds the label assignment which minimizes the loss function due to misclassification. Especially when costs due to misclassification can be specified, optimization of the labeling might be very valuable. These costs can often not be determined in a general purpose land cover mapping.

In the supervised approach, the spectral class are directly linked to the cover classes through the sampling procedure. The actual per pixel label assignment will generally be optimal one. More relevant is the question whether the appropriate cover classes have been defined. The content of the classes might change, especially during the evaluation of the training statistics, if classes are deleted, if - in exceptional cases - merging of classes occurs, or if new classes are added. Then it could be useful to evaluate the land cover class definition on the basis of classification result. The final phase in the classification process would then not only consist of the evaluation of classification accuracy. But at the same time it would be directed to the definition of the final land cover classes at the appropriate level of generalization. A loss function, as provided by Aronoff, could then be used to evaluate alternative sets of land cover classes.

CHAPTER 4

DEFINITION OF FARM SIZE CLASSES FOR SUB-REGIONAL LAND USE ANALYSIS AND MAPPING



1. INTRODUCTION

Farm size is an important socio-economic variable, as well as an important criterion for the characterization and classification of farms, because choice of crop and farm management are strongly related to farm size. Data on farm size can often be obtained per administrative unit. This will generally be statistical data on farm size distribution and will not provide very specific information.

The aim of stratification is to define units with more homogeneous features. Since farm size is an important socio-economic parameter for land use analysis and mapping, the land use zones are assumed to represent units with less internal variability on farm size. Furthermore, there should be clear differences in farm size characteristics between these zones.

With respect to the Atlantic Zone, information on the size and location of farms is not readily available. Data can be obtained directly through farm survey or through aerial photographs, for instance. The latter approach would be more efficient as it requires only limited fieldwork. Under certain conditions, fields size is expected to be related to farm size. Thus, field size could be used in inventory of farm size. The validity of this indicator is verified in this chapter.

A different but related question concerns the class definition. Land use is inventoried by means of the delineation and classification of LUZs. It implies that field size and farm size distribution, among other characteristics, is investigated for the LUZs. The question is whether LUZs can be discriminated on the basis of their field size and farm size characteristics. If so, the field size and farm size classes should be defined such that they reflect the degree of distinction which can be reached with the material available. In other words the classes should represent the appropriate level of generalization.

At the same time, the investigation of the distinction between the LUZs serves the evaluation of the result of the photo interpretation result. If no differences are found in farm size between the LUZs, then obviously the LUZs do not represent relevant spatial units. The procedure is depicted in figure 4.1. If the data does not permit the required distinction between objects, the result of the photo interpretation result is rejected.

2. DATA ACQUISITION: SIZE OF FACES, FIELD SIZE, AND FARM SIZE

In agricultural areas, LUZs can generally be considered an aggregation of agricultural fields. In the natural vegetated areas, no field pattern is observed. Nor are fields often recognized in the agricultural frontier area. Therefore, these latter areas should be ignored when field size is investigated in relation to farm size. It is assumed that the size of agricultural fields is correlated with farm size. Since we want to use aerial photos to obtain information on field size, we exclude all those units where the faces were not considered to represent agricultural fields (see chapter two).

The size of fields will be related to farm size only for certain land uses. This is especially likely in areas dedicated to grazing. There, size of the fields will be related to the size of the herds and thus to size of the farm. Such a relationship is also likely with respect to arable cropping and other uses. Small farmers who depend on family for labour will cultivate maize on smaller plots if. Larger farms, with possibility of employing machinery or hired labour, will tend to have larger plots.

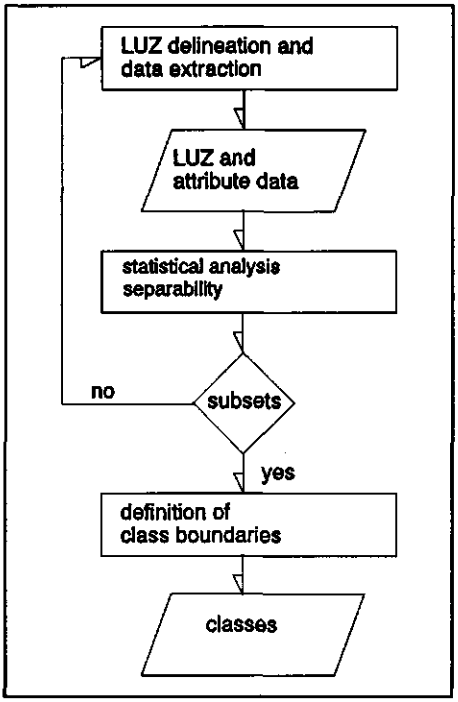


Fig 4.1 *Testing of the LUZ and class boundary definition.*

This assumption does not hold for intensive crops, like ornamental plants. These crops are grown on comparatively small plots, eventhough the farms might be large. Another example is the cultivation of bananas. The plantations are organized in production sectors. The size of these sectors does not give an indication of the size of the plantation. The number of sections or the number of stations for washing and packing of the fruits might be a more meaningful indicator of size.

Therefore, we should select only the LUZs with mixed land cover (arable cropping and grazing) or areas predominantly used for grazing. These LUZs are sampled with respect to field size. Data on field size is obtained through a grid point count, whereby a grid point corresponds to a certain area size (in hectares). Statistical data is recorded per land use zone. The data includes minimum value, maximum value, mean, median and standard deviation.

Information on farm size distribution was obtained through farm interviews (see Chapter Three of Part One). The interviews were carried out in the same LUZs for which data on field size was obtained. Also with respect to farm size, the minimum, maximum, mean, and median values were recorded per land use zone, in hectares as well as in farm size class. In this procedure, farm size classes were only used for the purpose of scaling the farm size data. The variance or standard deviation in farm size will increase with increasing mean farm size. (For a land use zone with large farms the difference between the size of the individual farms will be high also, in absolute terms.) To correct for this effect the farm size data is scaled according to the two schemes shown in Table 4.1. Two different schemes were applied to account for possible effect of the scaling pattern itself.

Table 4.1 *Scaling of farm size data.*

Scheme I:	Class:	Scheme II:
0 - 5 ha.	I	0 - 5 ha.
5 - 10 ha.	II	5 - 15 ha.
10 - 20 ha.	III	14 - 45 ha.
20 - 50 ha.	IV	45 - 135 ha.
50 - 100 ha.	V	135 - 440 ha.
100 - 200 ha.	VI	> 440 ha.
> 200 ha.	VII	

Based on the collected, we attempted to answer the following questions:

1. Does the data on size of faces, as obtained from the aerial photo, allow for discrimination between the LUZs?
2. Are field size and farm size related?
3. Can from the distribution of fields size conclusions be drawn on the distribution of farm size within the LUZs.

3. METHODS

The class definition for field size, with respect to the LUZ involves defining clusters in a one-dimensional feature space. For each LUZ many observations are made. Differences between the LUZs are evaluated by the one-way analysis of variance. The comparison of all possible pairs of group means is called the multiple comparison (Scheffé, 1959). It produces a matrix in which all the significantly different group means are indicated. Those LUZs with a small number of field size observations were excluded from the analyses. These zones were generally small in size. On the basis of this matrix homogeneous subsets are defined. These consist of groups (i.e. LUZs) with means that do not differ significantly. The minimum and maximum values of these subsets can be used to represent the class boundaries.

Before the LUZs are subjected to analysis of variance, the results of the photo interpretation are checked by evaluating the distribution characteristics. LUZs with a wide spread in data values, bi-modal distribution, or skewed distribution patterns (as indicated by a large difference between median and mean values for size of faces) where traced on the aerial photo. If indeed an interpretation error was found, the boundaries of the land use zone were redefined.

The relation between field size and farm size was analyzed by means of regression analysis, though at land use zone level. This implies that the regression analysis concerns object characteristics like mean or median fields size and farm size. It does not concern the individual measurements of field size and size of farms. It is investigated which of the statistics expressions of field size distribution (mean, median and other), best indicates farm size.

3. RESULTS

The highest correlation was found between the land use zone mean field size and mean farm size class (with referring to the first classification scheme for scaling of the farm size, Table 4.1). An R-square of 0.87 (N=37) was obtained using non-linear regression (see Figure 4.2).

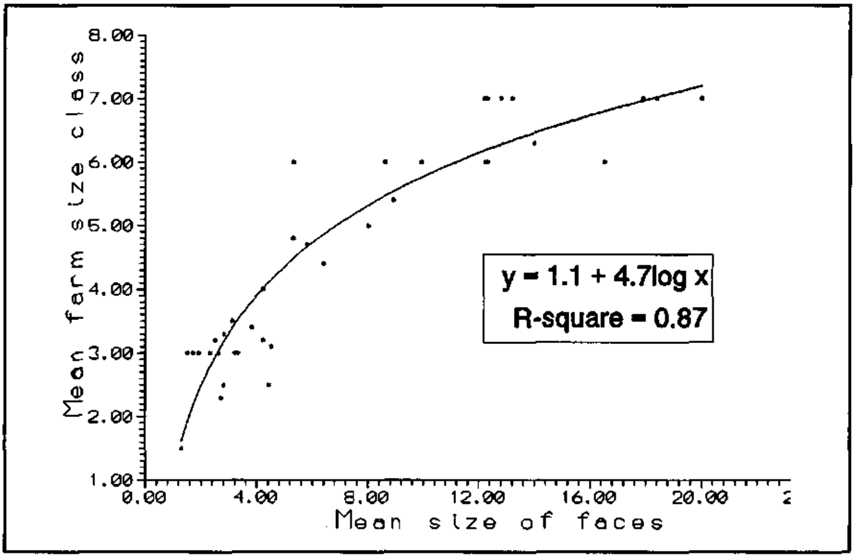


Fig 4.2 *Mean farm size class as function of the mean size of faces of the land use zones.*

The linear regression produced an R-square of 0.81. The points on the graph represent LUZs. The lower correlation for the linear regression is simply explained by mean farm size class reaching an artificial determined maximum of seven, while the mean size of faces is not delimited. (The highest class represents a range in farm size from 200 to a few thousand hectares). By leaving out the four LUZs with highest mean size of the faces we obtained almost the same correlation coefficient (R-square = 0.86 at N=33) for the linear regression. The use of the median values yielded lower results for both size of the faces and farm size.

The one-way analysis of variance indicated a significant difference between the LUZs with respect to their mean size of faces. The matrix below (Table 4.2) indicates the results of the multiple range test. The asterisk denotes pairs of LUZs with significant difference in mean size of faces. On the axes the LUZs are indicated by their identifier. Also listed is their mean value for the size of faces. This result was obtained with Duncan's multiple range test applying a significance level of 0.10. Comparable results were obtained with the application of the Least significant difference (LSD) range test (see SPSS, 1983) and an α of 0.05.

On the basis of these results, we can define homogeneous subsets of LUZs that do not differ significantly as regards their group means. The matrix should be read as follows:

LUZ no. 14 (first column, the number should be read vertically), with a mean field size of 1.12 hectare does not differ significantly from LUZ no. 17 (row) with a mean value of 4.25. Together with all the LUZs that have a value in between, they define a homogeneous subset. In the same way, a subset of LUZs with no statistical difference in mean field size can be defined for each column (i.e. each land use zone).

Many of these subsets will show strong overlap. They can be identified through the matrix. For example, the subsets corresponding to the first seven columns show a very strong overlap in their group members. All subsets contain the same LUZs with means equal to or less than 3.81, referring to LUZ no. 10. These subsets can be merged to define one homogeneous <u>set</u> whereby LUZ no. 10 represents the highest mean value.

Table 4.2 *Significantly different land use zones with respect to the mean size of the faces, as determined by the Duncan's multiple range test.*

LUZ Identifier.[1]

Mean	LUZ Id.	14	40	88	26	75	99	53	34	93	45	60	145	33	94	80	23	76	10	89	17	36	100	56	84	39	65	01	91	13	58	124	30	138	113	72
1.12	14																																			
1.14	40																																			
1.20	88																																			
1.26	26																																			
1.28	75																																			
1.36	99																																			
1.64	53																																			
1.66	34																																			
1.86	93																																			
1.92	45																																			
2.32	60																																			
2.40	145																																			
2.56	33																																			
2.77	94																																			
2.84	80																																			
3.20	23																																			
3.30	76																																			
3.81	10						*																													
4.00	89		*		*		*	*																												
4.25	17		*	*	*		*	*																												
4.98	36	*	*	*	*	*	*	*	*	*		*																								
5.02	100	*	*	*	*	*	*	*		*																										
5.46	56	*	*	*	*	*	*	*	*	*		*																								
6.04	84	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*																				
7.08	39	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*															
7.36	65	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*															
8.90	1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*											
9.92	91	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*											
10.40	13	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*											
12.11	58	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*								
12.16	124	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*								
13.76	30	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*							
14.00	138	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*							
17.60	113	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		
23.04	72	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	

* denotes pairs of groups significantly different at the 0.100 level/ [1] Number should be read vertically

The next columns that represents strongly overlapping subsets are the columns of LUZ no. 34 through LUZ no. 60. These are also merged to define a homogeneous set with a minimum mean value of 1.6 hectares (corresponding to LUZ column no. 34) and a maximum mean value between 4.25 hectares corresponding to LUZ row no. 17 and 4.98 ha. corresponding to LUZ row no. 36. The asterisks indicate that LUZ 36 is significantly different from LUZs 34, 93 and 60, because of which it is not part of the homogeneous set.

The next sets of strongly overlapping subsets can be defined in the same way. These sets are indicated by vertical and horizontal lines. The vertical lines denoting the start of a new set of overlapping subsets. The first LUZ in the new set determines the lower boundary. The horizontal lines denote the upper limits of the sets. The corresponding mean field size of the first and the last LUZ of the homogeneous set denote the minimum respectively the maximum boundaries of the classes for the mean size of faces. The classes are listed in Table 4.3a. However, the classes still show too great an overlap in value. The second class, therefore, was discarded, and the fourth and fifth class were merged, resulting in classes as listed in Table 4.3b.

Table 4.3 *Class boundaries for the mean size of the faces. A shows the total number of homogeneous subsets; B gives the reduced set with less overlap between the classes.*

A.	B.
1.0 - 3.8 ha.	1.0 - 3.8 ha.
1.6 - 4.6 ha.	2.4 - 5.5 ha.
2.4 - 5.5 ha.	4.9 - 11.0 ha.
4.9 - 8.0 ha.	9.9 - 17.6 ha.
6.0 - 11.0 ha.	> 14.0 ha.
9.9 - 17.6 ha.	
> 14.0 ha.	

The above mentioned classes of mean size of the faces can be translated into classes of mean size of farms. This can be done by using the regression equation for the correlation between size of face and farm size. However, the classes are intended to describe the range of farm size occurring within a specific land use zone and not the mean value. Therefore, the deviation from the mean is taken into account in translating upper and lower limits of the field size classes to farm size classes boundaries. This will generate the classes as listed in Table 4.4. For the translation of field size to farm size use was made of the linear regression equation obtained for the reduced set of 33 observations.

Table 4.4. *Farm size classes.*

Class 1.	0 - 29 ha.
Class 2.	10 - 60 ha.
Class 3.	30 - 260 ha.
Class 4.	85 - 1000 ha.
Class 5.	> 400 ha.

This class definition represents a set of farm size classes, appropriate to classification of LUZs. 'Appropriate' means that the size of the classes reflects the internal variability of the LUZs in terms of fields size (and farm size) and that it effectively describes the difference between LUZs. The definition of more specific classes would not be very useful. That is because the internal variability of the LUZ as regards to field size does not justify such detailed description of farm size. Definition of more general classes would not make use of the full discriminating potential of field size data. More general classes would therefore imply a loss of information. The result reflects the quality of the aerial photo interpretation (the LUZ map). Different class boundaries would result from a different aerial photo interpretation.

Since our objective is to develop the aerial photo interpretation as a tool for land use inventory, the emphasis is on the aerial photo characteristics (field size, in this case). The width of resulting farm size classes fully depends on field size characteristics of the LUZs: difference in mean field size between LUZs; mean standard deviation of field size of the LUZs. Also, the number of observations of fields size per LUZ will influence the outcome of the multiple range test and herewith the width of the classes. In defining the farm size classes the only farm data used is the LUZ mean farm size. Distribution of farm sizes within the LUZs has not been explicitly taken into account in defining the farm size classes. The relevant question then of course is to which degree the individual farms within a LUZ correspond in size to the farm size classes predicted on basis of the field size data. To answer this question, the original farm size data corresponding to the individual farms spread over nine LUZs were compared with the farm size class determined on the basis of the field size data. It proved that for 83 percent of the farms (N=78) farm size was within the boundaries indicated by the farm size class. The percentage per land use zone varied between 60 and 100 percent.

The relation between field size and farm size is determined on the basis of data for the relevant LUZs in a study area of 30 by 30 kilometers. The area is considered representative for the total northern part of the Atlantic Zone. Application of the approach to mapping of farm size to other parts of the Atlantic Zone requires the investigation of the validity of the relationship established between mean field size and mean farm size for those parts. This could not be verified because of lack of data at this time. For regions in which different land use characteristics and other socio-economic conditions prevail, the relationship at issue should be established independently.
In this chapter it is demonstrated that:

1. A clear difference in field characteristics exists between the LUZs.
2. The LUZs exhibit clear differences in farm size composition (see Chapter Three, Part One);
3. Farm size can be mapped on the basis of mean field size, given the relationship between mean field size and mean farm size.

4. DISCUSSION

The class boundaries for farm size classes that are used in socio-economic (regional) analysis are seldom subject to discussion. Choice of class boundaries is either based on convention or is made rather arbitrarily. For example, in the socio-economic analysis of the Rio Jiménez area in the Atlantic Zone of Costa Rica, the following farm size classes are used without further reference (Waaijenberg, 1990):

< 4 ha.
4 - 20 ha.
20 - 50 ha.
50 - 200 ha.
> 200 ha.

In a related study concerning a different area, two different classification systems are used. No further clarification is provided on the class boundaries used (Wielemaker, 1990):

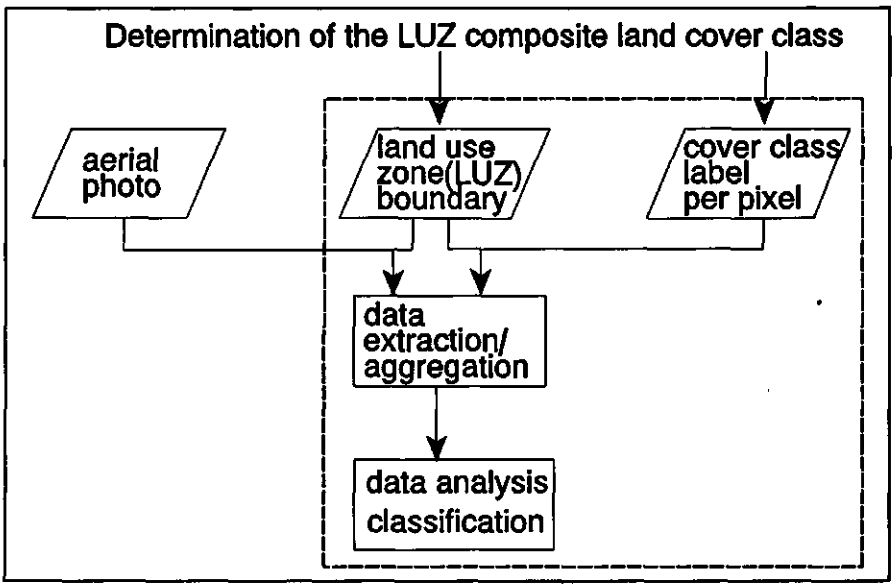
< 20 ha.	< 30 ha.
20 - 50 ha.	30- 60 ha.
50 - 200 ha.	60 - 120 ha.
> 200 ha.	120 - 250 ha.
	> 250 ha.

In these studies the farm size classes are defined a-priori. These studies investigate the forms of land use, agronomic or socio-economic characteristics associated with the given farm size classes. Areas or regions are characterized and compared with respect to their distribution pattern for farm size, among other variables. The areas which are compared are sometimes defined using different criteria (as was the case in the mentioned examples).

The land use zone approach is founded on a different concept, namely that of homogeneous LUZs. The focus is on the definition of units or regions relevant to the mapping of land use, instead of having to resort to more or less arbitrarily defined regions. The classes are chosen to express characteristic difference between the units of interest, as well as the variation within the units. It is a data-driven approach, which reflects the type, scale, and quality of the material used in defining the appropriate units and classes. Since the farm size classes are not a-priori defined but result from the procedure, the relevance of the classes for the characterization of farms and land use should be investigated afterwards. This aspect is evaluated in Chapter Three, Part One.

CHAPTER 5

DEFINING COMPOSITE LAND COVER CLASSES, TAKING ACCOUNT OF GEOMETRIC ACCURACY AND ACCURACY OF THE LAND COVER CLASSIFICATION



1. INTRODUCTION

This chapter, like Chapter 4, deals with the problem of class definition concerning aggregated objects. Aside from the field size distribution, land use zones (LUZs) are characterized by their land cover composition. Also in this case the issue is whether the land cover composition allows us to discriminate between the land use zones.

On the one hand, the analysis of the separability of the LUZs serves to evaluate the photo interpretation, since the objective of the photo interpretation is the identification of LUZs with distinct characteristics. On the other hand, the analysis serves to define the attribute value classes that should adequately reflect the discriminative power of the composite land cover attribute.

The problem at hand is different from the one treated in the previous chapter. Here we are concerned with the classification of a multi-dimensional discrete variable. This task requires different analytical techniques. Nonetheless, the principle remains the same, namely, clustering in an n-dimensional space.

An important aspect of the class definition is the error and uncertainty of the input data in relation to the class definition. With respect to the mean field size, the classes were defined such a way that they reflected the internal variability of the land use zones, while maintaining the power to discriminate between the zones (see Chapter 4). As such reflecting the appropriate classes for describing the variation in field size.

We want to define also appropriate classes for the classification of land cover composition of the LUZs. With respect to the accuracy of the input data for this classification two factors are of importance:

1. The assignment of the pixels to the regions. Hutchinson (1982) acknowledges the problem of discrete transitions in stratified maps. This problem emerges in the context of combining Landsat and ancillary data. The object thematic content is dependent on that geometry of the object. The accuracy with which the geometrical characteristics are determined will influence the reliability of the thematic attribute data and thus the class definition.
2. The assignment of the pixels to the classes. The cover classes are the result of an earlier classification process. Thus they have a certain associated reliability. This influences the reliability of the aggregated object classification.

Therefore, the main questions addressed in this chapter are:

1. What is the degree of inaccuracy of the thematic attribute values pertaining to the LUZs? What are the causes of inaccuracy?
2. How can the inaccuracy be accounted for in the definition of the appropriate set of composite land cover classes?

2. MATERIALS AND METHODS

Data on land cover composition of the land use zones is obtained by overlaying the land cover classification with the LUZ map. The land cover classification is described in Chapter 3; the delineation of the LUZs is described in Chapter 2.

2.1 Geometric transformation

Before the land use zone definition and the land cover classification result can be overlaid, they have to be referenced to the same coordinate system. Both the satellite imagery and the aerial photo interpretation were transformed to the local coordinate system (Lambert projection).

With regard to the registration of the satellite imagery, an upper limit for the RMS (Root of Mean Squares) is normally defined for the transformation result. This is generally set from 1.0 to 1.5 pixel. In the case of Landsat-TM this corresponds to between 30 and about 50 meters. Errors are caused by deviations in the identification and location of control points. Errors also depends on the degree of the transformation polynomial (see Buiten, 1988).

For the Guacimo-Rio Jiménez-Siquirres (GRS) study area in the Atlantic Zone of Costa Rica, a single photo numerical restitution was carried out. The application of this method requires information on the elevation for every point. A Digital Elevation Model (DEM) was constructed by first digitizing the contour lines from the topographic sheets. This was followed by rasterization and interpolation to obtain elevation data per raster cell. Transformation parameters were determined on the basis of carefully selected reference points. The procedure is schematized in Figure 5.1.

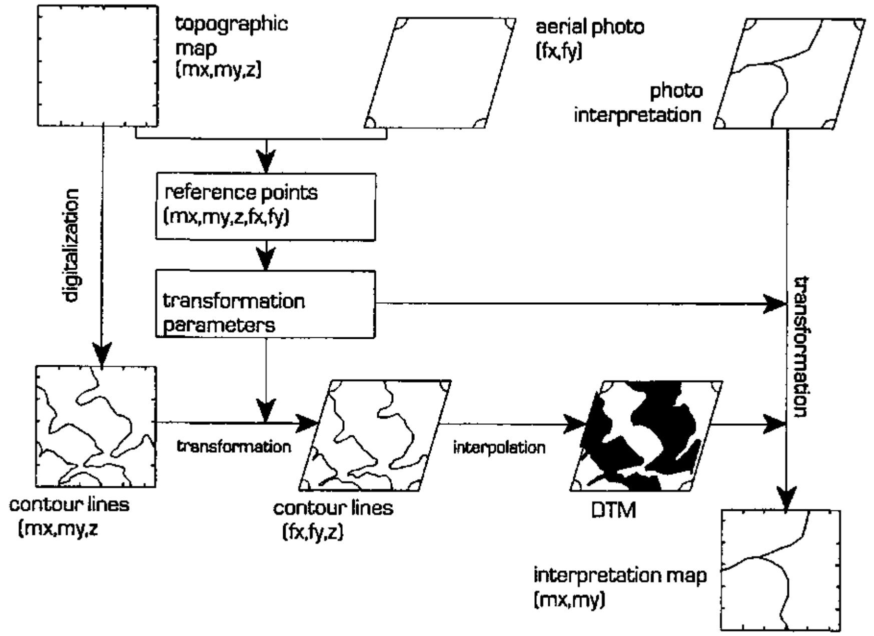


Fig. 5.1 *Single bundle resection and monoplotting operation.*

The procedure was carried out with a program developed at the Agricultural University Wageningen. It operates on a personal computer under MS-DOS with peripherals (a digitizing tablet and a plotter). See Molenaar and Stuiver (1987) for a description of the monoplotting procedure.

2.2 Overlay and data aggregation

The GRS study area in Costa Rica consisted of 158 LUZs. This map was overlaid with the land cover classification. For every LUZ, the percentage of each of the 16 land cover classes was obtained (illustrated in Table 5.1). The land cover classification of the GRS area, with the LUZs overlaid, is presented in the appendix.

The percentages had to be corrected for the effect of cloud cover and for the unclassified area (result of incomplete matching of the coverage boundaries). The area of the LUZ not classified or corresponding to clouds was distributed over the other cover classes, in accordance to the portion each cover class occupies in the LUZ.

Table 5.1 *The matrix of land cover class percentages*

LUZ Identifier	Land cover class percentages					
	LCC 1	LCC 2	---	---	---	LCC 16
1	$X_{1,1}$	$X_{1,2}$	--	--	--	$X_{1,16}$
---	--	--	--	--	--	--
---	--	--	--	--	--	--
158	$X_{158,1}$	$X_{158,2}$	--	--	--	$X_{158,16}$

2.3 Class definition through cluster analysis and by defining a threshold value for the cluster solution

The discrimination between the LUZs by land cover composition is analyzed by means of a hierarchical cluster analysis. And on the basis of the cluster solution, the composite land cover classes are defined.

The hierarchical cluster analysis evaluates, in a hierarchical order, the distances between cases or groups of cases. Cases or groups thereof are linked in order of increasing distance. Two linked cases represent a group. The group, with its associated group values, will be compared to other cases or groups to form new groups, etc. At the end of the clustering procedure, one group is left. That group incorporates all the original cases.

For each distance, a corresponding cluster solution can be found. The distance at which to accept the cluster solution is the minimal distance at which two groups can be considered different. This minimum is determined by the accuracy with which the distance between groups can be calculated. Once the threshold value is determined and the cluster solution accepted, the minimum and maximum values of the resulting clusters are used to define the class boundaries.

We selected average linking between groups as cluster methods, using the city block as the measure of distance. The city block measures the 'walking distance'. It means that the distance between LUZs is determined by the sum of the difference in percentage for each cover class.

The minimal required distance between LUZs is determined by the accuracy with which the land cover percentages can be calculated. Once this threshold value is defined, the cluster solution will serve as a basis for the class definition. The LUZ with the minimum value in the cluster will thus define the lower boundary of the composite land cover class. The land use zone with the maximum value will then define the upper boundary of the composite land cover class. (For each of the 16 land cover classes, the minimum and maximum values are determined per composite land cover class.)

The land cover composition of a LUZ is determined by the pixels pertaining to it. Accordingly, the reliability of the cover class percentage is related to the certainty with which a pixel's membership in the LUZs can be determined. It is also related to the reliability of the land cover class label that is assignment to a individual pixel.

The reliability of the assignment of a pixel to a LUZ depends on the geometric accuracy. And this, in turn, depends on the precision of the transformation process, as well as the precision with which reference point can be identified and located. The accuracy is expressed in terms of discrepancies between measured and calculated points. Maximum errors are used as a basis for determining the threshold value. This is done to assure that the final classes will capture the variation in the composition of land cover class as consequence of the geometric inaccuracy.

The reliability of the LUZ land cover class percentage is depends on the accuracy of the land cover classification. The listed land cover percentages are in fact estimates of the actual land cover percentage. For each estimation of a land cover percentage a confidence interval can be calculated, taking into account the reliability of the per pixel assignment of that specific land cover label. The confidence interval can be calculated on the basis of the probability of class 'A' occurring *r* times, given *s* occurrences of pixels classified as 'A'. This probability is based on two terms, namely:

1 The probability of class 'A' occurring, given that the pixel is classified as 'A' [P(A | A)], and;
2 The probability of occurrence of class 'A', given that the pixel is not classified as 'A' [P(A | not A)].

The formula is as given below:

$$P(A_r)=\sum_{i=0}^{r}[(\binom{s}{i}(p_1)^i(1-p_1)^{s-i})*(\binom{n-s}{r-i}(p_2)^{r-i}(1-p_2)^{(n-s)-(r-i)})]$$

where:

n = total population (number of pixels of a LUZ)
r = number of occurrences of 'A'
s = number of occurrences of pixels classified as 'A'
p1 = P(A|A) is the conditional probability of A given A, with A being the event of class A (classified as A)
p2 = P(A|¬A) is the conditional probability of occurrence A given ¬A (not A), with ¬A being the event of not class A (not classified as A).

3. RESULTS

3.1 Accuracy of the geometric and thematic object attributes

Geometric accuracy

With respect to geometrical accuracy, the orientation of the satellite imagery relative to the LUZ map must be coordinated, since these are overlaid. The possible relative error is ascertained by adding the registration error of the satellite imagery to the possible deviation in geometry of the land use zone map. The former was set to 1.5 pixel (given as a parameter in the transformation process), corresponding to 50 meter in the field.
As regards the geometric accuracy of the land use zone boundaries, the following sources of error have been identified (see also Buiten, 1988):

- The accuracy of the LUZ delineation. The line thickness gives an indication of the accuracy with which a line (boundary) can be drawn on the aerial photo. It is taken to be 0.3 millimeter, which corresponds to about 25 meters in the terrain, considering the 1:80,000 scale of the aerial photos.
- The accuracy of the subsequent digitalization is taken to be 0.2 millimeter (when done accurately), corresponding to 16 meters in reality. (This is slightly more than the 0.15 millimeter mentioned by Buiten (1988) for the digitalization accuracy; that, however, concerns the digitizing of control points.)
- The accuracy of the transformation itself. In the case of Costa Rica a deviation of 40 meters was accepted for the results of the numerical restitution.

A discrepancy of 40 meters between measured and calculated reference points was accepted. This margin takes the quality of the topographic sheets into account. Their quality affects both the accuracy of the DEM and the precision with which the coordinates of the reference points could be determined. An accuracy of about 80 meters is obtained for the geometric precision of the LUZ boundary, by adding the errors from the different sources. This implies that when we compare the land cover classification and the LUZ map, the maximum relative error,is 130 meters (80 plus 50 meters). This figure corresponds to a maximum deviation of 4 to 5 pixels, which was indeed observed in some places. The expected standard error is 70 meters (obtained by $\sqrt{\Sigma\sigma_i^2}$, whereby sigma denotes the standard deviation in meters associated with factor *i*).

Land cover classification accuracy

The accuracy of the land cover classification was estimated by doing field checks. This methods was preferred to assessing the probability of class membership as derived from the conditional probability for the spectral values, given a certain spectral class. These can often be requested as additional output of the classification, depending on the type of classifier applied (Wu et al., 1988; Kenk et al., 1988; Wilkinson and Mégier, 1990). The probability that a pixel 'x' belongs to a certain spectral class does not need to correspond to the probability that 'x' correctly identifies actual land cover. Actual land cover was recorded along three transects in the field and for a number of randomly selected sites. The observations were compared with the classification values. Table 5.2 presents the results.

The figures lie within the 50 to 90 percent range generally reported for classification accuracy. Janssen et al. (1990) mentions 72 and 76 percent classification accuracy obtained in two agricultural areas in the Netherlands. Kenk (1988) mentions accuracies of 70 percent or more in terms of identifying the correct class. The classification performance was better than the classification accuracy mentioned by Ioka (1986); the performance ranged form 46.8 per cent to 76.8 percent, depending on the level of generalization. The percentages were used as a measure of the probability of correct classification.

Table 5.2 *Percentages of correct classification for a number of land cover classes.*

Banana	80 %
Bamboo	82 %
Ornamental crop	64 %
Forest	94 %
Secondary vegetation	89 %
Wooded Area	84 %
Pasture land	70 %
Secondary grasslands	72 %

Reliability of the LUZ land cover percentage

The LUZ land cover percentage is determined through aggregation of pixels (elements). The reliability of this percentage depends on the reliability of the elementary observations (the per pixel land cover class). The error in relative positioning introduces error in the composition of the land use zone. Pixels and their associated land cover class are included in LUZs to which they do not belong. The error it introduces will be referred to as 'pollution'. The grade of 'pollution' of a unit will depend on the size and form of the unit. A possible error of 130 meters in relative position would result in a pollution of 25 % for a land use zone corresponding to a square kilometer. Therefore, a square kilometre was set as the minimum required size for the land use zone, in order to restrict the maximum pollution to 25 percent. Increase in size of the LUZ reduces the percentage of pollution. The more irregular the shape, the higher the percentage of pollution.

The formula presented in the methods section is used to estimate the effect of the land cover classification accuracy on the assessment of the land cover class percentage. The confidence interval has been calculated for a number of fictitious cases, to obtain an indication of its general size.

In view of the classification accuracies (presented in Table 5.2) and the results of the contingency analysis (see Chapter 3), we took 0.8 as the average probability of correct classification (the probability that class 'A' indeed occurs if the pixel is classified as 'A'). And we took 0.02 as the probability of the occurrence of cover class 'A', if the pixel is classified as other than 'A'.

Given a total population of 100 pixels, of which 30 have been classified as 'A', the probability for each number of occurrences of class 'A' (r) can be calculated. (For example, the probability of class 'A' occurring 20 times is 0.016.) Based on these results, the confidence region is determined. In this case a level of significance of 0.95 would would correspond to a confidence region where 22 (occurrences) is the lower boundary and 30 the upper boundary, corresponding to a range of 8 percent.

The same situation, but with 70 percent of the units classified as 'A', would result in a confidence region of 12 percent. It indicates that the confidence region will generally vary between a 8 and 12 percent for the major land cover classes. The general accuracy of the land cover class percentage was taken to be 10 percent.

3.2 A threshold as criterion for accepting the appropriate cluster solution

These effects of inaccuracy add up to a critical distance in land cover composition between LUZs of about 45 percent (using the city block distance measure). (Because the LCC composition is given in percent the distance between LUZs is also expressed in percent.) Of this critical distance, 25 percent corresponds to the maximum allowed pollution as a consequence of geometric imprecision and 20 percent corresponds to the maximum distance as consequence of the inaccuracy of the land cover classification. A distance of 20 percent is ascribed to the land cover classification inaccuracy because an error of 10 percent in the percentage of one LCC implies an additional error of 10 percent in the remainder of the LCC's of the LUZ. (Together the cover class percentages of a LUZ add up to 100 percent.)

This means that for a distance less than 45, the LUZs may not be assumed to represent different characteristics with respect to their land cover composition. Differences less than 45 are related to noise. They are an effect of geometrical inaccuracy and limited reliability of the land cover classification. This corresponds nicely to the distance found between LUZs representing the same object class. For example, a maximum distance of 47 percent was found, considering all possible pairs of LUZs corresponding to banana plantations. The critical distance of 45, therefore, resulted in the grouping of all the banana plantations with corresponding characteristics. Two banana plantations showed clearly deviating percentages for the area that was classified as banana. These discrepancies corresponded to deviant characteristics in the field (with respect to plant variety, management and production level; see Chapter 3 and 4 of Part 1). They were not included in the resulting cluster of banana plantations (see Table 5.3, the composite land cover classes BAN1 and BAN2).

A threshold value of 45 resulted in 41 clusters, on which basis 41 composite land cover classes were defined. The resulting clusters denoted clear differences in land cover composition. Not only did the clusters reflect obvious differences in land cover composition, for example between banana plantations and other land use patterns, but also smaller, but significant, differences were indicated. For instance, between the banana plantations and between LUZs with mixed land cover (compare WAFORP2, WAFORPB, FORWAPAS, FORWA1, etc. in Table 5.3).

The relevance of the composite land cover classes with respect to the land use inventory is evaluated in Chapter 3 of Part 1. Table 5.3 shows the 20 cluster solution. It is to be considered a generalization of the 41 cluster solution (for example, all LUZs with a forest cover of more than 50 percent were grouped). It serves to illustrate the distinction obtained in land cover composition between the LUZs.

The data in Table 5.3 indicate that the percentages of five cover classes were generally sufficient to distinguish between most of the composite cover classes. These correspond to the following cover classes:

- Banana (BAN);
- Forest (FOR);
- Wooded area (WA);
- Grassland (PAS);
- Bare soil and built-up area (BB).

In some cases the percentage of other cover classes was used to define a composite cover class. This was done, for example, in the case of bamboo plantations and some composite classes in which secondary vegetation occupied a considerable part of the area.

Often only one criterion was sufficient to define the composite cover class. For example, the percentage of the LUZ classified as banana allowed us to distinguish between both the banana cover classes and the other composite cover classes and between the two composite banana cover classes. To differentiate between the composite forest cover classes, only the percentage of the forest cover and the percentage of wooded area was required.

Table 5.3 *Percentages of the land cover classes for the composite land cover classes corresponding to the 20 cluster solution.*

Class name[4]	Forest	Wooded area[1]	Pasture	Grassland[2]	Bare soil & BA[3]	Banana	Bamboo
BAN 1						> 60	
BAN 2						45 - 60	
WA	< 20	> 40			< 20		
WAFORP2	15 - 35	30 - 50		10 - 20			
WAFORPB	15 - 35	20 - 45	10 - 20		10 - 20		
WAPAS		30 - 50	10 - 20	10 - 30	10 - 20		
PASWA		20 - 40	25 - 35		8 - 20		
PAS1			> 60				
PAS2		< 30	40 - 60				
BBWAPAS		20 - 35	15 - 30		20 - 45		
BBWA		(WA+SV) > 30			20 - 35		
BB1					40 - 60		
BB2					> 60		
BBFOR	> 15				> 30		
PASBB			> 35		15 - 30		
BAM							> 15
FORWAPAS	20 - 35	15 - 20		10 - 25			
FORWA1	30 - 50	20 - 40					
FORWA2	35 - 50	< 20					
FOR	> 50						

[1] Wooded area (WA) denotes an area with a high density of trees, whether they are tree crops, tree plantations or wooded vegetation along river courses, for example.

[2] Grassland here represents neglected pastures, invaded with shrubs and grass vegetation up to 3.50 meters in height, or the grass vegetation encountered in inundated areas, bottom lands or otherwise poorly drained areas.

[3] BA denotes Built-up area. Bare soil and built-up area is defined as one cover class.

[4] Names are abbreviations; BAN stands for banana, WA for wooded area, PAS or P for pasture, BB for bare soil and built-up area, BAM for bamboo, FOR for forest and SV for secondary vegetation.

4. DISCUSSION

The classification of the LUZs with respect to the land cover composition can be described mathematically by 'x $\in$ S' (Klir and Folger, 1988; Molenaar, 1991). 'X' in this case corresponds to the land use zone and its associated value for the composite land cover. The subset 'S' corresponds to the composite land cover class. For the classes 'S' we can define the following requirements, analogous to those defined by Swain and Davis (1978) for classification systems (see also Chapter 3, Part 2):

- The list of classes must be exhaustive.
- The classes must be distinctive.
- The classes must contain informational value.

The second requirement signifies that the members of the same class should form homogeneous subsets, while members of different classes should form heterogeneous groups with respect to the relevant properties. A too narrow definition of classes in comparison to the variation of the attribute values will result in a low reliability of the classification result. A too broad definition of the classes will result in loss of information upon classification.

The method presented in this chapter aims to fulfill the first two requirements. It is a quantitative approach to understanding the variation in land cover composition and the distinctions between LUZs made on that basis. In our case the classification concerns composite objects characterized by a multi-dimensional feature.

The method represents a data-driven approach to the classification of land cover and land use. (The steps are illustrated in figure 5.2). This means that the classes are not a-priori defined. Instead, they depend on the accuracy with which the object characteristics can be determined (the 'x' in the expression above). That, in turn, depends on the methods applied and the quality of the data sources used to describe the objects. Errors in topography and classification can be evaluated or estimated. Their effect on the determination of the cover percentages can be assessed and accounted for in the definition of the composite cover classes. As such, a certain degree of accuracy and reliability of the classification is assured. The composite cover classes represent the appropriate level of generalization.

The data-driven approach also means that the objects are described and classified in terms of characteristics of the material used (e.g. size of faces and composite land cover). A final step is needed in which semantic content is ascribed to the classes defined (Fig. 5.2). This is called the mapping of the data classes (e.g. composite land cover classes) to the information categories (e.g. land use pattern). This phase in the inventory process refers to the mapping of the model entities to the real world. It is the subject of Chapter 3, Part 1. In this final phase, the relevance of the distinction between the LUZs is evaluated. The translation of the data classes is an additional source of error and should be evaluated independently.

The classification process is thus divided in two parts:

1. The first part concerns the definition of the data classes and the classification of the LUZs in terms of these classes. This is the more quantitative part, in which sources of errorcan be assessed quantitatively and accounted for.
2. The second part concerns the definition of the relevant information categories and the mapping of the data classes to these information categories. It involves context-dependent reasoning, which is often difficult to express in formal and quantitative terms.

A different approach to handling error and uncertainty is associated with each part. Much has been written on the reliability of information (e.g. Richards, 1986, with respect to remote sensing) and many authors have dealt with the problem of error in spatial data(bases) (Openshaw, 1989; Chrisman, 1989; Heuvelink *et al.*, 1989). Much of this work concerns the description of error and uncertainty (purity measure and other, Marsman and De Gruijter, 1986; Bregt *et al.*, 1987).

This chapter deals not as much with the description of error and uncertainty. Rather it adresses the question of how the error and uncertainty pertaining to the basic data affect the composite object characteristics and henceforth the classification of the aggregated objects. Errors are associated with all steps in the classification process (e.g. registration and classification, aggregation, see Fig. 5.2).

The aggregated nature of the object allows for a probabilistic approach in handling error

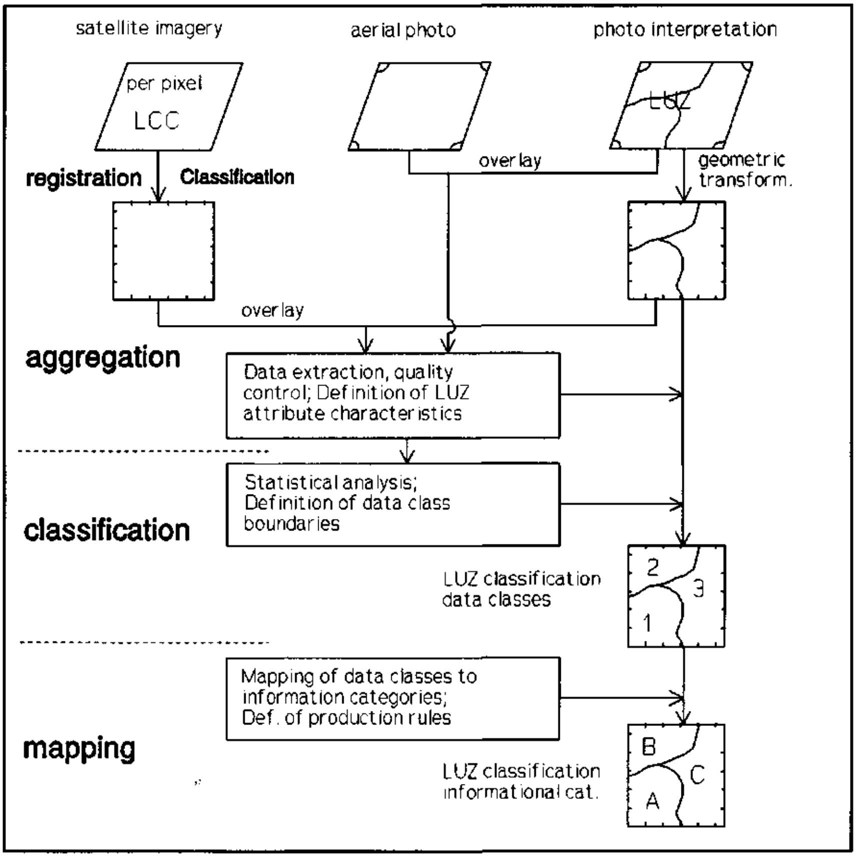


Fig. 5.2 Scheme of the Land Use Zone Classification process.

and uncertainty (both this chapter and Chapter 4 are examples). This is because each object can be regarded as a subset of the total set of picture elements (the total pixel population) or of other elementary objects (agricultural fields). When the error and uncertainty are described by a probabilistic model, quantitative (statistical) criteria can be defined for the class definition. There is always a trade-off between accuracy and relevance. This is demonstrated by Ioka (1986), who shows the monotonous increase of classification accuracy with decreasing number of classes (corresponding to levels of generalization). The method presented in this chapter aims to provide a solution with respect to the selection of the appropriate level of generalization, as far as the data classes are concerned.

As a tool for handling uncertainty, evidence reasoning (see Lee, 1987; Srinivasan, 1990; Kenk, 1988 for application hereof in the field of remote sensing) suffers from subjectivity (Middelkoop *et al.*, 1989). A probabilistic approach is therefore preferred whenever possible. The application of evidence reasoning is considered more appropriate with respect to this last phase of the classification process. At this point context-dependent reasoning is applied for the mapping of the data classes to the information categories.

The data-driven approach will provide the appropriate land use information, given the scale and quality of the material used. The land use information will be appropriate assuming that land cover and spatial characteristics reflect relevant differences in land use and provided that adequate methods for information extraction are applied. The adoption of a data-driven approach might be suitable if materials are not widely available and when exact information requirements are not known. The data-driven approach then assures reliable information on land cover and land use. Further use of this information will then depend on the level of generalization it represents. The data-driven approach is the strategic answer to the classification problem of land cover and land use in the Atlantic Zone of Costa Rica.

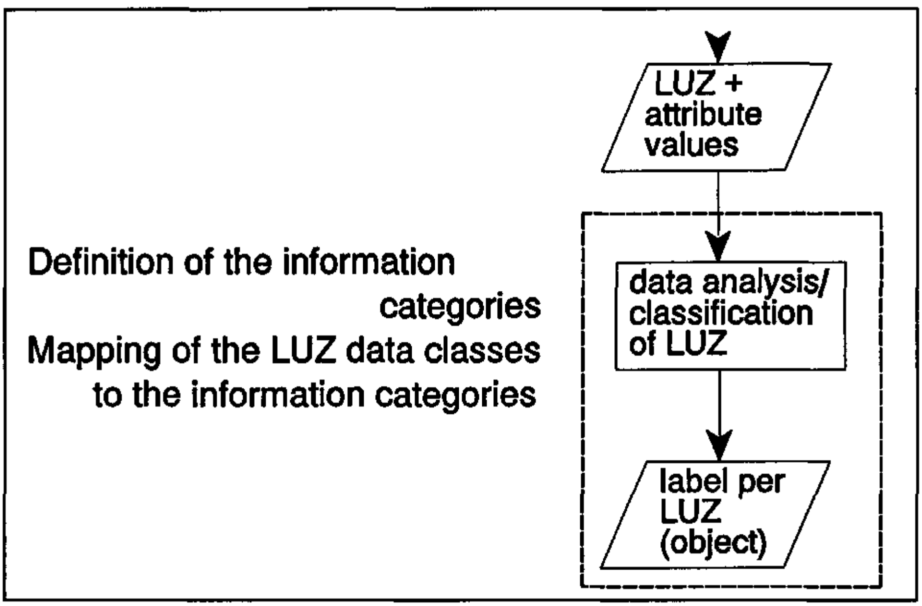
5. CONCLUSIONS

1. Hierarchical cluster analyses provides a quantitative method for the definition of distinct classes to describe characteristic compositions of objects, that consist of elements with each a specific nominal value (e.g. class value). It fulfills the requirement that the set of classes must be exhaustive,

2. The appropriate level of generalization of class definition, can be ascertained by taking several factors into account. One of these is the geometric accuracy of the material used to extract data for the description of the objects. Maximum alowed error in determining the object characteristics as consequence of geometric inaccuracy is translated into minimum size required for the objects. The other is the quality of the input data used to characterize the objects (accuracy of the per pixel land cover classification). These factors are pertinent to the definition of a threshold value to indicate distinction or no distinction between objects.

CHAPTER 6

A DECISION TREE FOR THE CLASSIFICATION OF LAND USE ZONES, REPRESENTING COMPLEX SPATIAL OBJECTS



1. INTRODUCTION

The land use zones (LUZs), corresponding to evident spatial patterns, are assumed to represent relevant geographic units as carriers of land use information. The LUZ displays a number of characteristics (attributes). These are described in the previous chapters. Chapter 4 and 5 gvie the class definition for two important characteristics. The resulting classes reflect the degree of discrimination between LUZs, that can be obtained on the basis of the single object attribute. This might correspond to a multi-dimensional feature, such as composite land cover. In this chapter, many object characteristics are considered for the classification of the LUZs. The LUZ classification is a multi-source classification problem (e.g. the LUZ map, aerial photo characteristics and land cover data derived from satellite imagery).

Aside from its multi-source aspects, the classification has to consider the semantic content of its geographic units and its classes. The data classes alone do not provide the required land use information. Accordingly, the data classes have to be mapped to the information categories that do provide relevant land use information. This involves the interpretation of the data with respect to land use. And that interpretation is context-dependent. An example of this mapping procedure was presented in Chapter 4. It refers to the interpretation of size of faces (through field size) relative to farm size. This relation is specific for the region and only valid under certain conditions (for specific LUZs).

Two aspects of the classification of the LUZs should be noted:

1. This is a complex classification problem, involving many object characteristics (multi-source or multi-feature).
2. It involves context-dependent decisions, as mapping involves the interpretation of the object characteristics for land use information.

The multi-source classification problem is the subject of many studies concerning the integration of remote sensing and geographic information systems. With respect to the multi-source analytical procedure, Lee *et al.* (1987) mention, the need for a method by which inferences about information classes can be drawn from the collection of data classes. Wilkinson and Mégier (1988) propose contextual rules, which might refer to likelihood of higher level entities in a taxonomy, rather than just the lowest level classes. For a discussion of the application of spectral and spatial rules in a hierarchical classification approach see Ton *et al.* (1991). Hierarchical classifiers or decision tree classifiers have long been recognised as a powerful classification tools (Swain and Hauska, 1977).

This chapter reviews the possibilities for application of a hierarchical decision model. In that model the attributes are evaluated at different levels of the classification hierarchy. There are two main questions: how to structure such a complex classification problem, and how to structure the use of context- dependent knowledge. The structure has to enable the transfer of the keys for the interpretation of the data (classes). The use of contextual evidence requires the specification of the context in spatial and thematic terms. And the associated rules must be specified as well. The context should be understood in the sense of a semantic context determined by way of object relations. However, given the geographic nature of the objects, a spatial dimension is associated with the contextual information.

Most of the studies mentioned concern the per pixel classification. They employ contextual evidence to improve reliability of the class label assignment. The LUZ classification refers to another level of aggregation. Yet, context information is equally relevant to the classification of the LUZ. The regional context, e.g. whether the LUZ belongs to the agricultural area or to the area of agricultural penetration, determines the interpretation of the land cover class (LCC) composition of the LUZ relative to the LUZ land use characteristics.

2. CONTEXT INFORMATION

With respect to the LUZ classification, the use of geographic context information is used to improve the interpretation of object features in order to obtain reliable land use information. The context rules encode facts relating land use to a certain geographic context. The geographic context itself refers to an area with certain thematic characteristics. In the present case, these provide information on the occurrence of land utilization types or farming systems. The geographic and thematic characteristics of the context need to be specified. Then the context can be identified and the context rules can be triggered.

Let us define a set A, representing the data classes and attribute values, and a set B, representing the information categories. Then the mapping of data classes to the information categories can be written as the mapping from A into B. This is written as:

$$f:A \text{-->} B$$

In this expression A is called the domain of the mapping and B the co-domain. The context can be seen as specifying a subset of B or limiting the range of f. This means that the contextual knowledge can be described in terms of spatial objects with geometric and thematic characteristics and the associated image of f.

LUZs represent a relevant context with respect to the mapping of the size of faces classes to the farm size classes, as well to the interpretation of the spectral classes in terms of specific land cover classes and to interpretation of the composite land cover classes relative to the land use patterns (LUPs). Land cover composition may provide important evidence for the interpretation of the size of faces. For instance, land cover information might indicate the cultivation of ornamental crops; consequently, no relation between field size and farm size would be expected. On the other hand, the information on field size may provide evidence for the occurrence of a specific land cover type. For example, cacao is predominantly grown on small and medium-sized farms. The field size, as an indicator of farm size, might provide evidence for the mapping of the cover class 'wooded area' to the specific land cover type 'cacao'. Also land use information might be inferred from the percentage of land the LCC occupies. For example, very high percentages for 'wooded area' indicate the presence of tree crops (cacao, macadamia or other), if the LUZ belongs to the agricultural region. In contrast smaller proportions of the LUZs occupied by 'wooded area' are generally attributed to wooded river bank, homesteads, wooded pastures and other.

Two problems emerge in using context information:

- One concerns the definition of the context rules itself. This is treated in Chapter 3 of Part 1.
- The other concerns the design of the system that triggers the rules. This is the subject of current chapter.

Field size may aid in the interpretation of LCC and, vice versa, land cover might aid in the interpretation of field size. Therefore, it is important to define rules for the classification at the different levels in the classification hierarchy (defining backward loops which allow for a stepwise refinement of the classification). Doing so will prevent circular reasoning.

3. THE TREE CLASSIFIER

The production rules (or decision rules) for the label assignment have the following structure:

IF {condition} THEN {conclusion} fi

Such rules have been widely applied in the field of remote sensing (Desachy, 1990; Wu *et al.*, 1988). Janssen (1990), for instance, assigns agricultural field objects a land cover label on the basis of a majority rule. He reports a substantial increase in classification accuracy.

In the LUZ classification, the rules are of a more complex nature. A set of conditions have to be fulfilled before a class label is assigned. The conditions may refer to characteristics associated with the land cover classification. Otherwise, they refer to the data obtained by aerial photo interpretation and sampling of the aerial photographs. These conditions concern:

- land cover class percentages;
- type of LUZ boundary;
- pattern of faces;
- size of the faces;
- type of boundary of the faces;
- presence of trees dispersed over the area.

Several rules are applied when assigning a class label. The application of the rules is hierarchically structured, corresponding to the classification hierarchies (taxonomic levels). Wilkinson and Mégier (1988) describe the use of production rules in a hierarchical classification system. The procedure proposed in this chapter is a step-wise specialization of the object class. Theclass value at a specific hierarchical level (the superclass) defines the context for the classification step on the next lower level. It allows for the application of context specific (object class dependent) production rules. The different types of objects have specific attribute structures and therefore require different production rules for their classification. For instance, for the further sub-division of banana plantation other decision rules apply (they only consider the percentage of banana cover) than for the further subdivision of forested area (for which the subdominant LCC is considered). The class structure of objects can be schematically represented as in Figure 6.1 (Molenaar, 1989). The classification procedure is illustrated in figure 6.2.

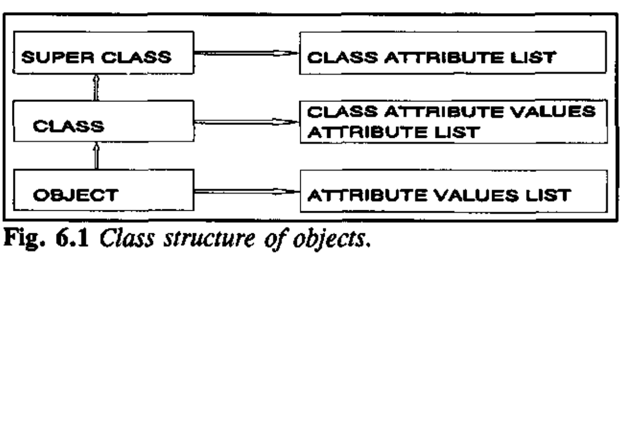


Fig. 6.1 *Class structure of objects.*

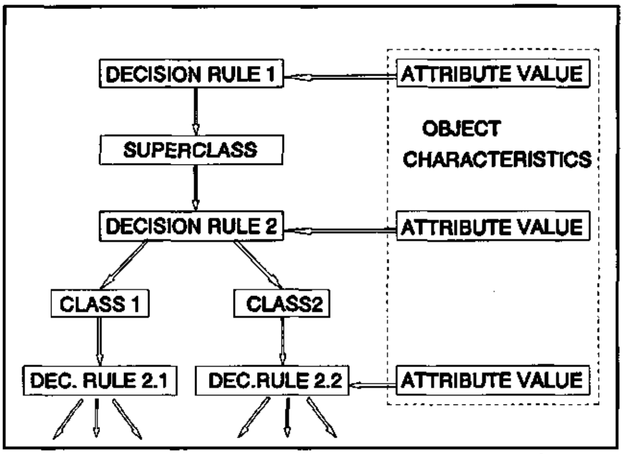


Fig. 6.2 *Structure of the classification tree as related to class hierarchies.*

4. A DECISION TREE FOR THE CLASSIFICATION OF LUZS: EXAMPLES.

This section introduces the decision tree as applied to the classification of LUZs. At the highest level, a distinction is made between LUZs corresponding to water bodies, artificial constructions, and the remaining LUZs. Information on water bodies, towns and villages can be obtained from maps. In this case, it is derived from the aerial photo.

Decision rule A1:
IF
('presence of faces' equals 'no') OR ('changes' equals 'gradual') OR
('LUZ boundary' not equals 'rectilinear')
THEN
('LUZ label 1' equals 'Natural and Semi-natural vegetation').

Water bodies and cities or villages are easily recognized on the aerial photos scale at a scale of 1:80,000. For the remaining LUZs, a distinction is made between areas with natural and semi-natural vegetation and the (partly) cultivated areas. The decision is based on spatial characteristics. These concern: the homogeneity of the LUZ, the presence of faces and, if present, the type of transition and type of LUZ boundary. (See decision rule A1.)

4.1. Natural and semi-natural vegetation

Figure 6.3 depicts the decision tree for the classification of the LUZs with natural and semi-natural vegetation. The decision at the highest level (A1) has just been discussed. At the next level (A2 and A3) the dominant LCC is the criterion for class label assignment. This attribute distinguishes between the forested areas and the natural grass vegetation in and around the lagoons in the north of the Atlantic Zone. The predominant LCC was determined on the basis of the land cover percentages. This was calculated with the ILWIS software[1].

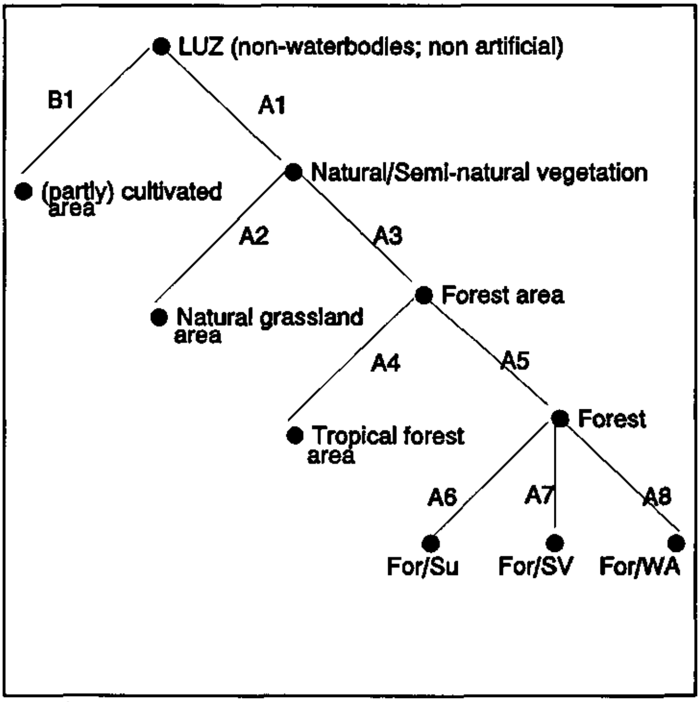


Fig. 6.3 *Decision tree for the classification of natural and semi-natural vegetation.*

The natural grassland area can be distinguished from other grassland area exhibiting the same vegetation type and structure and showing similar spectral response. This is possible because of the different context as defined by the spatial characteristics. The second decision rule is given on the next page.
Pas2 is the name for the LCC characterized by grass vegetation up to 3 meters in height, often with presence of shrubs or herbs. The other natural and semi-natural vegetation areas all show forest as the dominant LCC. Subdivision is done according to the percentage of the area classified as forest.

1. With more than 75 percent forest (the threshold value being based on the results of the hierarchical cluster analysis, see Chapter 5), the larger forested areas in the coastal plain and on the slopes and footslopes of the volcanos are differentiated and labelled as 'lowland and submontane humid tropical forest'.

---

[1] Integrated Land and Watershed Management Information System, developed at the International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede, The Netherlands.

Decision rule A2:

Given the classification of the LUZ at the first level of classification as 'n-atural and semi-natural vegetation' the following decision rule (A2) applies

IF

('dominant LCC' equals 'Pas2')

THEN

('LUZ label 2' equals 'area with natural grass-land vegetation with or without open water').

2. Areas where 50 to 75 percent of the land cover is classified as forest were further differentiated. The subdivision was based on the value for the sub-dominant LCC (see decision rule A8).
   A. The entire entire coastal area was mapped as 'lowland tropical humid forest incorporating peat and swamp area'. That decision was based on 'swamp vegetation' as the sub-dominant cover class. This is in agreement with the many swampy parts and peat areas actually present.
   B. When 'wooded area' represents the sub-dominant cover class, the smaller forested areas within the agricultural regions are identified. The 'wooded area' percentage might be (partly) explained by the boundary effects. That is, it could be a result of mixed pixels at the forest limit, classified as 'wooded area'. 'Wooded area' in this case should not be construed as a different vegetation structure; rather it is related to the size of the forested areas.
3. The area with parts classified as secondary vegetation corresponds to forest area of which parts have been cleared in the past and on which secondary regrowth has occurred. It refers to only one LUZ.

IF the LUZ belongs to an area with natural and semi- natural vegetation and IF the dominant land cover equals 'forest' and IF 50 to 75 % of the area is classified as forest, THEN:

Decision rule A8:

IF

('sub-dominant LCC' equals 'wooded area')

THEN

('LUZ label 4' equals 'remnant forest area, disturbed and partly cut forest')

4.2 Plantations

Within the (partly) cultivated areas, four categories are differentiated:

1. Wastelands, identified by the absence of faces and the presence of wooded area (WA)

and forest (For) in conjunction with secondary vegetation (SV) or secondary (abandoned) grassland (Pas2);
2. Partly cultivated lands corresponding to areas of agricultural penetration, exhibiting a characteristic pattern of deforested patches. Forest covers more than thirty per cent of the area;
3. Plantations, showing a homogeneous cover of the LUZ, with medium to coarse texture (referring to photo characteristics);
4. Area in agricultural use, where certain conditions, such as the presence of faces, should be fulfilled.

The structure of the decision tree is given in Figure 6.4.

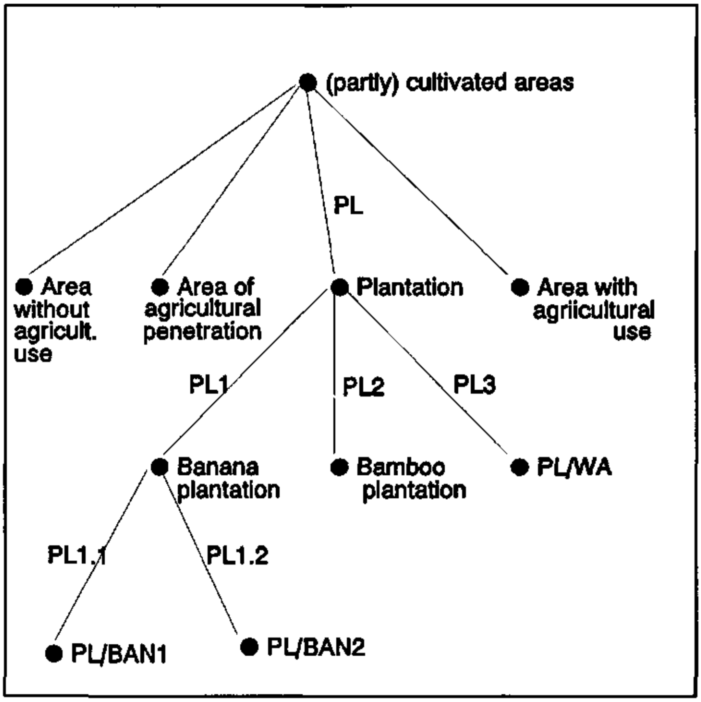


Fig. 6.4 *Decision tree for the classifcation of plantations.*

The plantations are subsequently differentiated according to their dominant land cover. This can have the value of:

1. Wooded Area. In the land cover classification, no distinction could be made between the different use and cover types that involved a high density of trees. However, within the context of 'plantations', the cover class 'wooded area' can only correspond to cover types like Macadamia (*Macadamia integrifolia*), citrus or cacao. Yet, of the latter two, no larger plantations were encountered in the area under consideration.
2. Bamboo, corresponding to a few large groves that exist in the zone;
3. Banana, referring to the many banana plantations present in the zone.

Based on the result of the hierarchical cluster analysis (Chapter 5), two classes of banana plantations were defined. One class correspondings to more than 60 percent of the LUZ being classified as banana. The other class corresponds to between 40 and 60 percent of the area classified as banana. Two plantations belonged to the latter class. Many pixels within both plantations were classified as secondary vegetation.

Both plantations showed rather deviant characteristics as regards the production of bananas. One of the plantations (the finca Santa Maria) had the lowest production in boxes per hectare per year in the northern Atlantic Zone in 1986 (corresponding to the year of the satellite image on which the land cover classification was based). Its production amounted to 1494 boxes/ha./year compared to an average production in the zone of 2122 boxes/ha./year[2]. A detailed study was carried out in this plantation to explain variation in yields by spectral reflectance data and soil data (Veldkamp *et al.*, 1990). The other plantation (one of the three plantations belonging to the 'Hacienda Bremen', also showed a rather low production (1769 boxes/ha./year). Furthermore it was traversed by roads, to transport the fruits. At the other plantations, the fruit is transported by cable, allowing a much more efficient use of the area. Also, Hacienda Bremen was one of the few plantations at that time to cultivate a dwarf variety of banana (gran nain). used. In case of Hacienda Bremen the combination of these characteristics explains the low percentage of the plantation area classified as banana. The examples illustrate that a low percentage of the plantation area being classified as 'banana' can be an expression of different phenomena. This should be taken into account in interpreting the data classes. It illustrates also the need for separate classifications phases: one for the assignment of data class labels and one for the mapping of the data classes to the information categories.

IF the LUZ belongs to the 'cultivated area', and IF the area is classified as 'plantation', and IF the dominant cover class equals 'Banana', THEN the following decision rule applies:

Decision rule PL1.2:

IF ('%ban' larger than '60')
THEN ('LUZ label 4' equals 'PL/BAN1')
ELSE
IF ('%ban' between '40 and 60')
THEN ('LUZ label 4' equals 'PL/BAN2')
ELSE ('error')

The two composite land cover classes for banana plantations are defined on the basis of results of the hierarchical cluster analysis. The classes prove to provide relevant distinction between the banana plantations. However, this distinction does not refer to the actual percentage of the plantation in use for the cultivation of banana (between 40 to 60 percent and respectively more than 60 percent). The low percentages of the area classified as banana

[2] Source: ASociación BAnanera NAcional (ASBANA) of Costa Rica.

might be seen as the result of poor management (as in the case of Santa Maria) or deviating varieties (as is the case for Hacienda Bremen).

4.3 Classification of agricultural areas

Figure 6.5 presents part of the decision tree for the classification of agricultural areas.

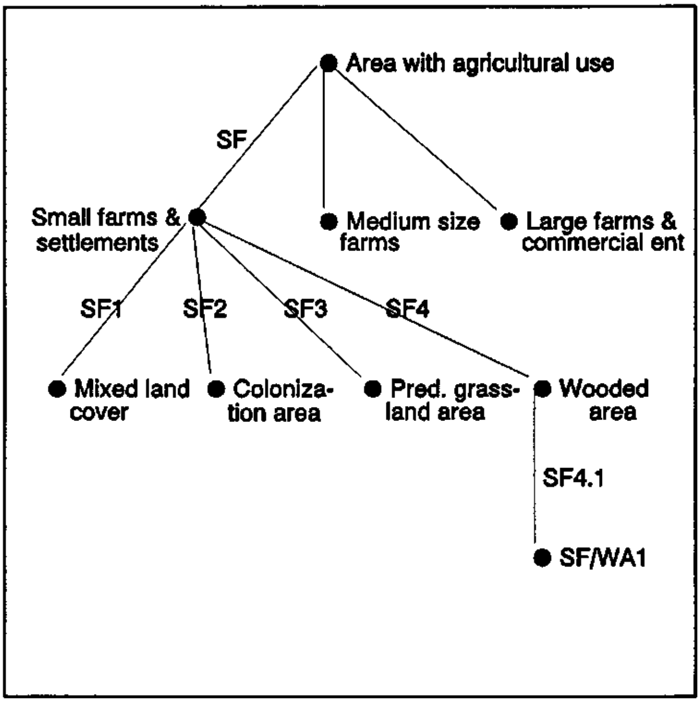


Fig. 6.5 *Decision tree for the classification of LUZs in agricultural areas, with only the branch of small farms and settlements fully described.*

The first distinction between the LUZs is based on the size of the faces. This allows for the distinction on mean farm size between LUZs. The rule 'SF' therefore is formulated as follows:

Decision rule SF:
IF ('mean size of faces' less than '3.9')
THEN ('LUZ label x' equals 'small farms')

See Chapter 4, for the determination of the correct threshold value. Instead of using the size of the fields as the criterion also the class values can be used when LUZs have already been classified for their mean size for faces. The rule is expressed as follows:

IF ('Class size of faces' equals '1')
THEN ('LUZ label x' equals 'small farms').

The resulting classes are further subdivided, according to conditions concerning the land cover composition:

1. Small farm area with a mixed land cover (i.e. presence of the cover classes bare soil, grassland and wooded area). It indicates areas with an agricultural use involving well arable cropping (predominantly maize), animal husbandry and perennial crops. A variety of crops is generally found.
2. Areas with presence of forest (15 to 30 percent of the LUZ should be classified as such) indicate the colonization areas. In some cases the areas correspond to recent state- controlled settlement schemes.
3. Area with grassland as dominant LCC, although considerable parts are classified as 'wooded area' and to a lesser extent as 'bare soil'. Animal husbandry is the most important agricultural activity. This land use pattern is not very common among LUZs where small farms are found.
4. A few LUZs show very high percentages for 'wooded area', due to which they form a separate subclass. A high concentration of small cacao plantations is found in all these areas.

See Chapter 3 of Part 1 for a more detailed description of the land use categories. The definition of the fourth class clearly illustrates how context information can be used for the interpretation of a rather general cover class. 'Wooded area' can correspond to a number of cover types or land utilization types like macadamia, citrus and cacao. Furthermore, tree plantations (for instance, Laurel (*Cordia alliodora*)) are likely to be classified as 'wooded area'. Also wooded areas along streams and rivers are classified as such. Other densely wooded cover types like wooded grasslands and homestead gardens are identified as 'wooded area'. Therefore, in most agricultural areas the percentage of 'wooded area' can vary considerably. It sometimes covers 45 per cent of the area.

The cluster analysis revealed a few LUZs with very high percentages for 'wooded area'. The 'wooded area' class is interpreted as 'macadamia' when the LUZs represent a plantation and as 'cacao fields' when the LUZ corresponds to small farms. It does not exclude the occurrence of other cover types related to 'wooded area'. These, however, explain only part (25 to 35 percent) of the total area classified as 'wooded area'. Indeed, all three LUZs, that have small farms and more than 40 % classified as 'wooded area' showed high percentages of cacao plantations. Moreover, all the existing areas in which land use was characterized by the presence of cacao were identified by the procedure. The condition for deciding on the presence of cacao is that the areas pertain to the older agricultural regions, because the cacao plantations are often found as a remanant of former agricultural activities.

Within the context of small farms and a differentiated land use, the 'bare soil and built-up area' is interpreted as corresponding to the area for the cultivation of maize. This interpretation is based on the date of scene recording of the scene and the period of field preparation. The area used for the cultivation of maize occupies only part of the LUZs. Therefore, very high percentages of 'bare soil and built-up area' are not reached. Where high percentages are found, the cover class should be interpreted differently. For instance, it could be identifeid as recently (in comparison to the date of scene recording) deforested area, when forest occupies part of the LUZ. Or the area may be cleared for the planting of banana, which can sometimes be inferred from the spatial pattern. In all these case the high percentage 'bare soil' denotes change in land use.

5. DISCUSSION.

This chapter presents a conceptual model of a decision tree for the classification of LUZs. The decision tree is described for a number of cases. This approach makes use of quantitatively verifiable criteria for the class assignment. A computer program has been written to perform an automated classification. It improves the flexibility. New insights the regional land use can be easily incorporation and other requirements for the presentation of the data (categorization) can be easily fulfilled. It allows also for checking the consistency of the decision rules.

The appendix presents the LUZ classification map and the associated classification system for the Guacimo-Rio Jeménez-Siquirres study area is given. The reader will recognize that structure of the classification system as corresponds to the structure of the decision tree depicted here.

An important objective of the design of a decision tree for classification purposes, is to provide a structure for the transfer of the interpretation keys and the context rules. It forces the interpreter to specify the link between observed object characteristics (such as field size and land cover percentages) to real phenomena in the field.

At this stage only the satellite image-derived LUZ characteristics and aerial photo characteristics are considered in the decision model. At a later date it might be useful to incorporate information concerning neighboring LUZs as additional contextual information or ancillary data (for example, soil data) in the decision model as additional evidence.

Also in future work the aspects of error and uncertainty with respect to the mapping process deserves attention, especially because of the context dependent nature of the decision involved. In Chapter 3 of Part 1 the classification of the agricultural area is further commented upon. This represents the most complex area for the classification of LUZ. Also in that chapter the method we employ is validated to the extent that the actual land use corresponding to the categories defined is investigated.

PART III

LAND USE EVALUATION

CHAPTER 1

EVALUATING LAND USE AT SUB REGIONAL LEVEL IN THE ATLANTIC ZONE OF COSTA RICA, CONSIDERING BIOPHYSICAL LAND POTENTIALS.

E.J. Huising and W.G. Wielemaker

1. INTRODUCTION

Inadequate land use causing land degradation and empoverishment of rural populations is becoming a worldwide concern and stresses the importance of a land use policy which leads to viable and sustainable land use scenarios. The new land use code, adopted by the costarican government, is an example of such a policy. Credits will be granted to farmers only if their land falls within the correct land class (The Tico Times, 1991). The code is intended to develop a more sustainable use of land resources.

Procedures for the definition and selection of sustainable land use alternatives require an analysis of actual land use in relation to land characteristics in order to identify problem areas and to formulate viable alternatives. The present study in the Atlantic Zone of Costa Rica is an example of such a land use analysis. Soils and land use were mapped independently at a scale of 1:100.000 (reconnaissance). GIS techniques were used to process and combine the information. Such an analysis would fail if information on either land use or soils were lacking, which is often the case.

The purpose of the evaluation was to investigate to which degree agricultural use corresponds with land capability, indicating areas where land use has a degrading effect on land qualities (over-use) and also areas with opportunities for more intensive use of the land (under-use). In the evaluation only biophysical aspects are considered.

When evaluating land use with respect to land capabilities it is important that the level of aggregation of both the soil information and the land use information corresponds. E.g., evaluating yield at field level is not realistic when soil data are obtained from a reconnaissance soil map. Due to the variability of the terrain such a small scale map presents the information on soil types in an aggregated or associated form so that the map does not provide a decisive answer as to the particular soil type of the field concerned.

Which statements can be made on land use with respect to soil and land suitability depends, of course, on the accuracy and the detail of information provided by both the soil and land use map. The type of statements is strongly related to the level of aggregation (scale level, see also the levels of detail defined by Bouma, 1989). Aggregation level and its implication for the statements on land use, are therefore also a topic of discussion in this paper.

2. TERRAIN UNITS, LAND USE PATTERNS AND LAND CAPABILITY CLASSES

Soil information of the Atlantic Zone of Costa Rica

In the Atlantic Zone of Costa Rica a reconnaissance soil survey has been carried out at a scale of 1:100000 (Wielemaker and Oosterom, 1992). For the storage and retrieval of maps and associated data, the SIESTA soil information system was designed (Oosterom et al., 1992).

The data structure

The 162 mapping units are physiographic in nature and are referred to hereafter as Physiographic Mapping Units (PMU). The mapping units generally have a composite character, because within one mapping unit up to five terrain units (TUs) may occur, corresponding to the physiographic elements of the PMU. The PMU composition table gives the PMU-identifier, the TU identifiers and estimated area percentages of the terrain units for each PMU. The terrain unit table lists for all the terrain units identified, the terrain properties and the soil type. Properties refer to major and minor physiography, lithology, parent material, slope, stoniness and soil type. The soil types and associated data are stored in a separate table (SU). The data structure is illustrated below.

Table 1.1. The data base structure of the soils data base

PMU-ID	TU1-ID	TU1 %	- - -	- - -	TU5-ID	TU5 %

TU-ID	Physiogr.	- - -	Soil	Land Capability class

SU-ID	Depth	pH	- -	- -

For every terrain unit an associated land capability class is defined. The land capability classes serve to group the terrain units in a limited number of classes on the basis of the terrain unit characteristics (a data reduction process). The land capability classes are also stored in a separate table.

Land capability classes

In order to reduce the amount of data, nutrient status, depth and drainage and slope and stoniness were used to group soils in land capability classes, as listed below. The land capability classes denote general agricultural potential. It decreases from class 1 (LC1) to class 5 (LC5) as follows:

Class	Potential for agriculture
1	Deep, fertile soils, suitable for all land use types of the zone;
2	Moderately fertile soils, moderately suitable for nutrient demanding crops;
3	Nutrient poor soils, only suitable for acid tolerant and little nutrient requiring crops;
4	Shallow or sandy soils, with severely restricted use;
5	Steep (5.1), poorly drained (5.2) or soils with extreme stoniness (5.3), soils not suitable for agriculture use.

The first three classes are subdivided according to:
1. Soils with no further limitations;
2. Soils with impeded drainage (requiring artificial drainage);
3. Soils on slopes from 0 - 8 % and stony;
4. Soils on slopes from 8 - 30 %.

Class four is subdivided in soils less than 30 cm (4.1) and less than 10 cm deep (4.2). Class 5 is subdivided as indicated in the description.

The study was conducted for the GRS area. From southwest to northeast three landscape types can be distinguished:

(1) Sloping areas on the Turrialba volcano with lava and lahar deposits of andesitic composition. Higher up the volcano, volcanic ash mantles those deposits. Near the coast and in the Tortuguero national park, the Tortuguero hills occur, remnants of Tertiary to early Quaternary volcanism.

(2) Gently sloping plains at the foot of the volcano where rivers and lahar flows deposited fine grained material from the same volcano.

(3) Very gently sloping to flat areas near the coast with fine textured to peaty deposits especially further away from rivers where areas become swampy. Sandy deposits occur alongside rivers.

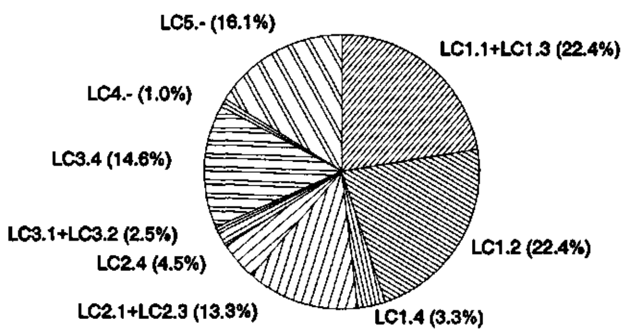


Fig. 1.1 Area percentages of land per suitability class for the study area.

The distribution of the land capability classes is illustrated in Figure 1.1. Fertile soils (class 1) dominate with an area percentage of 48, but half of it suffers of impeded drainage (class 1.2), particularly in the lower part of landscape 2 in transition to landscape 3 near the coast. Fertility status as reflected by classes 1 to 3, is directly linked to the age of the deposits which varies greatly particularly in landscapes 1 and 2. The younger deposits there are covered with nutrient-rich and non-acid soils, but older deposits are covered with nutrient-poor and acid soils (Wielemaker and Oosterom, 1992). The last ones occupy in landscape 2 the sloping remnants of older lahar deposits. The combination with strong slope (class 5.1) makes them unsuitable for agricultural purposes, because erosion is too great a risk just as on the steeper parts of the Turrialba Volcano in landscape 1 and the Tortuguero volcano near the coast. Younger lava and lahar deposits on or near the Turrialba volcano can be extremely stony which makes agriculture impractical (class 5.3); further away from the volcano some soils on very recent deposits are so shallow and sandy (class 4) that agricultural production is marginal due to droughtiness and lack of foothold for the plants.

The swamps and peat areas of landscape 3 near the coast should be protected because the peat and clay soils have no bearing capacity and would degrade on reclamation (class 5.2).

Land use information.

For the mapping of land use at reconnaissance level, Land Use Zones (LUZ) are defined. The land use zones represent geographical units, which are classified on the basis of characteristics like form and shape of fields and land cover composition. Table 1.2 presents a generalized version of the classification system.

Table 1.2. The Atlantic Zone land use classification system.

1. NATURAL VEGETATION
   1. *Forest land* (FOREST)
      Lowland and premontain humid tropical forest.
   2. *Natural grassland vegetation.*
2. AREA OF AGRICULTURAL PENETRATION (AP/GRASS)
   1. *Small farms*; forest covering more than 30 % of the area.
   2. *Medium sized and large farms*; forest covers more then 50 % of the area.
3. AGRICULTURAL LAND (AL)
   1. *Small farms*
      1. Mixed land cover
         1. Homesteads.
         2. Annual crops (predominantly maize) in combination with perennial crops and grassland. (AL/MIXED1)
         3. Perennial crops (fruit) and livestock, less annual crops. (AL/MIXED2)
      2. Mixed land cover and forest (AP/MIXED)
         Perennial crops and cattle breeding, some arable cropping.
      3. Predominant grassland area (AL/GRASS)
         Grazing in combination with tree crops.
      4. 'Wooded area' as dominant land cover type.(AL/SFOV)
         Cacao, in cases with livestock production and cropping for home consumption.
      5. With 'Bare Soil' as dominant cover type.(AL/SFOV)
         Annual and perennial crops (maize, beans, cacao, etc.) or recently deforested area.
   2. *Medium sized farms*
      1. Mixed land cover(AL/MIXED3)
         Livestock production and arable cropping (maize) in cases perennial crops and timber production.
      2. Predominant grassland cover. (AL/GRASS)
         Cattle breeding, in cases perennial and annual cropping.
   3. *Large farms*
      1. Grassland cover(AL/GRASS)
         Cattle breeding, different types of pastures occurring (improved, degraded)
      2. With 'Bare soil' cover.
         1. Arable crop production.
         2. Ornamental crop production.
      3. With forest cover.
4. PLANTATIONS (PL)
   1. *Banana plantations* (PL/BAN)
   2. *Macadamia plantation.*
   3. *Tree plantation/Reforestation*
   4. *Bamboo.*
   5. *Other.*
5. TOWNS AND VILLAGES
6. WETLAND
   1. *With forest cover*, peat and swamp areas, with high concentration of yolillo and other palm vegetation.
   2. *Mixed cover*; Wooded area, cultivated grassland and bare soil. Areas for parts of the year inundated and very poorly drained areas with yolillo and grass vegetation.
7. OTHER, WASTELAND. (WL)
   1. *Forest and wooded area.*
   2. *Mixed land cover composition.*
      Wooded area, non-cultivated and/or secondary vegetation, in cases some cultivated grassland or bare soil.

The 'area of Agricultural Penetration'(AP) represents areas, partly forested and characterized by a process of agricultural expansion: The colonization areas. The colonization of the Atlantic Zone started at the end of the last century but gained momentum especially in the last three decades (Waaijenberg, 1990; Hall, 1984).

With regard to agricultural land (category 3), land use is described by the Land Use Pattern. The land use pattern reflects characteristics of the occurring farming system(s), such as farm size, most important crops, presence of livestock and other. It is in fact a specific combination of land utilization types, occurring within a particular land use zone (the land utilization types themselves refer to a lower level of aggregation). Considering the mapping scale of 1:100.000, individual farms (except for the larger ones), let alone fields, cannot be mapped. The fact that land use zones are represented by specific land use patterns is explained by the occurrence of farms of a specific size class and by their agricultural history.

For evaluation of land use at reconnaissance level we refer to the land use pattern. The recognition of different aggregation levels implies the definition of a hierarchy in agricultural systems (Fresco *et al.*, 1990) as is also reflected in the data structure, presented in Table 1.3.

Table 1.3. *The structure of the land use data.*

LUZ identifier	Land cover composition	Average field size	- -	- -	Land Use Pattern

LUZ identifier	Farm identifier

Farm-Id	Farm size	Capital items	Labour	- -	- -	Farming system

Crop type	Farm Id	Area	Inputs	- -	Cropping system

Husbandry type	Farm_Id	Herd composition	- -	Livestock system

Figure 1.2 shows that the areas corresponding to the land use categories are not equally distributed over the study area. Agricultural land (AL), excluding the plantations, occupies slightly more than 50 percent of the area, equally spread between areas dedicated purely to livestock production (AL/GRASS) and areas in which we find a combination of arable cropping, tree cropping and pastures for grazing (AL/MIXED). Banana plantations occupy only a small part of the area, but banana production is most important in economic terms.

The land use zone map also provides information on the distribution of farm size, but only for the agricultural areas (Fig. 1.3), because for the area of agricultural penetration (AP) and the forest area no indication of the size of the properties could be obtained. In nineteen percent of the area very small farms occur, corresponding mostly to (former) settlement schemes or areas distributed by the government. The farm size in this class ranges from 10 to 15 hectares with a maximum of 29 hectares. The category small farms comprises farms up to 60 hectares. The large farms are generally a few hundred hectares in size and sometimes up to a few thousand hectares. Noteworthy is the high percentage of the area corresponding to large farms. It indicates the skewed property distribution pattern.

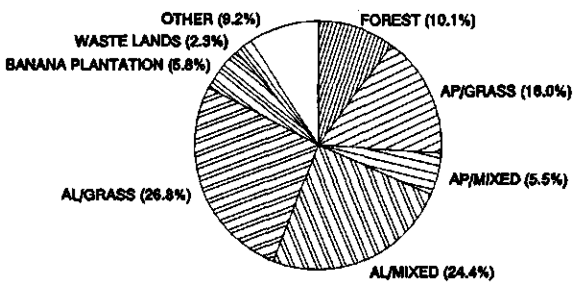


Fig. 1.2 Area percentage of land per land use category for the study area.

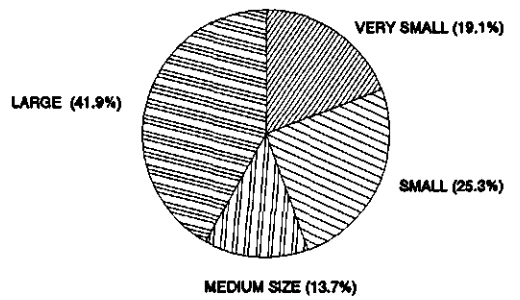


Fig. 1.3 Are precentage of farm size classes of the agricultural area.

3. METHODS FOR COMPARING LAND AND LAND USE

Combining soil and land use information

Overlaying the soil and land use map

The overlaying (in ARC-INFO) of the land use zone map and the soil map resulted in nearly 1000 mapping units with each a specific LUP-PMU combination, despite the specification of a rather high tolerance value of 100 meters for the union of line elements. (Fig. 1.4 shows the result of the overlay for a part of the area.) The large number seems therefore, to a large extent, due to the mapping procedures: on the soil map the river valleys are mapped to demonstrate the drainage pattern, while on the land use map the river valleys are drawn as lines, because their surface is small and the information considered irrelevant.

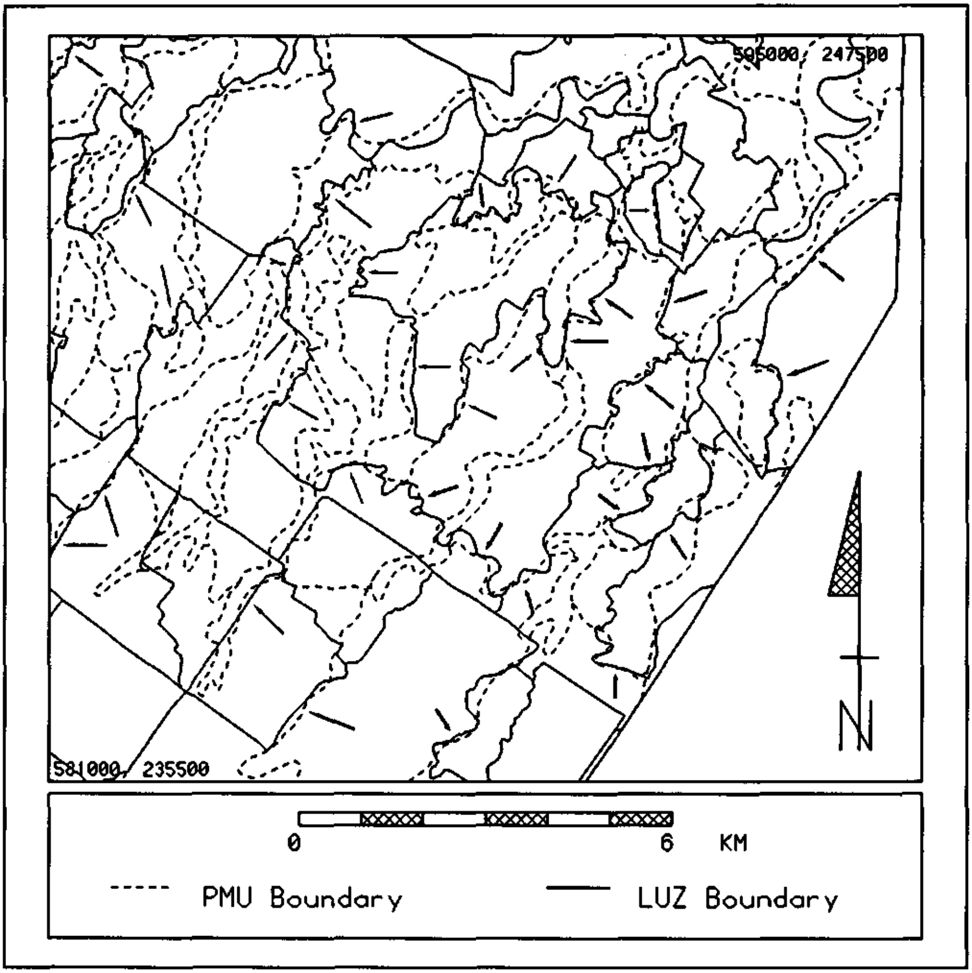


Fig. 1.4 Part of the land use zone-physiographic map unit overlay.

Upon automatic union, many small units result with a rather unreliable listing of the content. Through editing of the LUP-PMU map (correcting for the slivers) the number of polygons could be reduced to 566, corresponding to 314 unique combinations of LUP and PMU. Remarkable are then the many cases in which LUZ and PMU boundaries coincide, despite the difference in mapping procedure. Arrows in Figure 1.4 show where boundaries correspond. A calculated estimate indicates that an average of 63 per cent of the land use zones contours corresponds to boundaries of physiographic mapping units. This shows that the land and soil characteristics are important in explaining the geographical distribution of the land cover and land use patterns.

Matrix calculations

The LUP-PMU map data can be organized in a $m * p$ matrix A_{ij} with m = 43 rows representing the LUP's and p = 62 columns representing the 62 PMU's, with the elements indicating the area in hectares. For every PMU the land capability class distribution can be calculated, given its composition of TUs and associated land capability class. These data can also be ordered in a $p * n$ matrix B_{ij}, with p = 62 corresponding to the number of PMU's and n = 17 corresponding to the 17 Land Capability Classes (LCs). The elements of the matrix denote the fraction of MPU *i* with land capability class *j*.

Multiplying both matrices A and B yields a $m * n$ matrix C with information on the area of LC *i* for LUP *j*. The data structure as implemented in the SIESTA soil information system did not allow for retrieval of the required information through a data base queries operation. A programme was therefore written in ARC-INFO programming language to obtain the necessary information.

The latter matrix 'C' was used for further calculations with respect to the evaluation of the land use. Based on the matrix C not only the distribution of LUP or associated characteristics per suitability class can be determined, but also the distribution of the land capability class per specific land use pattern.

Matching land use requirements with land capabilities

Soil suitability rating

A land suitability class in the FAO (1976) land evaluation procedure, reflects the degree to which land can satisfy the requirements of a particular land utilization type. In this case study the requirements of a number of land utililization types were compared with land qualities as reflected by the land capability classes. The rating is based on expert judgement. Results are presented in Table 1.4, where Class 1 means very suitable, Class 2 suitable and Class 3 not suitable for the particular LUT considered.

The selected Land Utilization Types considered are relevant for the area; for evaluation of their suitability only soil and terrain characteristics were taken into account. Climate was not considered although the high annual rainfall of 3000-5000 mm, plays a role in depressing yields of certain crops through a high desease incidence, occurrence of pests and storage loss. Also economic criteria were left out of consideration.

Evaluating actual versus potential use

To judge whether land is used in accordance with its potential, actual land use per capability class is compared with its potential use (Table 1.4). Two conditions are specified namely (1) over-use and (2) under-use of the land.

<u>Over-use</u> is defined as land where the requirements of a particular land utilization type exceeds the particular land quality of a land unit, as such presenting a risk of degrading that particular land type and its vegetative cover because requirements of that particular land use type can not be met by the land qualities of the land type. In the suitability rating it is indicated as not suitable.

To illustrate what is meant by risk for degradation of the vegetative cover, the improved grasslands on poor soils might serve as an example. In the suitability rating, distinction is made between improved pastures and 'natural' grasslands. The first consist of improved grass species like Estrella (*Cynodon nlemfuensis*), Tanner (*Brachiaria sp.*) and Guinea (*Panicum*

Table 1.4. *Suitability rating of the land for specific land utilization types.*

	Suitability classes (SC)															
	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	4.1	4.2	5.1	5.2
Maize non mechanized fertilizer appl.	1	2d	1	3e	2n	3dn	2n	3e	3n	3nd	3n	3ne	3no	3no	3	3
Maize, mechanized fertilizer appl.	1	2d	2l	3me	2n	3dn	2l	3e	3n	3n	3n	3n	3o	3o	3	3
Ornamental crop	1	2d	1	2e	1	2d	1	2e	3n	3dn	3nl	3ne	3o	3o	3	3
Root and tuber crops	1	2d	1	2e	1	2d	1	2e	1	2d	1	2e	2o	3o	3	3
Palm heart	1	1	1	1	1	1	1	1	1	1	1	1	3	3	3	3
Cacao	1	1	1	1	1	1	1	1	2n	2n	2n	2n	3o	3o	3	3
Banana plantation	1	2d	1	3me	2n	2nd	2n	3m	3n	3n	3n	3n	3o	3o	3	3
Macadamia plantation	1	2d	1	1	1	2d	1	1	2n	3dn	2n	2n	3o	3o	3	3
Improved pasture	1	2d	1	1	1p	2pd	1p	1p	3p	3pd	3p	3p	2o	3o	3	3
'Natural grasslands'	1	1	1	1	1	1	1	1	1p	2pd	1p	1p	2o	3o	3	3
Pasture with fruit trees	1	2d	1	1	1	2d	1	1	2n	2dn	2n	2n	2o	3o	3	3
Timber plantation	1	2d	1	1	1	2d	1	1	1	2d	1	1	2o	3o	3	3
Natural forest	1	1	1	1	1	1	1	1	1	1	1	1	2o	3o	1	1*

l = Limitation for mechanical tillage of the land.
o = Limitation ragarding depth of soil, limited rooting.
d = Limitation regarding drainage of the soil.
p = Soil compaction and limited recuperative power of the soil, degradation of the grassland.
e = Erosion risk, limitation for mechanized tillage of the land.
n = Limitation with regard to the nutreient status of the soil, pH.
m = Limitation with regard to tillage and other cultivation practices.
*) Suitable for peat and swamp vegetation; unsuitable for natural forest.

maximum). Especially on the less fertile soils they are easily suppressed by species more adapted to these poorer circumstances. Therefore, the improved grasslands do not represent a suitable use on these soils.

Under-use is defined as land offering possibilities for more intensive or more requiring use, because requirements of present land use are more than satisfied by the land qualities of this particular land type.
To decide whether land is over- or under-used, a ranking order was established which reads as follows:
- Annual crops, banana plantations and some ornamental crops are considered most requiring as regards soil fertility and to represent the most intensive use of the soil;
- Perennial crops like palmheart, some ornamental crops, cacao and coffee are judged second in requirements with respect to soil and management;
- Grassland is considered third, although there need not be much difference with perennial crops. Improved grassland which requires fertile soils, is demanding as regards Pasture management.
- Reforestation and tree plantations are considered fourth with respect to the level of land use intensity;
- Forest, swamp vegetation and waste lands are least requiring.

When indicating possibilities for more requiring use, socio-economic conditions were also taken into account; for example in the case of areas with small farms banana production was not considered a viable option.

Evaluating land use at sub-regional scale

The evaluation procedure, as presented above, refers to land utilization types in relation to soil and terrain conditions. At sub-regional level, land use is described in terms of land use patterns, usually referring to a combination of land utilization types. Therefore, to apply the above described procedure for the evaluation of land use patterns, the following procedure is followed:

- The most demanding land utilization type is used to assess whether the land use pattern represents an over-use of the land.
- To evaluate possibilities for more intensive use the least demanding land utilization type of the land use pattern at issue is considered.

E.g. in case of an LUP consisting of maize in combination with grassland, maize is used for the evaluation of a possible over-use and grassland for the evaluation of a possible under-use.

Based on the above mentioned considerations, the soil suitability matrix is translated in two matrices D_1 and D_2; the first indicating LUP-LC combinations with over-use, and the second indicating combinations with under-use. The elements of the matrix can only have two values (1 or 0). The two matrices represent expert knowledge.

Data on under- or over-used areas (in hectares), are obtained by multiplying the matrix C with the matrix D_1 or D_2 respectively.

4. RESULTS AND DISCUSSION

Over-used areas

The Land Use Zones with over-use cover an area of 13977 Ha, corresponding to 17.5 % of the total area. Figure 1.5 shows the distribution of land use patterns for the over-used areas. Of those areas 37 % is explained by the area of Agricultural Penetration (AP), both with grazing (AP/GRASS) and with a combination of grassland and perennial crops (AP/MIXED) as most important use.

The area of AP/GRASS is found in the coastal plain as well as on the slopes of the Turrialba volcano. However, a large proportion of the latter area in the south western part of the study area (Fig. 1.6) has slopes of over 30 % which makes the soils very vulnerable to erosion and degradation under any type of agricultural use.

The area of AP/GRASS indicated as over-used, represents 25 % of the total area of AP/GRASS. In the coastal plain the areas of AP/GRASS cover some peat and swamp areas (appr. 400 Ha.), which also represent over-use. Further agricultural penetration would involve the colonization of peat and swamp areas.

The 31 percent of the area of AP/MIXED also represents a risk of degradation; such areas are found in the central part of the study area (Fig. 1.6) where perennial crops are grown on acid and poor soils (LC3.4) on the sloping remnants of old lahar deposits. It corresponds to 15 % of the total over-used area.

In the agricultural areas with a combination of LUT's (AL/MIXED2 and AL/MIXED3) over-use is due to the growth of maize, which is a very requiring crop, on soils with low fertility (Class 3), on sloping land (LC2.4 and 3.4), on very shallow (Class 4) and very poorly drained soils (Class 5). It accounts for 19 % of the total over-used area. Evaluation is, however, difficult, given the composite nature of the units. Figure 1.6 shows therefore only the most evident cases.

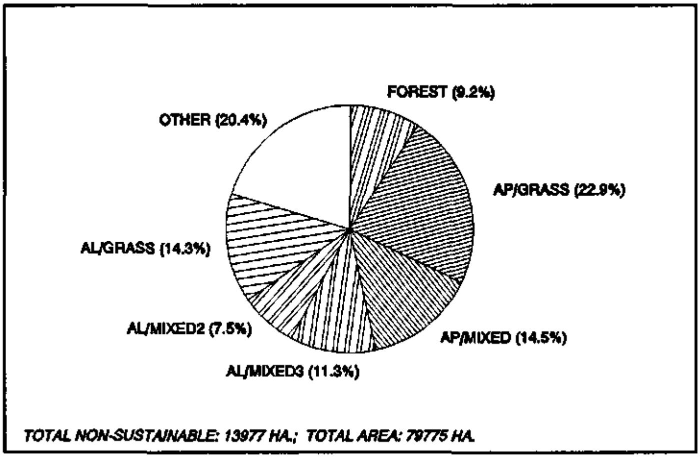


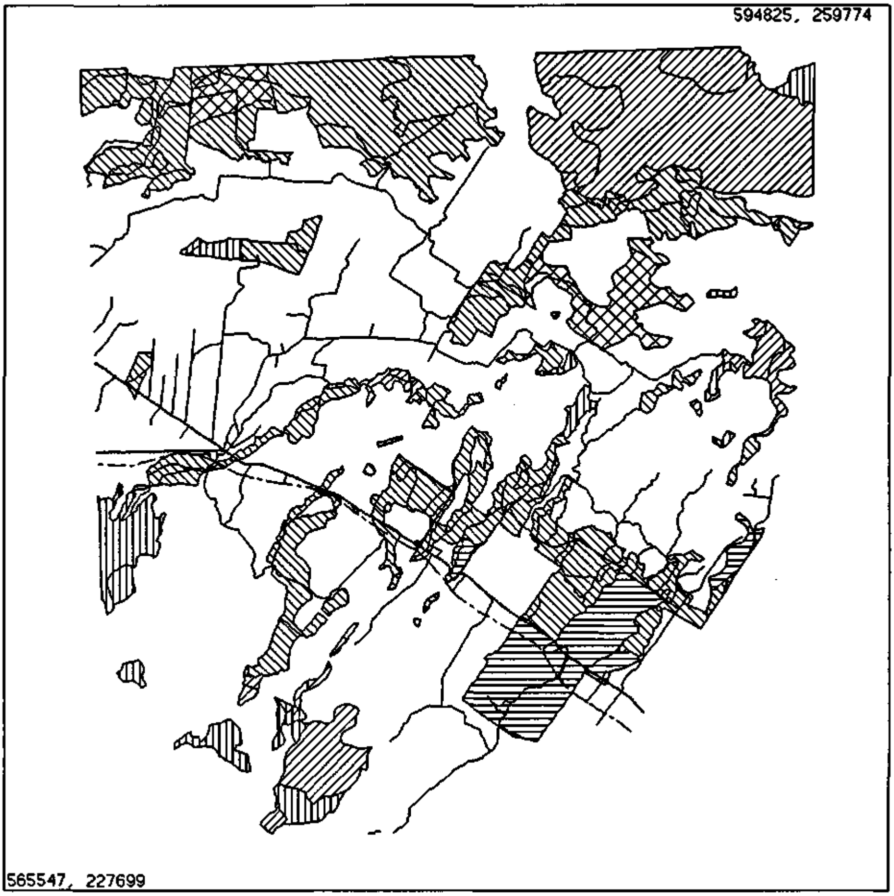
Fig. 1.5 Area percentages of land with non sustainable use.

An important percentage of the area with over-use is due to grassland used for grazing in areas with slopes of over 30 %. These are areas to be protected because of a great risk of erosion. The presented figures show that the area where actual land use presents a severe and acute danger for land degradation, is relatively small. It occurs in areas of agricultural penetration (AP) on the middle slopes of the volcano, posing a severe risk of erosion but also in the recently colonized land on old lahar deposits in the central part of the area and in the area of AP containing swamp and peat areas. Those areas are not suitable for any type of agriculture and should be protected.

Under-used areas

A large part of the study area is indicated as having potential for more intensive use of the land (under-use). In total 40673 hectares, or 51 % of the total area, is labelled as such. The percentage distribution of the Land Use Patterns (LUPs) in the under-used area is presented in Figure 1.7. The following categories are concenred:
Grassland. This occupies 50 % of the under-used area of which 24 % (AP/GRASS) is on fertile soils with drainage problems (LC1.2) and 26 % (AL/GRASS) on fertile soils without drainage problems. They are found in the northeastern and northern part of the study area (Fig. 1.8).
Mixed land cover (Al/Mixed1). This LUP consists of small farms with crops such as maize, root and tubers, perennials and grassland used for cattle breeding and grazing (usually with various kinds of fruit trees). Soils belong to the best of the area. Under-use is especially obvious on the grasslands in these areas.
Cacao (AL/SFOV). Cacao with a low level of management predominates. The plantations are often neglected. Soils could be used more profitably for crops such as roots and tubers.

In summary the under-used areas consist for the major part of fertile land used for grazing; they usually have a young land use history or experience active colonization.



Areas of which the land properties allow for a more intensive use of the land than that is represented by the actual land use.

- Areas with forest cover
- Areas of agricultural penetration, with deforested parts used for grazing
- Agricultural land, used predominantly for grazing purposes
- Agricultural land used for perennial crops (Cacao) and grazing
- Agricultural land; land use directed to arable cropping (maize) and livestock production

Fig. 1.6 Areas of which the land properties allow for a more intensive use of the land (only those areas referring to the listed land use patterns).

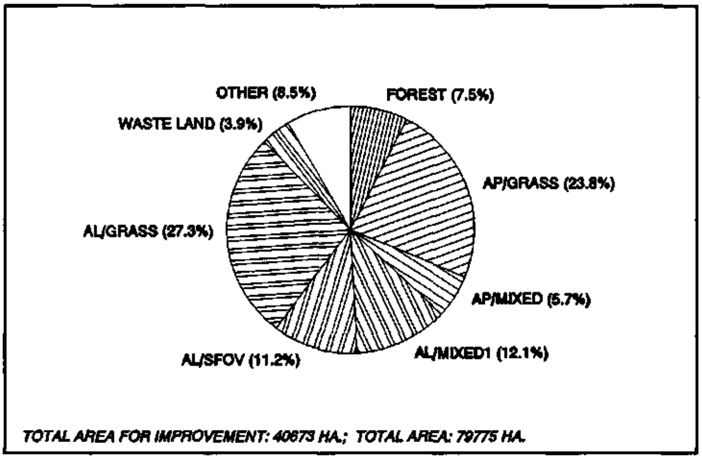


Fig. 1.7 Distribution of the land use categories for the areas indicated as having potential for more intensive use of the land.

Evaluating land use per land use zone

While the former section evaluated present land use and possible alternatives for the whole area, this section focusses on land use zones with one particular land use type. Compare individual land use zones when they represent the same land utilization type, with respect to land capability allows for the evaluation of yield levels in relation to land capability characteristics. In this case an evaluation is provided by four land use zones representing banana plantations. The analysis supplies information on one or more attributes of the land use category one is interested in, in this case management of the plantation and yield level in relation to land capability class.

Figure 1.9 shows that the proportions of capability classes, LC1.1 and LC1.2, can differ considerably. This is, however, not reflected in the yields (given in boxes/ha/year, corresponding to the year 1986). The yield levels of 'Rio Frio' and 'St. Clara' are comparable, while the first has the highest percentage of first class soil and the second the lowest of the four. 'St. Maria' has the lowest yield level. It indicates that the St. Maria farm possesses potential for higher yields, given the present socio-economic conditions, because its soil suitability is similar to that of the other plantations. I.e. managerial factors explain the poorer peformance. In 1986 the plantation was put under new management and by 1988 yields had increased from 1494 to 1935 boxes/ha/year (source: ASBANA, Dirección Estudios Económicos, 1990). A detailed study of this plantation with respect to yield in relation to soil type was published by Veldkamp *et al.* (1990).

Land capability class and farm size distribution

Although economic data are crucial in the evaluation of land use for sustained use, it was not possible to use them because they were lacking for the region.
However, there are some parameters, provided by the land use map which can be used to

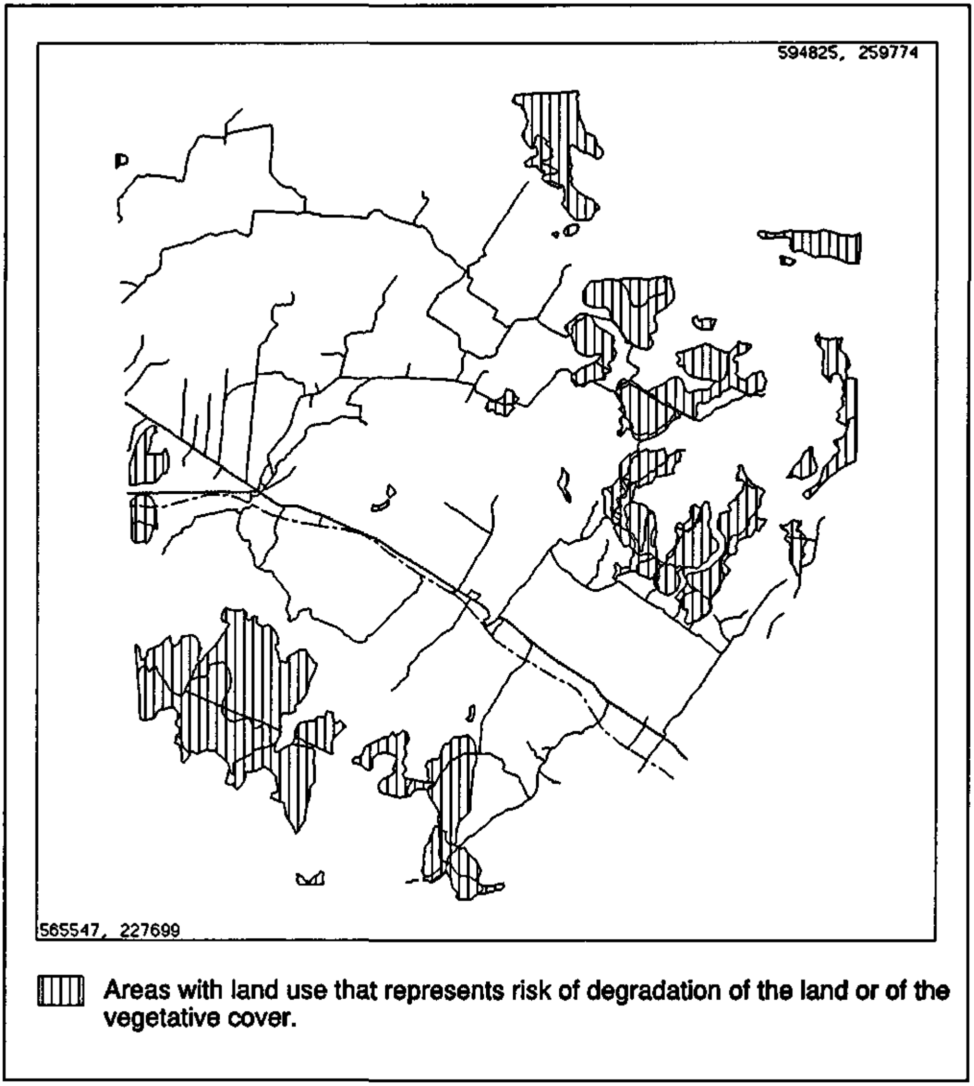


Fig. 1.8 Areas representing land use with higher demands than can be met by the land properties.

analyze land use in socio-economic terms. One of those is farm size (Fig. 1.3). Farm size distribution in relation to land use and land capability classes, provides information on how economic opportunities in the area are divided. Fertile soils in the area offer a wide range of land use alternatives for virtually any type of farmer. The less fertile soils are, however, more restricted as regards viable land use alternatives, while the more profitable ones like Macadamia require large investments without any profit in the first years of growth. This implies that economically attractive land use alternatives on less fertile soils are only within reach of larger and/or economically strong farmers.

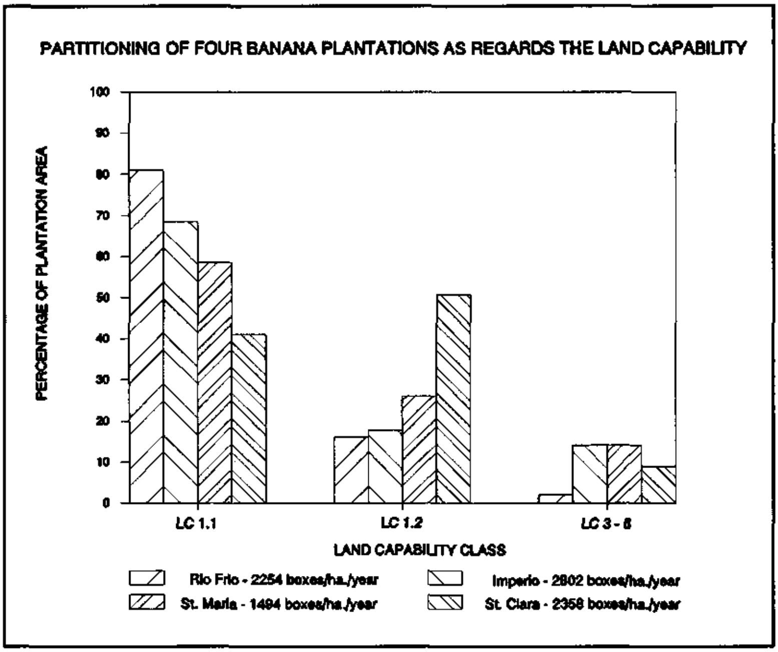


Fig. 1.9 Suitability class distribution of four banana plantations with their yields (Boxes/Ha./Year).

A comparison of data from Figure 1.10 and 1.3 shows that the area percentage of large farms is bigger on first class soils (Fig. 1.10) than when considering all capability classes (Fig. 1.3). For small farms the reverse is true. It illustrates how opportunities are divided for both farm size classes.

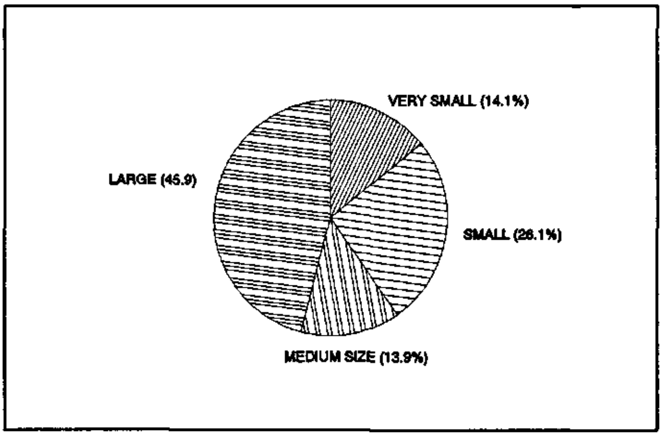


Fig. 1.10 Area percentage of farm size class on fertile soils (suitability class 1.-).

Banana is an example of a land use type almost exclusively found on first class land (Fig. 1.11). All land for banana production is artificially drained so that impeded drainage (class 1.2) does not present a limitation. The strong expansion of the area under banana cultivation in the recent years is observed in areas indicated in this study as having potential for more intensive use. These areas are characterized by the presence of larger farms dedicated to cattle breeding and first class soils, however suffering from impeded drainage. Improving the drainage conditions is not a viable option for the existing cattle farms. It is one factor which explains the penetration of the banana plantation in this area. The other factor is related to the larger size of the farms encountered in the area, which make it possible for the banana plantations to acquire larger tracks of land. It illustrates how the evaluation of land use considering biophysical as well as socio-economic factors might be used to predict and evaluate land use change.

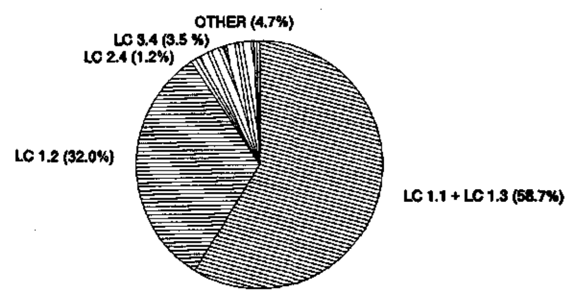


Fig. 1.11 Distribution of the suitability classes for the total of banana plantations.

Evaluating sources of error and uncertainty

As explained in the introduction, maps were used in which information is aggregated. Overlaying of such information leads to complications as regards the accuracy and validity of data. It is important to discuss this because such data might be used in a GIS for land use planning with consequences for farmers in the area.

The following sources of error and uncertainty upon overlaying might occur:

(1) Inaccuracy due to matching and reduction of the PMU-LUZ combinations. This procedure helped to minimize the number of combinations. It concerns here a relatively small area so that we may neglect this source of inaccuracy.

(2) Uncertainty as consequence of working on different levels of aggregation.

Over use and under use refers to land utilization types in combination with terrain units. The determination of the corresponding areas of over- and under-use refers to combinations of land use zones and physiographic mapping units; both composite entities either in terms of LUTs or TU's. This has certain consequences:

A) In case of a composite LUZ, over- or under-use is decided on the basis of either

the most exigent or least demanding LUT, constituting only a part of the LUZ of concern. The results, as such, over estimate the area of over- or under-use.
B) The PMU are described in terms of percentages of the constituting terrain units. In calculating the area of over use or under use, these percentages are accounted for. The percentages might be very accurate for the PMU as such. This need not be true for that part of the PMU intersecting with a specific LUZ (illustrated by Fig. 1.12, where A ∩ B represents the intersection of PMU A and LUZ B), since there is no information on the location of the TU within the PMU.

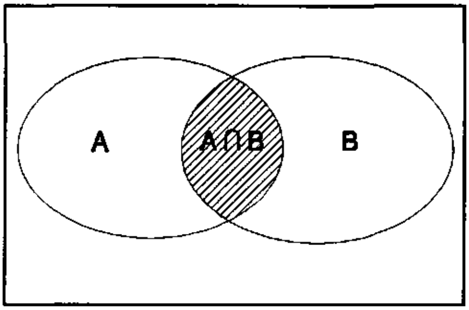


Fig. 1.12 Intersecting Physiographic Mapping Unit and Land Use Zone.

This implies that the figures denoting the percentages of the different TU's represent probabilites or predicted values in case of the intersection. It introduces uncertainty with regard to the statements on the area of non-appropriate use.
The following might illustrate this. Suppose a mapping unit of the physiographic soil map consists of 50 % swamp soils and of 50 % well drained soils. After overlaying the land use map, 50 % of the mapping unit is covered by commercial forest and 50 % by grass vegetation. Overlaying then automatically creates PMU-LUZ units in which both land use types would grow for 50 % on well drained and for 50 % on swamp soils. In reality it could also be that forest would grow exclusively on the well drained soils and the grass vegetation in the swamps.

This type of uncertainty can not be eliminated and will always exist when aggregated information is combined. The percentages, and corresponding hectares, indicating over- or under-use of the soil, should then be read as areas corresponding to LUP-PMU units, which possibly represent inadequate use. The degree of error or uncertainty depends on the number of components of the PMU or LUZ and on the degree to which the Land Uze Zones and the Physiographic Mapping Units correspond. This allows for excluding units with a strong composite character in order to exclude less reliable results. For the area of over- and under-use, as presented in figure 1.6 and 1.8, this was indeed done.
Given the 63 % of the land use zone boundaries corresponding to PMU boundaries and given the in general rather homogeneous character of the mapping units (by majority the PMU consist of one or two Terrain Units), the results are thought to give a reliable indication of the area of over-use and under-use in the area of concern.
The presented procedure serves to identify possible problem areas, indicating their relative importance. It draws attention to areas where the magnitude of the problem should be studied in more detail.

6. CONCLUSIONS

- Nearly twenty per cent of the area is classed as over-used. These areas are predominantly found on poor soils in the colonization areas and in areas colonized during the last 10 years. The agricultural frontier reached a point where its further extension will mean a severe risk of degradation for the land, because the surrounding forest areas are not suitable for more intensive use. The colonization area has reached a limit, beyond which no appropriate agricultural use is possible.

- The land use in the area is not as much characterized by over-use as by not using the land and soils to its full potential (under-use). In 51 % of the area more intensive use of the land is possible, predominantly because fertile soils are now used for grazing which is considered under-use.

- A comparison of farm size distribution with land capability classes shows that the area percentage of large farms or plantations is bigger on first class land as compared to other land capability classes. For small farms the reverse is true.

- The aggregation level of the different sources has to be comparable to allow for meaningfull comparison. In this regional study, Physiographic Mapping Units were therefor copared with Land Use Patterns within Land Use Zones.
When aggregated information from two sources is overlain as in this study on land and land use (i.e. when the units are composite in nature) the information serves to indicate areas where more detailed studies are likely to be relevant when comparing soils or Terrain Units with Land Utilization Types.

CHAPTER 2

VARIATION OF MEASURED BANANA YIELDS IN A COSTA RICAN PLANTATION AS EXPLAINED BY SOIL SURVEY AND THEMATIC MAPPER DATA

E. Veldkamp, E.J. Huising, A. Stein and J. Bouma

Geoderma 47: 337-348.

Abstract. Variation in yield was studied on an 370 ha banana plantation in Costa Rica. Yield for 42, 6-12 ha large sections, were expressed in terms of: number of bunches/ha, weight/bunch, gross production/ha. Yield variation was explained by readily accessible soil and thematic mapper (TM) information, using three models. First, a TM image was used. A number of band combinations (vegetation indices) and Near Infrared Band (TM4) were tested to characterize each section. The TM4 band and the "Greenness" band combination appeared to be the best estimator with 46% of the variation explained. Second, yield variation was explained on the basis of a detailed soil map that was analyzed for each section. Of the variance in gross production, 67% could be explained by production per soil unit. Gross production of the units varied from 504 to 896 kg/ha/week. In the third model, the TM and soil data were combined. The explained part of the variation in a combined model remained 67% or was even lower. The study indicates that the soil map of this area was a good estimator for banana yields in terms of gross production. Since the production figures were only available per section, i.e., for relatively large rectangular areas, the averaged TM image for a section did not provide extra information. Site specific, stratified sampling per soil map unit is proposed to improve results of the type of models used in this study.

INTRODUCTION

Bananas are a major cash crop in Costa Rica. The banana crop is grown on large plantations with uniform management. Nevertheless, large variations in yield exist within various parts of the plantations. Difference in soil conditions are considered to be the main cause of these variations.

Soil surveys have been used to relate yields to soil conditions. Variation in soil units may differ from that in others. Soil map delineations can therefore, be relevant when preparing predictive maps (Stein *et al.*, 1988). Since delineated areas on soil maps may have an internal variability up to 40% (e.g. Marsman and de Gruijter, 1986), the *first* objective of this study was to test the feasibility of relating actual banana yields with soil map units. The soil map used in this study, was prepared in a traditional way, using borings and interpreting physiographic features.

Yields can not only be measured, they can also be estimated by using remote sensing data. Considerable experience has been gained by using spectral data for the estimation of canopy biomass and leaf area index (LAI) of grain crops and rangelands (e.g, Tucker, 1979; Williamson, 1988; Plummer, 1988). Estimating green canopy biomass, using spectral information, had been most successful for wheat and other grains (e.g. Gardner *et al.*, 1985). The periodic mapping of green canopy biomass has been used as a basis for forecasting yields of grains crops, as yields of these crops are a function of the development curve of the green canopy. The *second* objective of this study was therefore the estimation of yields in a banana plantation, using TM images. As the banana crop differs considerably from a grain crop, a different approach was required.

Most studies relating spectral information to plant parameters of various kinds have used field spectrometry and very specific information such as LAI and biomass (e.g., Richardson *et al.*, 1983; Tucker, 1977, 1979). In this study only information was used that can be easily obtained for any banana plantation (i.e. actual yields and spectral information from TM images). For management purposes, most plantations keep records of the yields per section.

The *third* objective of this study was to test whether a combination of remote sensing and soil survey data would improve the regression models of each of these variables separately.

MATERIALS AND METHODS

The study area

The study was carried out on the plantation "Santa Maria" which covers 370 ha and is located in the Guacimo district of the Atlantic Zone of Costa Rica. The climate is tropical and humid with a mean annual temperature of about 27°C (isohyperthermic) and an average precipitation of about 3300 mm (perudic) (Soil Survey Staff, 1975).

The Atlantic Zone consists of a broad alluvial plain. The deposited sediments originate from the Central Cordillera, a mountain range mainly consisting of volcanic rocks. The study area is at the footslopes of the Turrialba volcano. The sediments in the study area have been deposited by the river Jiménez.

The soils

During the soil survey five different soil units were distinguished (Fig. 2.1), which are briefly described below. classifications are according to ICOMAND (1988) and Soil Survey Staff (1975).

Milano soil. Classification: Andic Humitropept. This soil, located at the highest positions of the study area, is considered to be developed in the remnants of Pleistocene fluvio-laharic deposits. The Milano soil is a deep, moderately well to well drained soil with dark reddish-brown colours and slightly andic properties. Structure is moderately strong to strong crumb and sub-angular blocky. The Milano soil is very porous and the texture class is very fine.

Santa Maria soil. Classification: Aquic Tropudalf. This soil is situated on the younger river terrace than the terrace remnants on which the Milano soil occurs. The Santa Maria soil is a deep, moderately well drained soil with dark-brown to brown colours and no andic properties. Structure is moderate to strongly developed angular blocky. Texture class is fine to very fine.

La Lucha soil. Classification: Aquic Hapludand. This soil has been developed in recent alluvial material. The La Lucha soil is moderately deep and moderately to poorly drained, with clearly expressed andic properties. The colour is brown to greyish brown. Structure is weak angular blocky and the porosity is high. Texture is sandy loam to loam.

Dos Novillos soil. Classification: Melanic Udivitrand. This soil is situated on the recent natural levees of the Rio Jiménez. The Dos Novillos soil is deep and well drained with clearly expressed andic properties. The colour is dark brown to brown. The structure is weak sub-angular blocky and the porosity is high. The soil has a loamy texture with a sandy sub-soil.

Bosque soil. Classification: Fluventic Tropaquept. This soil is situated in the backswamps of the Rio Jiménez. The Bosque soil is a deep, poorly drained soil with no andic properties. The colour is yellowish grey to dull brown. The structure is strong angular blocky and the porosity is high. The texture is very fine.

Yield data

The banana grown on the plantation is the clone "Valery". Throughout the year fruit bunches are produced at a rather constant rate. All possible stages of development can be

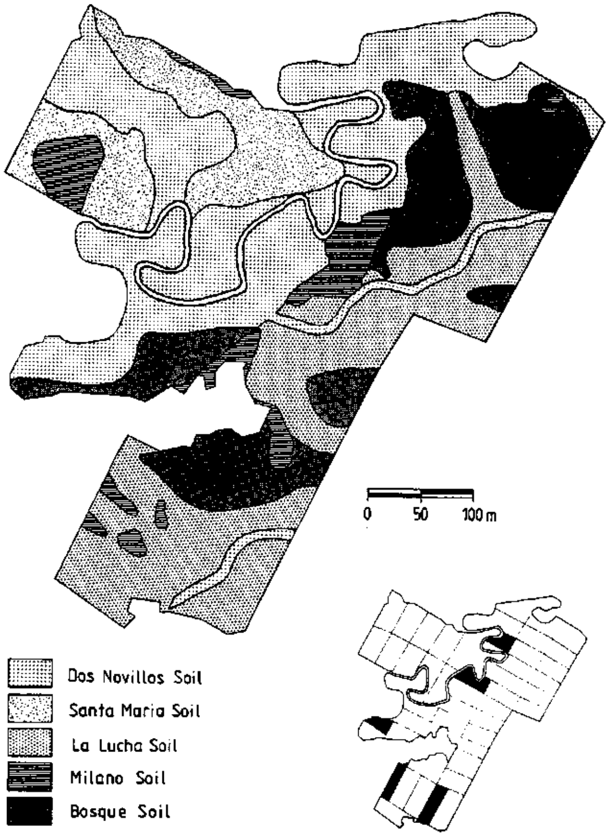


Fig. 2.1 *Soil map of the study area. Below right, section map of the study area. Shaded areas were excluded.*

found within the same plantation. Fruit are produced about 10 months after the pseudostem starts to grow (Soto, 1985). The total fruit production depends on: (1) the number of fruit bunches per ha; (2) the weight per bunch; (3) the time lapse between subsequent bunches from the same plant (stool).

The plantation is divided into 47 sections (Fig. 2.1), varying from 6 to 12 ha. Yield data were collected per section. The following yield data were used in the study:

(I) The number of bunches produced per ha per week for each section during the period of May 1987-May 1988.

(II) The average bunch weight per section. This was collected in May 1988 by counting the "hands" per bunch (one hand consists of a number of bananas attached to the bunch at the same place) and using the relation between bunch weight and number of hands as found by Jaramillo (1979): $W=5.107HB-18.043$, Where HB=the number of "hands"/bunch and W=mean bunch weight (kg).

(III) The gross production, calculated from the previous two data.

In the study only 42 sections have been used (Fig 2.1). For two sections yields data were incomplete and the area of three sections has been enlarged since the TM image had been recorded.

The TM image

The TM image used was recorded in February 1986. Band 4, hereafter referred to as TM4 and a number of TM band combinations which are commonly used to estimate biomass and LAI were included in this study. The "raw" data, available in bytes (0-255), were transformed into reflection percentages using the method as described by Markham and Baker (1987). In order to rearrange the available spectral information, a principal component analysis was performed. The first principal component was also included in the study. Table 2.1 provides an overview of ratios and band combinations being used.

Combining data with a geographic information system

The interpretation of TM images and geographic information was done with the ERDAS system (ERDAS Inc., 1987). To compare TM information, soil and yield data maps were digitized and transformed geometrically to fit the TM image. For the classification use was made of supervised pattern recognition. Four training samples were selected, covering the entire range of banana crop reflectance characteristics. Mean, covariance and standard deviation were calculated to test homogeneity of the training samples. The classification of the picture elements (pixels) was done by means of a minimum distance classifier. Only the pixels classified as banana pixels were further considered. The classified TM image was overlain by the digitized soil and section maps. A routine was developed to calculate the mean and standard deviation of the pixels of each digitized polygon (sections or units). Pixels, situated on the boundary of the two polygons were discarded.

The following data were extracted: (a) the area of each section; (b) the fraction of the area of each soil unit per section; (c) the average value of TM4, band combination or principal component per section; and (d) the average value of TM4, band combination or principal component per soil unit per section.

Interaction of spectral radiance and banana leaves

Gardner *et al.* (1985) have summarized the interaction of the wavelengths of the Thematic Mapper with a vegetation canopy. The TM1, TM2 and TM3 can be considered to contain information concerning the chlorophyll and pigment concentration in the first layer of leaves near the top of the canopy. The TM4 responds to multiple leaf layers. Reflectance continues to increase non-linearly, with increasing number of layers, up to a value which is 85% higher than the reflectance from a single leaf (Myers, 1970). The TM5 and TM7 both respond to changes in leaf area or water content. The TM5 interacts with more than one leaf layer, though not with as many layers as the TM4, and the TM7 responds to the first layer of leaves only.

Since the TM4 is the most important TM waveband which responds to multiple leaf layers, Gardner *et al.* (1985) considered this band to be the principal vegetative response or leaf area indicator. Since the canopy of a banana crop is nor closed, a considerable amount of the TM4 reflectance originates directly from the soil. In order to minimize the soil effect, various combinations of TM4 with other bands were tested.

Only biomass and LAI can be directly estimated on the basis of spectral information (e.g. Tucker, 1979; Williamson, 1988; Plummer, 1988). Yields must be estimated indirectly, using correlations with biomass, LAI (leaf area index) or other crop characteristics. For banana it is assumed that the LAI and biomass were also correlated with yield.

Statistical procedures

The central question in this study is how to explain variations in yield data by readily available information like soil map delineations and TM images. As the number of available yield data was too limited to apply geostatistical methods, use was made of statistical methods on the assumption that observation were independent. Several regression analyses were carried out, depending upon the sets of explanatory variables.

(1) To explain variations in yield on the basis if TM data, a simple linear regression model was used. This approach seemed justified, given the rather small range of reflectance values in spite of the non-linear relationship between radiance and biomass (Tucker, 1979; Plummer, 1988). The regression model was applied to the TM4 band and several band combinations and is formulated as:

$$Yi=\alpha_0+\alpha_1\mathrm{SPEC}_i \tag{1}$$

where SPEC_i is the mean pixel value of a spectral band or band combination for section $i(i=1,\ldots,42)$, Y_i is the actual yield for section i and α_0 and α_1 are parameters, to be estimated by the model.

(2) To explain yield variations on the basis of soil units a multiple linear regression model was used:

$$Y_i=\mu+\beta_1\mathrm{BO}_i+\beta_2\mathrm{DN}_i+\beta_3\mathrm{LL}_i+\beta_4\mathrm{MI}_i+\beta_5\mathrm{SM}_i \tag{2}$$

where BO_i, DN_i, LL_i, MI_i, SM_i $(i=1,\ldots,42)$ denote the fraction of the area of the soil units "Bosque", "Dos Novillos", "La Lucha", "Milano" and "Santa Maria", respectively, in the ith section. Y_i represents the actual yield for section i, μ is the mean value, and the β's are parameters. Because the fractions for every section sum up to 1, contrasts of the parameters (e.g., $\mu+\beta_j$) provide an estimate for the average yield production in the jth soil unit.

(3) To explain yield by soil map units *and* TM data, two different multiple linear regression models were used. The first model is a combination of eqs. 1 and 2, where the reflectance, in the band combination as obtained from eq. 1, for section i is included as an extra explanatory variable:

$$Y_i=\alpha_0+\alpha_1\mathrm{SPEC}_i+\beta_1\mathrm{BO}_i+\beta_2\mathrm{DN}_i+\beta_3\mathrm{LL}_i+\beta_4\mathrm{MI}_i+\beta_5\mathrm{SM}_i \tag{3}$$

where the names of the variables correspond with the names mentioned above. However, different coefficients will be obtained. In fact, this model was applied only with the TM4 band, being the most promising variable of the first model. The extra variable was included to detect additional information from TM4. This model can be interpreted to be an analysis of how much variation in the residuals from eq. 2 can be explained by the TM4 band.

In the second model the explanatory variables are obtained by multiplying the average TM4 value per soil fraction with the fraction of the soil units per section:

$$Y_i=\gamma 0+\gamma_1\mathrm{TMBO}_i+\gamma_2\mathrm{TMDN}_i+\gamma_3\mathrm{TMLL}_i+\gamma_4\mathrm{TMMI}_i+\gamma_5\mathrm{TMSM}_i \tag{4}$$

Where $TMBO_i$, $TMDN_i$, $TMLL_i$, $TMMI_i$, $TMSM_i$ ($i=1,\ldots,42$) correspond with the fraction of the soil types Bosque, Dos Novillos, La Lucha, Milano and Santa Maria, respectively, in the ith section, multiplied by the TM4 value within this fraction. Again, Y_i represent the actual yield for section i and the γ's are parameters to be estimated by the model.

RESULTS AND DISCUSSION

Banana yields and thematic mapper data

The regression results (Table 2.1) indicate that there are substantial difference among the R^2 for gross production, bunch weight and bunch number on the basis of spectral reflectance values. The best explanatory variables for the actual yield appear to be the TM4 and 'Greenness'. In Fig. 2 TM4 is plotted against gross production.

Table 2.1 *Tested ratios and combinations of TM bands and results of spectral estimation on actual yield*

Name	Formula	R^2		
		gr. prod.	#bunch	weight
TM4		.46*	.30*	.24*
NIR/red[1]	TM4/TM3	.42*	.28*	.22*
√(NIR/red)[1]	√(TM4/TM3)	.42*	.28*	.22*
NIR-red[1]	TM4-TM3	.45*	.30*	.24*
VI[1]	(TM4-TM3)/(TM4+TM3)	.41*	.27*	.22*
1/VI[1]	(TM4+TM3)/(TM4-TM3)	.40*	.26*	.22*
TVI[1]	√(VI+0.5)	.40*	.26*	.22*
NIR/MIR[2]	TM4/TM5	.27*	.20*	.13
	(TM4-TM5)/TM4+TM5)	.25*	.19*	.12
	(TM4-TM3)/(TM5+TM7)	.32*	.23*	.15
Greenness[3]	-.27TM1-.22TM2-.55TM3+.72TM4+.07TM5-.17TM7	.46*	.30*	.24*
Brightness[3]	.29TM1+.25TM2+.48TM3+.56TM4+.44TM5+.17TM7	.04	.02	.03
Wetness[3]	.15TM1+.18TM2+.33TM3+.34TM4-.62TM5-.42TM7	.21*	.16*	.10
PC1	-.12TM1-.08TM2-.19TM3+.73TM4-.56TM5-.31TM7	.34*	.24*	.25*

NIR=Near Infrared; MIR=Middle Infrared; VI=Vgetation Index; TVI=Transformed Veg. Index; PC=Principal Component

[1] See Tucker, 1979; [2] See Williamson, 1988; [3] See Crist and Cicone, 1984;

* Significant at $\alpha=0.1$.

In all cases the regression model explaining variation in gross production has the highest R^2, followed by number of bunches and average bunch weight. It can be concluded that gross production is the best estimable yield characteristic.

A considerable part of the observed variation could not be accounted for, because interpretation of the TM image was restricted to a particular day and the yields were whole year averages. Also, soil reflectance may have a disturbing contribution in the total reflectance. However, band rations, normally used to minimize soil influence, did not lead to an increase in the R^2 when compared with the sole use of TM4. an explanation might be that the variation in reflectance values within some TM bands is very small. Even the TM4 band has a coefficient of variation of only 0.06, with values ranging from 33.5 to 42.6.

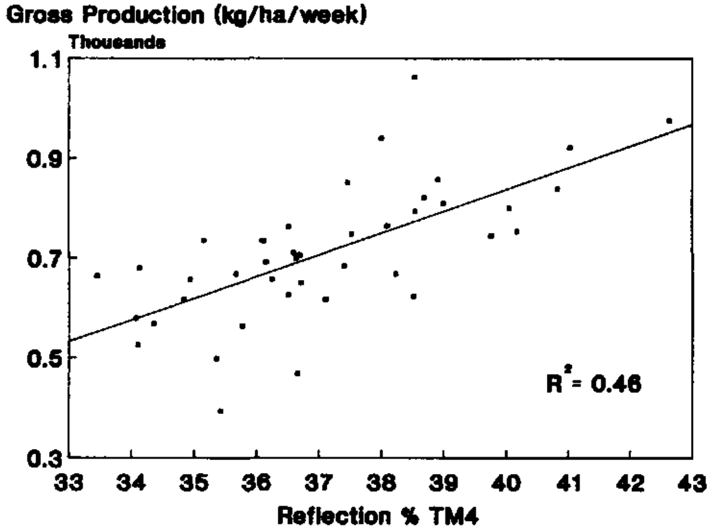


Fig. 2.2. *Gross production plotted against % reflection of TM4. Results of the best model obtained with eq. 1 (see text).*

Banana yields and soil survey data

Results (Tables 2.2 and 2.3) indicate that the sol survey data explain as much as 67% of the variation in banana yields. Moreover, there are large differences in production among the five soil units. Gross production, bunch number and bunch weight are the highest on the Dos Novillos Soil. The largest difference in gross production (392 kg $ha^{-1}week^{-1}$) and weight (5.7 kg/bunch) are encountered when comparing this soil with the Bosque soil, whereas the largest difference in the number of bunches (11.5 $ha^{-1}week^{-1}$) is encountered when comparing this soil with the Milano soil. Production on the other sol units is intermediate and differences are not significant.

Table 2.2 *Estimates and standard error of estimate (SEE) of the yields based on eq. 2*

Soil unit	Gross prod. (kg $ha^{-1}week^{-1}$)		Bunches (nr.$ha^{-1}week^{-1}$)		Weight (kg/bunch)	
	estimate	SEE	estimate	SEE	estimate	SEE
Milano	621	137	28.3	5.0	22.4	1.5
Santa Maria	783	58	37.9	2.1	20.6	1.5
La Lucha	710	59	35.4	2.1	20.1	1.5
Bosque	504	58	30.1	2.1	17.0	1.5
Dos Novillos	896	68	39.8	2.5	22.7	1.7

The regression model explains variation in gross production best, followed by bunch number and bunch weight. A model based on soil units provides better estimates for yield than a model based on the TM image (Table 2.3). Considering the internal variability of soil maps (up to 40%, Marsman and De Gruijter, 1986) the soil survey data are rather successful in explaining the differences in production. This stresses the relevance of soil survey information when explaining banana yields.

Table 2.3 *R2 of yield estimation by different equations*

	Gross production	Bunches	Weight
Eq. 1	0.46	0.30	0.24
Eq. 2	0.668	0.516	0.368
Eq. 3	0.672	0.516	0.372
Eq. 4	0.409	0.246	0.355

Banana yields and both thematic mapper and soil survey data

The results of the regression analysis (Table 2.3) emphasize that no variation in the residuals of eq. 2 is explained by TM4 values with eq. 3. Of interest is the high degree of collinearity between TM4 and the soil unit values. This implies that the TM4 data are strongly linked with the soil survey data, when explaining yield variation.

Although eq. 4 is promising as such, the R^2 did not increase (e.g. R^2=0.409) for gross production. This could be due to the often small number of 30m x 30m pixels per fraction of a soil unit in a section. Images with a higher resolution are potentially attractive to be applied in this type of study.

The above results lead to the conclusion the eq. 2 is the best linear model to explain variation in actual yield.

In the present study was compared at the level of sections because yield data were not available for smaller areas. However, both soil and TM data had a much smaller resolution. The average reflectance within a section, or within the fraction of a soil unit within a section, has to be used to compare TM data with yield data. Thus, variation of the TM image within a section could be accounted for. This may partly explain why only a small part of the variation in yield could be explained by the TM image.

Implications for future research

In this study a simple relation was assumed between LAI and yield or between biomass and yield. Until now, information on this relation is not available for a banana crop. Research to define this relation is necessary to allow use of remote sensing techniques for the estimation of banana yields. However, this explanatory study showed that yield differences can be related to reflectance data using published relations derived for other crops.

The TM image showed very clear spectral differences within the banana plantation. However, due to the scale at which yield data were available, average reflectances per section or per soil unit within sections, had to be used. In future, more site specific yield data should be obtained. Observations within the different soil units should be made and this does not present a problem since yields from individual plants can easily be recorded. The necessary number of measurements within each soil unit is a function of the internal variation, which can be characterized by geostatistical techniques, and of the complexity of the TM image.

Furthermore, crop characteristics like biomass, LAI and plant density need to be considered separately to explain the variation of the spectral response within sections.

Finally, eqs. 1-4 relate, of course, only to the study area. Independent tests will be made in other plantations were the same soil units occur to establish the predictive quality of the regression equations obtained in this study.

ACKNOWLEDGEMENTS

We thank the personnel of the Santa Maria plantation, where the study was carried out and the National Banana Growers' Association (ASBANA) of Costa Rica for their active cooperation in the collecting the necessary data. The cooperation of M.A. Mulders, W.G. Wielemaker, A. Otten and P.G.M. Versteeg is highly appreciated.

CONCLUDING CHAPTER

LAND USE ZONES TO DESCRIBE LAND USE AT SUB-REGIONAL LEVEL.

The LEFSA sequence (Fresco *et al.*, 1990) defines the sub-regional level as one of the hierarchies in agricultural systems. LEFSA is a conceptual model; It does not give an operational definition of land use at the sub-regional level. In the introduction we mentioned the lack of appropriate tools for the inventory and description of land use at this level. We also pointed out the possible application of land use and vegetation maps. These aim at the identification of vegetation types and crops, but lack information on farms (see, for example, the USGS land cover-land use classification system for use with remote sensor data [Anderson, 1976] or the CORINE land use classification [CORINE, 1989]). Farm information from statistical sources or limited field work has to be added. The LUZ approach, presented in this thesis, adopts such a strategy. The intention to explicitly describe land use at the sub-regional scale has been expressed earlier by Van Wijngaarden and Kooiman (1988). However they did not further elaborate the idea.

The land use zone (LUZ) and its associated land use pattern (LUP) give an operational definition of land use at sub-regional scale. The LUZ provides the geographical basis for the inventory of land use. The LUZs prove to represent stationary spatial features in the area of study, when they belong to the agricultural land. The stability of the LUZ boundaries in time do indicate that the LUZs are spatial objects representing meaningful entities in reality, in stead of LUZs being more or less arbitrarily defined units, having relevance only for the purpose of mapping. The permanent geometric characteristics also indicate that the LUZ may serve as a geographic reference units for monitoring land use change.

Land use is described in terms of the LUP. In case LUZ refers to one farm or part of a farm, the LUP may denote a single land utilization type (LUT) (e.g. cultivation of banana as a plantation crop) or a combination of LUTs. In case LUZ refers to an area comprising many farms, the LUP may denote a single farming system (e.g. large cattle farms) or a combination of farming systems. The LUZ approach has been applied to the Guacimo-Rio Jiménez-Siquirres (GRS) region, a representative part of the Atlantic Zone of Costa Rica. LUZs provided relevant insight in the sub-regional land use structure in this region. A total of 43 distinct LUPs were defined, distributed among the major land use categories (area of natural vegetation, area of agricultural penetration, agricultural area, plantation area, residential area, and other).

For LUZs that comprise many farms, the farming systems were investigated. Clear differences in farming system composition were observed. The investigation also showed that there is always a clear dominance of one farming system. The dominant farming system, sometimes in combination with the sub-dominant system, is generally sufficient to describe the LUP. In doing so, the description would comply with 80 to 90 percent of the farms in a LUZ (see Chapter 3, Part I).

Also land use change could be described adequately on the basis of LUZs. The LUZs satisfactorily indicated the place and type of land use change. Changes in land use were consistent for LUZs that showed comparable land use patterns (Chapter 4, Part I). It means that certain trends in land use change could be observed. These trends could be explained by change in market prices or were confirmed by production figures.

We conclude that the LUZ represents a relevant entity for the inventory and description of land use and land use change at the sub-regional scale. However, some aspects of the application of the concepts LUZ and LUP require further investigation:

1. The farming systems were described in somewhat general terms, since the objective was to prove that differences in farming system composition existed

between sub-regions. The definition of farming systems, especially with reference to the appropriate level of detail, deserves further attention. The farming system should express the correspondence between the farms within one sub-region, while the specific farm characteristics should be treated at farm level.

2. LUZ has been described mainly as a compilation of elementary objects. However, an LUZ also represents a single entity. The independent role of the LUZ (with its associated LUP) in the analysis of land use should be further investigated. Also, attention should be devoted to the justification of the LUZ as an agricultural system. The investigation should give an explanation as to why specific LUPs occur and why LUZs exhibit a particular dynamic behavior.
3. The LUZ approach is expected to be of general interest for applications that require information on sub-regional land use (e.g. for agricultural statistics). The LUZ represents a unit of observation for land use inventory. Its application as a functional unit for sub-regional land use analyses and planning has been addressed in Part III, though only with respect to bio-physical land potentials. The functional aspects of the LUZ require further attention (e.g. the relevance of the LUZ in view of developing alternative land use scenarios).

OBJECT-ORIENTED DATA MODELING

In recent years much attention has been devoted to object-structured data bases. The application hereof in a GIS environment has been investigated (Gahegan, 1988; Roberts *et al.*, 1991). The object-oriented approach has been developed for applications whose data are naturally viewed as a collection of objects. This is less evident for applications like the inventory of land cover, land use or soils. The inventory traditionally focuses on the demarcation of areas of land to describe specific land characteristics, rather than on the recognition of objects in the terrain. The basic model underlying the traditional approach to mapping is that of the attribute value linked to a specific location. The observations are grouped (or clustered) to form mapping units with associated thematic classes. The clustering (referring to spatial grouping of observations as well as the definition of thematic classes) makes use of intuitive and associative knowledge (e.g. by means of interpretation of aerial photos). Criteria for clustering are often not made explicit. The 'land unit' concept (Van Gils *et al.*, 1987) is illustrative in this respect. The 'land unit' represents a "holistically defined tract of land that is ecologically homogeneous". With holistic is meant that the land unit integrates the description of the terrain, soil, land cover and land use. An example of the application of land units for the mapping of agrotopoclimates is given by Zuviría (1992). The 'holistic' view is also expressed by Meijerink (1985) and van Wijngaarden and Kooiman (1988).

The object-oriented approach starts from a different concept, namely from objects that exist in reality of which the geometric and thematic characteristics can be determined independently. The relations between objects are defined through classification, aggregation and association, and the objects have a dynamic behavior.

We successfully modeled land use in the GRS area, by applying an object-oriented approach: the LUZ represents the basic spatial object of which the geometric (the LUZ boundaries) and thematic characteristics (land use pattern) were determined independently; relations between LUZs are defined by classes of corresponding land use patterns; LUZs and agricultural fields are related by aggregation and it is assumed that LUZs can be aggregated to land use regions (to be investigated still); LUZs and farms define an association; and we could identify the LUZs at the different points in time, which enabled us to investigate the dynamic behaviour of the LUZs.

The application of an object-structured data model implies that the data are stored per object. This provides opportunities to integrate data from different sources, dates, scale or resolution in a structured manner. In the conventional approach no structure for the integration of data is provided. We use data of different periods to determine the object characteristics of these periods. This means that we can compare or integrate the data from these different periods at object level. In the same way we can integrate data related to different scales by using this data for description of objects and relating the objects through aggregation. Also at the object level that we can integrate different data types or data from different data sources, because the different sources are used to describe the characteristics of the same object. The combined use of aerial photos of different type, scale and date, of remote sensing data of different dates and of tabular data (like farm survey data) to describe the LUZs might illustrate the opportunities for data integration. The opportunities for data integration, especially with respect to scale levels in land use inventory, should be further investigated and verified through implementation of the model.

The object-oriented approach also provides opportunities for methodological improvement, by making use of the mutual character of the relations between objects. For example, the composite objects inherits the attribute values of its elementary objects. Vice versa, does the composite object define a context of the elementary object. The context information can be used to better (more reliable) interpret phenomena related to the elementary object level. For example, the LUZ serves as a context for the interpretation of the land cover classes with respect to land utilization types. In the same way may the general LUZ class define a context for the classification of the LUZs at the lower level in the classification hierarchy. Think of the differentiated classification of LUZs that belong to the agricultural area and to the area of natural vegetation. The use of context information in this manner and the opportunities it offers for structuring the land use inventory process deserves additional attention.

With respect to describing land use change by means of object dynamic behaviour some difficulties were encountered. These pertain especially to the interdependent nature of the object characteristics and the intricate relation between the classification, aggregation and association structure of objects. The following observations were made:

1. In the FDS the object class defines a relation between objects on the basis of a common attribute structure. These can often be related to classes defined on the basis of corresponding attribute values. For example, the LUZ can be seen as an object class, such as the physiographic mapping unit, the farm, road and river. Each of these objects classes has its own attribute structure. All LUZs are described by the land cover composition. However, we can distinguish characteristically different land cover patterns and on that basis we can classify the LUZs. The classes refer to LUZs with corresponding values for land cover composition. However, the classes also denote specific LUPs (e.g. agricultural land, area of natural vegetation). The description of land use for each of these categories will require distinct attribute structures. They might therefore be considered to represent object classes. More so, because the aggregation structure (e.g. agricultural fields being present or not) and dynamic behavior of objects belonging to these different classes is also be characteristically different. The inclination to treat all LUZs as belonging to the same object class will hamper the adequate description (or modeling) of land use.

The distinction, in this case, must be sought in the different contexts the object characteristics refer to. The land cover data has relevance within the context of the system to the description of object characteristics, while the land use data has relevance within the users' context.

2. Change in object geometry has implications for the object thematic characteristics. This has to be taken into account in evaluating object dynamic behavior. For example, change in the LUZ boundary will affect the land cover composition. Therefore, change in the land cover composition can only be evaluated effectively if no change occurs in the object boundaries or if the effect of the change in LUZ boundary on the cover composition can be accounted for.
   Stationary characteristics are required to evaluate change. The object class determines which particular characteristics might be assumed to be stationary. The LUZs pertaining to the agricultural areas tend to exhibit stationary boundaries. This is not the case in areas of agricultural penetration, for instance. In the agricultural areas land use change can consequently be inventoried through change in the LUZ land cover composition. For monitoring change in the area of agricultural penetration, other procedures should be adopted. These examples demonstrates that the object dynamic behavior is related to the object class.
3. Object dynamic behavior, referring to existing objects, does not adequately provide for the description of changes as a consequence of new objects appearing or old objects disappearing. Split and merge (also called change in association structure) has been added to the types of change already defined (see Chapters 1 and 2, Part 1).
4. Different types of aggregation can be defined at the same level of aggregation. When these different types refer to the same class of elementary objects, than an association between composite objects of these different types can be defined through common elementary objects. An association of two or more elementary object classes is denoted by the composite objects class that contains these elementary object classes. The composite object in such case represents an association.
   For example, the relation between LUZs and farms is modeled through association. Both entities refer to the same level of aggregation; both are an aggregation of agricultural fields. Many farms may belong to one LUZ and vice versa many LUZs may belong to one farm. A LUZ may then be associated to a farm by the agricultural fields they have in common.
   LUZs that belong to the "area of agricultural penetration" are an example of objects that represent an association. The areas of agricultural penetration consist, on the one hand, of grasslands with the elementary objects referring to fields and, on the other hand, of forest and other areas of natural vegetation in which case the elementary objects refer to areas with homogeneous land cover.

The relations that exist between the object classes (of both composite and elementary objects), types of aggregation, types of association and object dynamic behavior, and the opportunities this offers for modeling of land use should be further investigated.

A DATA DRIVEN, KNOWLEDGE-BASED APPROACH TO THE RECOGNITION OF LAND USE PATTERNS

A third theme in this thesis is the use of satellite imagery (SI) and aerial photos (APs) for the inventory of land use and land use change. Because we adopted an object-based strategy, these materials are used to identify and classify objects: LUZs in this event. The identification and classification of the LUZs on the basis of these materials involves pattern recognition. With respect to the identification and classification of LUZs e.g. spatial patterns (or field patterns) and the land cover patterns were investigated.

The LUZs, however, were not defined at the beginning of the investigation. Only a generic object model was defined. Because the object definition is context-dependent, insight in the context is required. As regards to the object definition for land use inventory, insight is required in sub-regional land use and in the role of the objects in land use analysis and planning. These insight were lacking at the beginning.

The use of SI or APs seems conflicting. The conflict is due to the fact that these materials serves to provide the insights in the regional context, while these insight are required for its proper interpretation. This leads to an iterative procedure with feedback loops. This is illustrated by the procedure for land use inventory in the GRS area. First, the aerial photos were used to investigate spatial patterns. On the basis hereof the LUZs were identified. The LUZ characteristics are subsequently determined, which provides some insight in the sub-regional land use structure. This information is than again used to improve the aerial photo interpretation. Another example of an iterative procedure with feed back loops is given by the land cover classification.[1]

Our results demonstrate that remote sensing and aerial photography can be employed successfully as tools in land use inventory at the sub-regional level. However, equally important is the structuring and description of the land use inventory process. The process structure or logic is defined by the sequence of steps in the inventory process. We identified the following steps (see Part II):

1. Identification of the LUZ, i.e. delineation of the land use zones on the basis of the spatial (field) pattern. The land use pattern is assumed to be expressed in the field pattern. The AP interpretation serves to make an inventory of the occurring field patterns. On this basis the LUZ are delineated.
2. Land cover classification. The land cover classification serves to enable the quantitative description of the land cover characteristics of the LUZs.
3. Determination of the LUZ image characteristics (data extraction). This step concerns the land cover composition and field size distribution, both assumed to express the LUPs. The specific land use characteristics associated with the land cover composition and the field distribution are still to be determined.
4. Quality check. Statistical evaluation of the discrimination between land use zones as regards the characteristics mentioned under step 3 (analysis of variance, cluster analysis). Acceptance or rejection of the land use zone interpretation; in case of rejection, adaptation of the land use zone delineation if indeed interpretation error is concluded.
5. Definition of the data classes (e.g. composite land cover classes, mean field size classes) on the basis of the results of the cluster analysis conducted in step 4. To define the clusters, homogeneity criteria and criteria for separability (minimum distance) are specified.

[1] Information on land cover is inferred from spectral data: general information on land cover is obtained by using the SI; this information is aggregated to LUZ level; the land cover composition of the LUZ tells something about the occurring land cover types (or land utilization types); and this information is fed back to the original land cover classification to obtain information on the specific land cover.

6. Field and farm survey; definition of the farming systems and classification of farms.
7. Evaluation of the relation between measured image characteristics and field data. Specification of the information categories (land use patterns); specification of the mapping rules; land use zone classification.
8. Evaluation of results.

Much attention is devoted to class definition, being a very important element in the procedure. A "data-driven" approach is adopted to object classification. Also, a clear distinction is made between data classes and information categories in order to structure the classification process. The data-driven approach provides a solution to the classification of objects when classes are not a-priori defined. It is a strategic answer to the problem of the use of SI and AP stated earlier, namely that these materials are used to obtain insight in the context while this insight is also required for its proper interpretation. The classes are inductive: they result from the inventory process.

First, the LUZs are categorized in terms of data classes. The data classes refer to the image characteristics of the LUZ. The definition of the data classes is based on statistical analysis of differences between the objects (LUZs). As such, the classes reflect the degree to which objects can be distinguished on the basis of their characteristics (the classes represent the appropriate level of generalization). The quality of the input data has been taken into account, especially in determining minimum required distances between groups.[2] The set of data classes fulfils the two requirements: it must be exhaustive and the classes must be separable (see Chapters 4 and 5, Part II).

Secondly, the land use characteristics of the LUZs with corresponding data class values are investigated, and the LUPs (the information categories) are defined accordingly (Chapter 3, Part I). Finally, all LUZs are assigned an LUP on the basis of their data class values. The approach yielded LUPs that provided specific and accurate information on land use on a sub-regional scale. The assumption that field pattern and land cover composition are an expression of the land use pattern is confirmed.

In the literature little attention is devoted to the problem of defining appropriate classes, while this is necessary to obtain reliable and accurate classification results. The subject of class definition is generally not elaborated in reports. A data-driven (or unsupervised) approach to classification is applied in the field of remote sensing, where it concerns a per pixel classification. Such an approach has not yet been used to classify complex spatial objects.

By making a clear distinction between data classes and information categories, the context-dependent decisions could be made explicit. These decisions refer to the rules for mapping of the data classes to the information categories. The definition of these mapping rules is based on observations of whether objects belonging to the same data class represent meaningful information categories (Chapter 3, Part I). The rules in fact denote how the data (classes) are to be interpreted in the given regional and application context (e.g. interpretation of the spectral classes in terms of land cover; interpretation of the composite land cover in terms of land use pattern). Because we are not able, as yet, to describe this context formally and

[2] For example, the reliability of the land cover classification is expressed in the width of the composite land cover classes; The quality of the AP interpretation will influence the difference in field size characteristics between LUZ and herewith the definition of the mean field size classes.

adequately, expert knowledge is required to make these context-dependent decisions. In our case, expert knowledge is applied to the following tasks:

- The photo interpretation for the delineation of the LUZ;
- The definition of the appropriate land cover classes and the interpretation of the spectral classes with respect to land cover;
- The definition of land use patterns and the interpretation of the LUZ characteristics with respect to land use.

This expert knowledge is expressed by the mapping rules. These rules guide the conditional assignment of class labels to the objects. We could define a decision tree to structure a complex decision process in which many criteria need to be evaluated, such as the classification of LUZs. The decision tree imposes a hierarchical order on the application of the decision rules. Thus, what is provided is a strategy to support the expert in making decisions in complex situations for which no standard solutions are available.

The use of such decision rules is typical for knowledge-based (expert) systems. In recent years there has been increasing interest in developing knowledge-based systems to help automate remote sensing image interpretation or to analyze geographical data in general. Robinson and Frank (1987) provide an overview of expert systems for GIS. An example in the field of land evaluation is ALES (Rossiter, 1990). For the expert system approach to imagery interpretation, see Goodenough *et al.* (1987), Hadipriono *et al.* (1991), Skidmore *et al.* (1989), Srinivasan and Richards (1990), Wilkinson and Mégier (1990) and Wu *et al.* 1988. The approach generally consists of the application of context rules (also called decision rules, belief functions, or otherwise) for the interpretation of imagery, whereby information about the context is obtained from a GIS. Most of these approaches concern a per pixel based classification. Again, the application of an object-oriented approach to land use inventory is new. It is important to realize that different classification strategies exist and that there are different concept of data modeling. Attention should be devoted to criteria for selection of a particular strategy or concept, which will depend on the situation.

This thesis demonstrates that the process of inference of land use information from satellite imagery and aerial photos can be structured and effectively described. A clear process structure improves reproducibility of the results and enables the transfer of the methodology applied. Parts of the inventory process, namely the use of spectral data and aerial photo-derived data for the characterization and classification of the LUZs, can be described quantitatively (Chapters 4 and 5, Part II). Parts that concern the context-dependent classification of LUZs require expert knowledge; these parts are more difficult to formalize or describe quantitatively (Chapter 3, Part I). These parts involves the interpretation of data. The decisions can be described by mapping rules (Chapter 6, Part II) This, in turn, enables the transfer of knowledge and the verification of the decisions. The rules can be easily adapted, which will improves the flexibility of the system. In applying the LUZ approach to other regions, the mapping rules will need adaptation and tuning. The structure provided might also serve to evaluate existing land use inventories.

In this thesis, an onset is given to a knowledge-based system for land use inventory and analysis on a sub-regional scale. It provides an operational definition of land use, a data structure for the description of sub-regional land use, and an a method for the inventory of land use and land use change employing satellite imagery and aerial photos. The viability, flexibility and efficiency of the LUZ approach is to be proven by the implementation of the data model and land use inventory procedure.

APPENDIX A.

APPLICATION OF INDICES TO SPECTRALLY CHARACTERIZE GENERAL PURPOSE LAND COVER CLASSES: A CASE STUDY IN THE ATLANTIC ZONE OF COSTA RICA

1. INTRODUCTION

Chapter 3 of Part 2 discussed the spectral and spatial complexity of second generation satellite imagery. It is the source of the difficulty and tediousness of the task of locating and delineating training statistics with statistical properties suitable to the maximum likelihood classification. Reduction of the number of spectral features and parameters to describe the land cover would facilitate the definition of training classes. The question is, of course, whether the reduced set of spectral features allows discrimination between the cover classes and whether the features are related to relevant cover characteristics. The selection of the spectral features as well as the relevant parameters to describe land cover requires insight in cause-effect relationships concerning spectral response.

For the investigation of the spectral characteristics of remotely sensed imagery one should review the work done in the field of feature extraction. Various studies indicate that the spectral reflectance pattern of land cover in Landsat-TM is basically three dimensional. The application of indices representing these three spectral dimensions is tested for general-purpose land cover mapping in the Atlantic Zone of Costa Rica.

As regards the factors explaining spectral response, much work has been done on modeling canopy reflectance. This work pertains to the development of vegetation indices, e.g. for crop monitoring, as well as to the strides made in modelling vegetation canopies to predict spectral response. The models developed so far require too much specific information to be of use for general purpose mapping. Furthermore, the description of agricultural crops in optical and microwave remote sensing studies, proposed by Cihlar *et al.* (1987), is too detailed for general-purpose mapping. However, the insights developed in these disciplines could be used to improve existing procedures for land cover classification.

The aim of the present study is to investigate ways to use indices (or spectral features) for general-purpose land cover mapping. Study site is the Atlantic Zone of Costa Rica. The spectral data are from the Landsat Thematic Mapper (TM) satellite data. The scene recoding date was February 6, 1986.

2. FEATURE EXTRACTION AND SPECTRAL CORRELATION

For an explanation of the spectral variation, one should refer to work on feature extraction and spectral correlation. Feature extraction is done to transform the spectral data into spectral features in order to obtain a better descriptive basis for the spectral signatures (Mulder, 1988). Mulder shows that the spectral variation can be decomposed into intensity variation and two mixing factors: the spectral mixing of bare soil and vegetation, and the equivalence of mixing water and bare soil spectral reflectance (equivalent to the three principal components, or PCs). The intensity relates to structural information, mainly through shadow and shades (Mulder mentions geomorphology for example). The color picture (defined by hue and saturation) contains 'spectral reflectance' information.

Kauth and Thomas (1976) and Crist and Cicone (1984) define three spectral features which can be directly associated with physical scene characteristics. They are defined as brightness, greenness and wetness. Brightness is responsive to changes in total reflectance and to the physical processes which influence the total reflectance. Soil brightness is measured by the brightness index. Particle size and soil surface roughness are mentioned as factors affecting the soil brightness.

Greenness (determined by the reflection in the NIR in contrast to the sum of the visible bands) is reported to be correlated to canopy closure, green leaf area index (GLAI) and fresh biomass. GLAI can also be estimated from the spectral mixture of the vegetation and the underlying soil (Mulder, 1988).

Crist and Cicone (1984) mention that soil moisture status was found to be the primary characteristic expressed in wetness, which seems to be in agreement with Mulder (1988), who mentions the spectral mixing of water and bare soil as the second mixing factor. Given the high weights of the middle infrared bands and the sensitivity of these bands for leaf water content (Tucker, 1980), the wetness index will also be responsive to leaf water content and plant water stress, aside from being responsive to soil moisture alone.

The spectral reflectance patterns of land cover are thus basically three dimensional. This is also indicated by the first three principal components. These often explain more than 95 % of the spectral variance and show clear correspondance with the three indices mentioned. However, the principal components are highly dependent on the scene characteristics and the proportion of different cover types in the scene.

The above suggests a correlation between indices and specific cover characteristics. Brightness, intensity, or the first principal component all reflect intensity phenomena. The intensity is determined by structural properties, like relief for terrain and surface roughness for soils. Structural properties can also be defined for vegetation canopies, e.g. canopy surface roughness as determined by the size and spacing of the vegetation elements.

Greenness, the bare soil vegetation mixing factor, and often the second principal component (depending on the scene characteristics) reflect the amount of green vegetation. Aside from greenness, many indices have been developed expressing the same type of phenomena. Linear combinations of red and photographic infrared have been denoted to correspond to green leaf area index (GLAI) or biomass, generally referring to a specific crop or vegetation type (Tucker, 1979; Curran, 1983; Clevers, 1986; Plummer, 1988). Other authors report better correspondence with the GLAI when the middle infrared (MIR) is included in the band combination (Curran and Williamson, 1978; Williamson, 1988). However, the MIR does not contribute to greenness or to the second PC in general, but it has a strong weight in the wetness index and generally in the PC3. This means that greenness or vegetation indices based on only the red and NIR bands are not exclusive indicators of green biomass. The significance of both indices as such is still undetermined.

There are not many examples of indices being used for classification purposes. Haas (1985) used brightness and greenness values instead of the original spectral bands to represent cover types. The vegetation types were not defined in terms of properties explicitly related to the brightness and greenness values. The use of indices is then more for the purpose of data reduction.

3. MODELLING VEGETATION CANOPIES

Models of vegetation canopies have been made for predicting spectral response. The objective is to estimate biophysical parameters from reflectance data through model inversion. Initially optical models for homogeneous vegetative canopies were developed (Suits, 1972). The parameters of the model can be related to solar zenith angle and canopy parameters such as leaf area index (LAI), leaf angle distribution (LAD), and hemispherical leaf reflectance and leaf transmittance. Variation in directional reflectance of vegetative canopies is also a function of azimuthal viewing angle.

For forest canopies, similar properties are considered to determine the interception and reflection of solar radiation. Peterson (1987) lists the following six structural properties:

- Leaf area index;
- Vertical distribution of foliage;
- Distribution of leaf inclination angles;
- Leaf transmittance, reflectance and scattering properties;
- Grouping or clumpiness of the foliage;
- Distribution of the leaf azimuthal angles.

These parameters are far too specific to described land cover classes for general-purpose mapping. Distribution of leaf angles will be specific to plant species or vegetation types. The vegetation type might, therefore, serve as a general indicator of LAD as well as the vertical distribution of foliage. Leaf transmittance and reflectance properties have proven to be dependent on internal leaf structure. In this respect discrimination between monocotyledons (grasses and cereals) and dicotyledon plants and vegetation types is relevant (Clevers, 1992).

Franklin (1986) names basal area and leaf biomass as the structural parameters for forest canopies. She concludes that these characteristics are expressed primarily in the visible LANDSAT-TM bands 2 and 3, rather than in the TM band 4. She also states that the mid-infrared channels do not show a relationship with varying stand densities, whereas Butera (1986) concludes that the Thematic Mapper bands 1, 5 and 7 (the blue and mid-infrared bands) are most significant in relating canopy closure to spectral response.

Three-dimensional models for heterogeneous canopies describe vegetation elements as discrete objects. A geometrical optical model of a conifer forest canopy with snow as background was presented by Li and Strahler (1985). The canopy is modeled as an assemblage of large solid three-dimensional objects. Size and spacing of the objects and reflectance of sunlit and shadowed parts of the object and the background determine total reflectance. The spectral characteristics are expressed in brightness and greenness values. Brightness values decrease as trees are added, until a certain density is reached. Thereafter, overlapping occurs and the proportion of shadowed parts increases no further. Greenness values rise with increasing tree density. A similar approach is followed by Jupp *et al.* (1986) in modelling the vegetation structure of semi-arid eucalypt woodlands.

Goel and Grier developed a model for heterogeneous three-dimensional canopies, which is in fact based on Suits' canopy reflectance model. The model is most applicable to canopies whose architecture can be described by a repetitive pattern and to homogeneous fully covered canopies. Parameters are related to vegetation canopy spacing and size of the elliptical subcanopy, in addition to earlier-mentioned parameters of the Suits' model.

In general terms, these properties related to size and spacing might be described by the height (and variation therein) of the vegetation and by density of vegetation elements. The following general characteristics are proposed for the description of vegetation types:

- <u>Cover type</u>: This denotes whether the cover class belongs to vegetated area, non-vegetated area (bare soil, rock, built-up area) or water bodies, with reference to the associated characteristic shapes of the spectral curve.

  Concerning vegetative cover, further subdivisions are made according to vegetation type as defined by the characteristic geometry and with reference to the internal leaf structure. Thus, general vegetation types such as woody (or tree-like) vegetation, grass, herb and shrub vegetation are differentiated. For example, banana and banana-like vegetation (plants belonging to the family of the *Musaceae*, which is abundant in Costa Rica) is to be considered a separate cover, type due to the specific geometric characteristics (leaves of more than 2 meters in relation to plant height of 4 to 5 meters). Also palm trees, because of their specific geometry and internal leaf structure (a monocotyledon), constitute a separate cover type.

- Surface roughness: This relates to the variation in height of the surface (surface relief). It is relevant both to the non-vegetated surfaces and to the vegetative surfaces. It is intended as a measure for the amount of shadowing present. The variation in height should be evaluated in relation to the distance over which the change present itself.
- Density aspects: For homogeneous canopies, density can be interpreted as GLAI or biomass. For heterogeneous land cover types (e.g. woodlands), it can be interpreted as canopy closure. Density determines the influence of the background reflectance (e.g. bare soil or grass vegetation) on the spectral response. It is related to the spatial distribution of the elements within a composite land cover class.

The way in which the above properties, i.e. roughness and density, are measured depends on the cover type. For forest cover types, a measure of the surface roughness can be measured in terms of the distribution in stem diameter, rather than through the difficult estimation of crown heights. For grass vegetations, an indication of surface roughness can be obtained through direct measurement of maximum and minimum height of the vegetation.

A measure of vegetation density can be obtained by simply counting the number of trees or shrubs per unit of area. Measuring crown diameters and calculating the relative area (projection of the crown area) of the vegetative layers is less relevant to general inventories. For grass vegetation, biomass could be measured as an indication of the vegetation density; alternatively simple measurement of the height of the vegetation might suffice.

In the present study, the spectral values of the cover types are obtained through sampling of the image for known location and averaging for the corresponding pixel. The interpixel spectral variance, as an expression of spatial phenomena, is not considered. The average values are assumed to denote the reflectance of the cover type, representing the sum of reflectance of the individual cover components weighted according to their respective proportion in the area.

4. METHODS: BAND COMBINATIONS TESTED

The band combination listed in Table A.1 are evaluated for their discriminative power concerning the land cover classes, as well as for their correspondance with cover characteristics. The different band combinations were taken from the literature.

They constitute a brightness index, being the simple sum of bands and the first principal component, which will be discussed later. Furthermore, they form combinations expressing either the contrast between the NIR with the visible (like the second principal component, the ratio of TM3 and TM4 or the normalized difference of these two TM bands [the normalized vegetation index, NVI]) or the contrast between the MIR and the sum of the visible and the NIR (like the third principal component, the ratio of TM4 and TM5, or the normalized difference of TM band 4 and TM band 5 [NORMDIF4-5]).

Table A.1. *Band combinations tested for informative value as regards to the mapping of vegetative cover classes*

Name	TM Band or Band Combination
Sum Bands	TM1+TM2+TM3+TM4+TM5+TM7
Red	TM3
NIR	TM4
Red/NIR	TM3/TM4
NVI	(TM4-TM3)/(TM4+TM3)[1]
MIR	TM5
NIR/MIR	TM4/TM5
ND NIR-MIR	(TM4-TM5)/(TM4+TM5)[2]
	(TM4-TM3)/(TM5+TM7)[3]
PC1	
PC2	
PC3	

For 66 locations corresponding to various cover types, the spectral values were determined through sampling of the Landsat-TM scene. The values for the linear combinations are derived from these. In general, multiple samples were taken for the same cover type, representing small differences in canopy structure or density of the vegetation.
Testing was subsequently done by:

- Listing the cover samples in order of increasing index values for each of the band combinations to see whether a logical order was induced;
- Visual interpretation of the scatter plots; and
- Cluster analysis.

The index or set of indices should permit discrimination between the cover types as defined above. It should also provide for discriminative power within the cover type for differences in surface roughness and vegetation density. For example, the grasslands should be recognizable by their index values as a separate cover type, but the indices should also portray the differences between the grassland vegetation types. Whether indices reflect differences between grasslands is evaluated by listing the index values for all the samples belonging to the same general cover type and investigating the corelation with general vegetation parameters.

[1] Tucker, 1979

[2] Williamson, 1988

[3] Williamson 1988

5. RESULTS AND DISCUSSION

5.1. Principal component transform

The first three principal components (PCs) corresponding to the Landsat TM scene of the Atlantic Zone of Costa Rica are shown in table A.2. The principal components transform is calculated on basis of the covariance matrix.

For calculation of the principal components, part of the scene was selected, excluding the Atlantic Ocean. The area is predominantly flat, with forest covering about 40 % of the area. Wooded area and grassland each covered about 15 %. The remaining part was divided among bare soil or sparsely vegetated area (and built-up area to a limited degree), banana plantations and land cover categories. Water bodies represent only a small percentage of the area.

The first principal component explained 68.8 % of the total spectral variation of the scene. That is most of the spectral variation corresponds to brightness phenomena, being a weighted summation of the spectral bands. The second principal component explained 22.8 %. It is a contrast between the near infrared band and the visible and middle infrared bands. It deviates from the greenness of the tasseled cap by the negative load of the middle infrared bands. The third principal component explains 6.1 % and is a contrast between the middle infrared bands and the NIR and visible bands. This is more or less in agreement with the wetness index of the tasseled cap transformation.

Table A.2. *Eigenvectors corresponding to the first, second and third principal component (PC).*

TM band no.	PC1	PC2	PC3
1	0.176	-0.327	0.671
2	0.150	-0.159	0.322
3	0.193	-0.315	0.338
4	0.598	0.749	0.248
5	0.686	-0.307	-0.507
6	0.094	-0.098	-0.072
7	0.269	-0.321	-0.150
Eigenvalues	780.8	258.8	69.1
Variance %	68.8	22.8	6.1

5.2 Significance of indices for the discrimination of land cover classes

Brightness shows characteristic values for the different vegetative covers (see Figure A.1), which seems to be related mainly to the form or structure of the vegetative surface:

- The forest training samples show low brightness values;

- Secondary forest, tree plantations and densely wooded areas show increased brightness values;
- Banana plantations and dense scrubland have intermediate brightness levels;
- Shrub/grasslands and secondary vegetation with little variation in the height of the vegetative surface have medium to high brightness values;
- Pasture with very smooth vegetative surfaces show the highest values.

Brightness does not differentiate vegetation types with more or less the same characteristics for the vegetative surface roughness, such as certain grasslands, secondary vegetation and semi-perennial crops like yucca (*Manihot esculenta*). Other band combinations should permit discrimination of these classes.

The combinations which contained only the NIR and/or the red reflectance band did not give satisfactory results. Vegetated and non-vegetated areas could clearly be distinguished, but within the vegetated areas discrimination was poor. The index values for vegetation classes like bamboo, banana, ornamental crops, grassland and secondary vegetation could be found within the same range.

This is explained by the indices responsive to the mixing of bare soil and vegetation. Vegetation indices like TM band 4 or the normalized vegetation index (NVI) provide the best discrimination in the low range of the LAI or at low biomass levels. This can be concluded from the reflection curves of the red and NIR reflectance in relation to LAI (Curran, 1983). Investigation of the spectral variance for the cover classes showed that the variation of NIR reflectance within the classes is too high in comparison to the overall variation in NIR reflectance (or to the between-group variance) to allow discrimination of the cover types.

In the Atlantic Zone, areas are either bare or completely covered with vegetation. Mixed forms are of minor importance. For complete vegetative covers, factors other than LAI or biomass obviously play an important role in determining the reflectance in the NIR. LAI or biomass have not been measured, but clear differences are expected between the cover classes defined.

With respect to a specific crop or vegetation type (i.e. with invariable geometric characteristics), the NIR might denote clear difference in vegetation characteristics. This was shown in a study concerning the yield prediction of banana, where the NIR proved to be the best predictor of yield (Veldkamp *et al.*, 1990).

This suggests that a hierarchical approach should be followed, in which the cover type should first be identified before LAI or biomass can be estimated. It leaves the problem of the identification of the cover types unresolved.

Better results were obtained when the MIR spectral bands were included in the band combination. TM band 5 differentiates between the vegetated and non-vegetated areas, as well as between grasslands and other vegetation types (e.g. secondary vegetation and (semi-)perennial crops). The TM band 5 is less sensitive to internal leaf structure and possibly to differences in plant geometry, given the high absorbance of energy associated with leaf water content. The differences between grassland and the other vegetation types as regards their reflectance in the MIR might be explained by several factors: the grasslands are more susceptible to drier climatic circumstances due to the influence of the direct incident radiation, the more direct influence of winds, and the limited rooting depth; all these characteristics influence leaf water content. But another possible explanation might be the inclusion of dry grass in the grassland vegetation types.

The MIR in combination (or in contrast) with the NIR might then reflect the geometric

characteristics of the vegetation and density aspects of the vegetation. Combinations of NIR and MIR differentiate between the following (see the scatterplot in Figure A.1.):
- Vegetated and non-vegetated areas (as do the TM band 1,2 and 3 or the vegetation indices, though not so the brightness index);
- Grassland vegetation and woody and shrub vegetation, (which also reflects the gradual changes in proportion between the two);
- Different types of woody and shrub vegetation (based on the geometric aspect of the vegetation).

With respect to the woody and shrub vegetation, clear distinction is found between:
- Vegetation types with a closed and smooth canopy and relatively large leaves (like banana and the ornamental crop *Draceana massangeana*) which show a high reflectance in the NIR and low reflectance in the MIR;
- Forest vegetation types showing low values in both the NIR and MIR reflectance;
- Woody and shrub vegetation like secondary vegetation and bamboo, which show a large variation from medium to high response in the NIR but have medium values for MIR reflectance.
- Grassland vegetation also reveals marked differences in the NIR (varying from medium to high values) but shows high reflectance in the MIR.

The scatterplot in Figure A.1 shows the brightness values and the values for the NORMDIF4-5 of the land cover samples. The correlation between the brightness values and NORMDIF4-5 is very low: an R-square equal to 0.09, which confirms the independence of both dimensions.

The importance of incorporating TM band 5 in the band combination is obvious. This is in agreement with Williamson (1988), who stresses the importance of TM band 5 for vegetation monitoring. However, the data does not suggest a correspondence of green leaf area index (GLAI) or biomass to spectral features incorporating TM band 5, irrespective of the cover type. The data seems to confirm findings of the Butera (1986), who found that the MIR bands were most significant in relating canopy closure to spectral response.

The relevance of different combinations to the mapping of land cover was tested by means of a hierarchical cluster analysis, based on the normalized Z-values for Brightness and the different band combinations. The results were considered satisfactory when the analysis resulted in the grouping of related land cover types according to type and structure of the vegetation. For the following band combinations, satisfactory and comparable results were obtained:
- The NIR divided by the MIR (TM4/TM5);
- The NORMDIF of the NIR and the MIR [(B4-B5)/(B4+B5)];
- The NIR minus the Red divided by MIR [(B4-B3)/(B5+B7)].

5.3 Differentiation within the major cover types

Forest cover

Clearly deviant vegetation characteristics also exist within the major cover types. For forest vegetation types, a difference is found with respect to the canopy structure. It is reflected in the deviating values for the estimated height of the vegetation and in the values

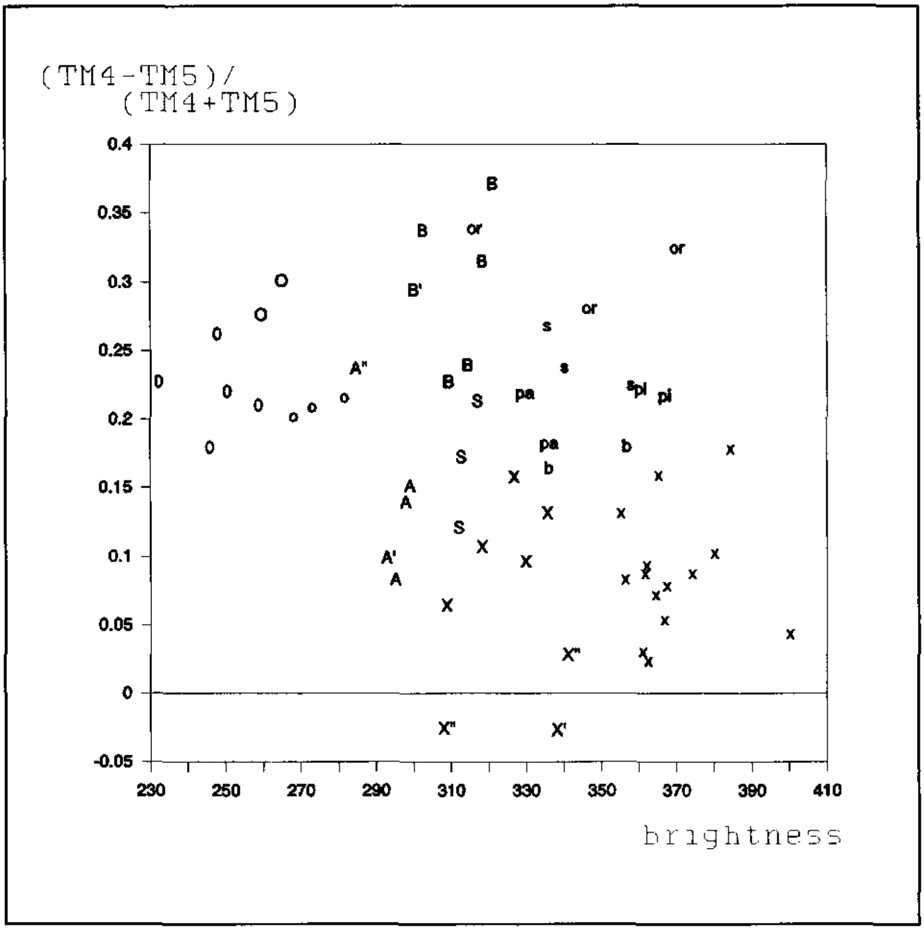


Fig. B.1 *Scatterplot showing the brightness values (summation of the six spectral TM bands) and the values for the band combination (TM4-TM5)/(TM4+TM5) of the vegetative land cover classes.*

Explanation of the symbols:

0	Forest area
O	Forest with dominance of palm vegetation
o	Secondary forest
A''	Pejivalle plantation (*Bactris gasipaës*)
A	Densely wooded area
B	Banana
B'	Plantain (*Musa AAB*)
or	Ornamental crop (caña india, *Dracaena massangeana*)
pa	Palm heart plantation
pi	Coconut plantation (dwarf variety)
s	Secondary regrowth (herbs and shrubs)
S	Secondary woody vegetation
x	Pasture ('estrella','ratana','natural', etc.)
X	Neglected pasture
X'	Grass vegetation in and near lagoons
X''	Pasture (guinea) both well kept and neglected

for stem diameter (see Table A.3.a). The TOR329 sample represents a primary forest type, periodically inundated. It is characterized by the presence of high emergent trees. Given the stem diameter of 120 cm, the maximum height of the vegetation might be somewhat underestimated. TOR429 represents a forest type with a more evan canopy structure. The emergent trees have been cut in the past. For3 represents a secondary forest with a rather smooth canopy surface.

Table A.3.a. *Listing of the values for brightness, PC1, normalized difference of band 4 and band 5, and PC3 for the forest areas.*

Training samples	Height of canopy (m)	Species	Diameter stem (cm)	Sum bands	PC1	TM4-TM5/ TM4+TM5	PC3
FOREST AREAS							
TOR 329	15 - 35		10 - 120	232	107.7	0.23	40.6
FOR 1	15 - 35		- - -	250	119.7	0.22	37.0
TOR 428	10 - 35		19 - 80	259	121.9	0.21	40.0
TOR 429	15 - 25		10 - 35	273	130.7	0.21	39.1
FOR 3	12 - 20		12 - 20	281	137.2	0.22	36.5
TOR 228	9 - 10		8 - 15	268	122.8	0.20	44.4
TOR 328	8 - 10	yolillo	- - -	265	123.3	0.30	47.6

The different vegetation characteristics seem to be reflected in the brightness values presented in Table A.3.a. The best correspondence is found between PC1 and the maximum height of the canopy (R-square equals 0.77). A similar correlation is obtained using the brightness values (R-square equals 0.75). Taking the difference between minimum and maximum height of the vegetative surface as a measure for surface roughness, a lower correlation was found (R-square equals 0.53). The minimum height varies little and is a rather arbitrary measure. It is therefore better to leave this out. Given the small sample population, the figures should be considered as a mere indication of possible correspondence.

The lower brightness values in this case are thought to be related to increased shadowing as a consequence of uneven canopy surface. The optical depth of the canopy of these forest cover types is assumed not to vary signifanctly.

Two samples (TOR328 and TOR228) deviate from the pattern, showing lower brightness values than expected on the basis of the small difference between maximum and minimum vegetation height. Both samples represent cover types, deviating in density and type of vegetation and may therefore not be compared to the other forest samples. Both vegetation types are found in peat areas. TOR328 is a rather open vegetation of palm trees (yolillo, *Raphia taedigera*). TOR228 corresponds to a stand of poorly developed trees (*Anacardium excelsum*) with a very open canopy structure. The open structure will increase the amount of shadowing and therefore reduce total reflectance. Both forest types have a grass cover as understory. Strong absorbance of the infrared radiation by a water-logged soil in comparison to the other forest types, which might explain the lower brightness values, is not observed.

The relevance of the NORMDIF of TM4 and TM5 or the PC3 to denote differences in type of vegetation is illustrated by the deviant values obtained for TOR228 and especially for TOR328 in comparison to the other forest types. The far more open and thin vegetation of TOR228 is reflected in lower values for the NORMDIF4-5.

The yolillo vegetation of TOR328 is characterized by relatively high values for the PC3 and the NORMDIF4-5. In this respect, the high value of the NORMDIF4-5 for the mature stand of pejivalle trees (0.24, denoted as A'' in Figure A.1) is remarkable in comparison to the other wooded areas with comparable density of trees (index values of 0.08, 0.14 and 0.15, samples denoted as A in the figure). Pejivalle (*Bactris gasipaës*) is also a palm tree (monocotyledon).

In general the transition of forest area to the less dense wooded area is reflected in lower values for the NORMDIF4-5 and lower values for the PC3.

Grassland vegetation

For grass vegetation a good correspondence was also found between brightness and height of the vegetation. The following sequence shows increasing brightness values:

- Wooded grasslands;
- Grasslands invaded by shrubs and herbs with height of vegetation up to 3.50 meter;
- Clumped grasslands (grasses growing in tussocks) reaching up to 2 meters as a result of poor maintenance and under grazing;
- Grasslands of medium height with somewhat irregular vegetation surfaces;
- Grasslands of short grasses and well-grazed pastures.

The grassland cover samples and their corresponding values are listed in Table A.3.b. The brightness shows a slightly better correlation with the maximum height of the vegetation than with the PC1 (R-square of 0.79 versus 0.74). Also here the difference between maximum and minimum vegetation height shows a lower correlation (R-square of 0.68). The fact that brightness decreases with increasing height of the vegetation might be explained as a possible effect of shading. The lower brightness values might also be explained as an effect of increasing optical depth of the canopy being associated with the height of the vegetation.

For grasslands, the NORMDIF4-5 provides more information than the PC3 with regard to the differentiation based on the dominant grass species. Guinea (*Panicum maximum*) clearly shows lower values than the other species. Taller grass species like 'king grass' and 'cola de venado' (*Pennisetum purpureum* and *Andropogon bicornis* - Pas 20, Pas 21, Pas 12 and Pas 06) show higher values than pastures characterized by the presence of 'estrella' (*Cynodon nlemfluensis*) or 'ratana' (*Ischaemum ciliare*) and 'natural' (*Axonopus compressus*). There is no explanation for the high value of Pas04. It might be due to the influence of soil background reflectance.

6. FINAL REMARKS

This investigation took the initiative to use indices for general land cover classification purposes, with the aim to make the process of training statistics more profitable. As regards the application of indices in the Atlantic Zone of Costa Rica, the following conclusions or remarks can be made:

1. Spectral differences between land cover classes are best expressed by the brightness index and the band combinations based on the contrast between the near infrared (and visible TM bands) and the middle infrared bands.
2. Vegetation indices discriminate between vegetated and non-vegetated cover types. They serve less to discriminate between the general vegetation land cover classes.

Table A.3.b. *Listing of the values for brightness, PC1, normalized difference of band 4 and band 5, and PC3 for the forest and grassland training samples.*

Training samples	Height (cm)	Species	Sum of bands	PC1	TM4-TM5/ TM4+TM5	PC3
GRASS VEGETATION						
Pas 23	25 - 150	guinea (neglected)	309	141.9	-0.03	32.0
Pas 19	230 - 300	jamalote	309	143.8	0.06	36.4
Pas 20	50 - 200	king grass	327	157.0	0.16	37.7
Pas 21	100 - 250	cola de venado	329	157.0	0.10	34.5
Pas 12	50 - 200	king grass	318	151.6	0.11	35.3
Pas 06	25 - 150	king grass	336	160.0	0.13	37.8
Pas 18	100 - 150	jamalote	339	155.4	-0.03	34.4
Pas 02	- - -	guinea	340	157.8	-0.02	32.0
Pas 22	- - 100	guinea	342	161.9	0.03	32.2
Pas 11	40 - 70	natural/ratana	359	168.0	0.08	38.0
Pas 08	40 - 80	estrella	361	166.6	0.03	38.2
Pas 01	20 - 60	natural/ratana	362	173.2	0.09	35.9
Pas 15	17 - 40	natural/ratana	362	174.3	0.09	33.8
Pas 10	65 - 80	estrella	364	170.5	0.07	39.2
Pas 03	10 - 15	natural/ratana	365	178.3	0.16	37.6
Pas 14	3 - 25	- - -	366	170.8	0.05	38.0
Pas 07	50 - 90	estrella	367	172.1	0.08	39.3
Pas 24	- - -	estrella	374	178.2	0.09	37.2
Pas 05	4 - 25	natural/ratana	380	184.2	0.10	34.0
Pas 04	- - 10	- - -	384	188.2	0.18	39.5

Differences in index values cannot be correlated with the general cover class.

3. Of the total spectral variation 68 % corresponds to brightness phenomena. The intensity of the total reflectance is governed by structural properties of the canopies or of the land surface. The difference in brightness values between the cover classes suggests a correlation with the structure of the vegetation. One can think of the shadowing induced by difference in size and height of the vegetation elements (vegetative surface roughness) as well as of the optical depth of the vegetation canopies. Brightness values were shown to be related to height of the vegetation for grasslands. The existence of such a relationship was also suggested for forest cover types.
4. The explanation for the response concerning the band combinations of TM band 4 and TM band 5 (or related feature combinations) proves to be complex. The data give some indication of dependence on vegetation or plant type (i.e. difference between palm trees and other trees, between grasses and other herbs, as well as between different grass species). One would also expect a relation with biomass or LAI, given the results of earlier studies. The general description of land cover types in relation to their spectral expression, in the context of the general-purpose land cover classification, should be investigated further. In this respect, attention should also be devoted to a hierarchical approache to the classification of land cover (i.e. the use of different spectral features at different levels of detail, corresponding to classes of different generalization level).

APPENDIX B

THE VARIANCE OF DIFFERENCE AS A MEASURE FOR A PRIORI FEATURE SELECTION

1. INTRODUCTION

In the initial phase of a classification process, spectral bands are selected for the production of color composites for visual interpretation and selection of training fields. The bands should be selected a priori. The feature selection aims at minimizing the loss of information while reducing the number of features. In the case of image classification, the aim of feature selection is to find a subset of bands which will provide an optimal trade-off between probability of error and cost of classification (Swain and Davis, 1978). When the spectral classes are known, the effectiveness of the separate sets of spectral bands can be evaluated by measuring the statistical distance between the spectral classes. However measures, such as the average pairwise transformed divergence (Swain *et al.*, 1971) might not be permitted by the available image processing software or statistical packages.

Measures for the estimation of the information content of band combinations are then derived from the spectral characteristics of the scene, i.e. from the variance-covariance matrix (hereafter simply called covariance matrix), corresponding to the total scene or a representative part of it.

2. THE INFORMATION CONTENT OF INDIVIDUAL SPECTRAL BANDS

Sadowski (1985) used the variance of a spectral band as a measure of information content. He assumed that a wide range of data values will occur in bands which are particularly responsive to variation in land cover. More recently, Mulders (1992) used standard deviation as a measure for single band information content.

The band variance might be a poor estimator when based on digital numbers (DN), since the digital values are obtained by calibrating and scaling spectral radiance measured at the satellite for each wave bands. When using the standard deviation as a measure of information content, the ranking of the individual wave bands with respect to information value would depend on whether the reflectance phenomena are expressed in DN, spectral radiance (SR), or planetary reflectance (PR) (see Table B.1). TM band 5 would have the highest information content when determined on the basis of digital numbers, but it would rank fifth when the values are expressed as spectral radiance, while ranking second when using planetary reflectance. The thermal band (TM band 6) was excluded from the analyses.

Irrespective of the type of image data, the question remains whether standard deviation (or variance) serves as an appropriate measure of information value. The information value of a spectral band for classification purposes is determined by the ability of the band to discriminate between land cover classes (discriminative power or single band class separability). It can be expressed by the 'within group sum of squares (WGSS) divided by the between group sum of squares (BGSS)', whereby the group stands for a spectral class. Lower ratio values indicate better separability. The values are listed in Tabel B.1 for the 50 spectral classes corresponding to vegetated and non-vegetated cover types, and for a reduced set of 44 classes corresponding only to vegetated cover types.

For the correlation between standard deviation and WGSS/BGSS ratio, a R-square of 0.81 was obtainedusing DN. An R-square of 0.90 was found when the reflectance was expressed in PR. The rather high correlation between the standard deviation (SD), expressed in both DN and PR, on hte one hand, and the discriminative power (DP), on the other, seems to confirm the assumption that variance is an appropriate measure of information content.

Table B.1. *Statistical data on spectral bands and spectral classes*

	Band1	Band2	Band3	Band4	Band5	Band7
Image data						
Digital Numbers (DN)						
Mean	67.97	26.73	24.44	89.33	70.27	21.84
St. Dev.	6.17	4.51	6.93	18.79	18.88	9.06
Spectral Radiance						
Mean	3.928	2.847	1.860	7.086	0.736	0.116
St. Dev.	0.370	0.486	0.561	1.522	0.208	0.054
Planetary Reflectance						
Mean	8.7E-2	6.7E-2	5.1E-2	29.4E-2	14.2E-2	6.3E-2
St. Dev.	0.82E-2	1.25E-2	1.55E-2	6.32E-2	4.02E-2	2.99E-2
Coeff. of Variance	0.09	0.17	0.28	0.21	0.27	0.41
Spectral classes						
Mean St.Dev.(DN)	2.00	1.36	1.73	5.44	4.21	1.92
WGSS/BGSS (N=50)	0.123	0.127	0.098	0.055	0.081	0.098
WGSS/BGSS (N=44)	0.268	0.107	0.134	0.083	0.072	0.085

However, the WGSS/BGSS ratio is not a very accurate measure for single band discriminative power. First of all, because the ratio is an average value. Secondly, the WGSS/BGSS value will decline with increasing distance between groups, also beyond the point where complete separation of the spectral classes is obtained. At that point it contributes no further to the improvement of discriminative power. The outcome of the WGSS/BGSS ratio is strongly influenced by the spectral difference between the vegetated and non-vegetated areas. The latter is represented by classes like 'built-up areas' and 'bare soil' and 'water bodies'. The classes represent widely deviating spectral patterns (large distances between groups). Moreover, the latter cover classes represent only a minor part of the area. They are the less relevant categories of the land cover in the Atlantic Zone of Costa Rica.

Therefore, attention is devoted to the discrimination of the land cover classes within the vegetated areas. When classes like 'built-up area','bare soil' and 'water bodies' were discarded, a considerably different outcome was obtained for the WGSS/BGSS (see the outcome for the reduced set of 44 observations). For the vegetated areas, TM band 5 proves to contain the most information instead of TM band 4; the relative importance of TM band 7 and TM band 2 is also greater. The correlation between SD and the WGSS/BGSS ratio drops considerably (R-square = 0.29). This is explained by the fact that a higher spectral variance for the total pixel population coincides with a higher spectral variance for the spectral classes corresponding to the cover classes. The standard deviation (or overall variance) is therefore not considered a reliable indicator of single band discriminative power.

The above suggests adopting the coefficient of variation as a measure to evaluate spectral band information value. Considering the reflectance measurement in the separate bands as independent, the coefficient of variation would be the appropriate measure for comparison of the spectral features. The coefficient of variance is also shown in table B.1.

The coefficient of variance does yield a higher correlation with the WGSS/BGSS than was obtained with the variance. But the variation in the WGSS/BGSS values explained by the coefficient is rather low (R-square = 0.45, N=44). The low correlation is the result of a rather deviant coefficient of variance obtained for TM band 7. The coefficient of variance does not provide an appropriate measure for single band discriminative power. This is because the WGSS and the BGSS are both correlated with the value level of the spectral band.

3. POTENTIAL INFORMATION CONTENT OF BAND COMBINATIONS

Measures and methods

The difficulty of selecting the optimum band combination lies in evaluating the information content of the individual spectral features in combination with the correlation between the features. Sheffield (1985) demonstrates that the total variance contained in the band combination is not an adequate measure of information content.

Ranking methods, based on the variance present in each feature and the degree of correlation between features, have been applied to select feature combinations (Lillesand and Kiefer, 1987, who mention using such an approach for the selection of band ratios). Mulders *et al.* (1992) uses only a 'correlation index' to select the feature combination with the highest discriminating potential. This 'correlation index' consists of the average correlation of the band pairs constituting a feature combination.

The discriminative power of feature combinations can be evaluated in terms of the average pairwise divergence (APD) or the average pairwise transformed divergence (APTD). The pairwise divergence between class pairs "i" and "j" is determined as follows (Swain *et al.* 1971):

$$D_{ij}=0.5\mathrm{tr}[(C_i-C_j)(C_j^{-1}-C_i^{-1})]+0.5\mathrm{tr}[(C_i^{-1}+C_j^{-1})(M_i-M_j)(M_i-M_j)^T]$$

where C is the class covariance matrix, C^{-1} is the inverse of the covariance matrix, M is the mean vector, T refers to the transpose of matrices, and tr is the trace of the matrices.

The transformed divergence is defined as:

$$TD_{ij}=2000[1-\exp(-D_{ij}/8)]$$

The measure reaches 2000 when the probability of correct classification saturates at 1. See Singh (1984) for a discussion on the pairwise divergence measure. The APD and APTD are obtained by calculating the mean over all possible class pairs for each feature combination. The APD and APTD values denote the relative importance of the feature combinations with respect to their information content. Therefore, these will be used as reference values. The following band combinations provide the best discrimination, in order of descending importance: 2-4-5, 3-4-5, 2-4-7 and 3-4-7 (see Table B.2). Note the drop in value for the reduced sample set, which corresponds to vegetative cover classes. Only values for the most relevant band combinations are presented.

For the sake of convenience and because color composites generally consist of three bands, the discussion is presented in terms of selection of sets of three bands. The method, however, is not specific for the number of bands to be selected. In general, the APTD values are high, indicating high probabilities of correct classification as far as the training samples are concerned. This is not surprising, because the sample set used in this study had been tested previously to see whether the classes jointly covered the total spectral range and whether the set of classes incorporated redundant spectral information (see Chapter 3, Part 2).

Table B.2. *The average pairwise divergence and average pairwise transformed divergence for the nine band combinations*

Sub matrix	APD (N=50)	APTD (N=50)	APD (N=44)	APTD (N=44)
1-4-7	186	1922	97	1900
2-4-7	205	1948	117	1933
3-4-7	230	1939	112	1922
1-4-5	234	1936	145	1919
2-4-5	250	1955	161	1943
3-4-5	282	1949	160	1935
2-3-4	206	1922	98	1900
3-5-7	222	1885	117	1853
4-5-7	224	1920	150	1898

Several measures can be derived from the covariance matrix for the information content of band combination. The band combinations are depicted as sub-matrices of the covariance matrix for the total number of bands.

The sum of the eigenvalues of the sub-matrices represents the total variance within the corresponding feature combination. When divided by the total variance of the entire scene, it gives the percentage of the variance explained by the band combination. However, the sum of eigenvalues does not account for the correlation between the bands. The determinant, used by Sheffield (1985) as a measure of information content, does account for the correlation between the spectral bands. Yet it is influenced by the proportion of the variance in the independent dimensions (i.e. proportion of the length of the principal axes defining the sub-space spanned by any particular band combination). It is therefore considered a biased measure.

What is needed is a measure taking account of both the total variance of a combination and the proportionality in the variance explained by the eigenvectors. Measures like λ_{max}-λ_{min}, $\lambda_{max}/\Sigma\lambda$ and $(\lambda_{max}-\lambda_{min})/(\lambda_{max}+\lambda_{min})$, where λ denotes the eigenvalue, do not fulfil this requirement. Therefore, we adopted another measure of information content of band combinations that is based on the variance of difference (VOD) between two populations (i.e.

two bands). A measure for combinations of more than two bands can be obtained by first summing the VOD of the band pairs that are part of the multi-band combination, and then dividing the result by the number band pairs.

$$\frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N}[VAR(i)+VAR(j)-2COV(i,j)}{0.5N(N-1)}$$

whereby i=1...N and j=i+1...N and j>i, N is the number of features

The determinant, the sum of eigenvalues and VOD for the sub- matrices is shown in Table B.3.

Table B.3. *Determinant, eigenvalues and variance of difference of the sub-matrices corresponding to three-band combinations, based on the covariance matrix.*

Sub-matrix	Determinant	Σeigenvalues	Var. of Diff.
1-4-5	1368249 (1)	745.012 (83%)	485.458 (89%)
3-4-5	1108682 (2)	754.959 (84%)	493.671 (91%)
4-5-7	613025 (3)	789.047 (88%)	475.481 (88%)
2-4-5	534346 (4)	727.638 (81%)	476.551 (88%)
1-4-7	312892 (5)	473.579 (53%)	389.869 (72%)
3-4-7	216141 (6)	483.527 (54%)	402.962 (74%)
2-4-7	127451 (7)	456.207 (51%)	375.069 (69%)
3-5-7	37215 (8)	482.249 (54%)	163.048 (30%)
2-3-4	30404 (9)	422.119 (47%)	351.906 (65%)
Total image		894.423	542.692

Results and discussion

The 3-4-5 band combination shows the highest score for the VOD measure, followed by the 1-4-5, 2-4-5 and 4-5-7 combinations. The sum of eigenvalues lists the 4-5-7 combination as most important. The determinant would indicate the 1-4-5 combination as most important. This is in agreement with Sheffield, who mentions that the 1-4-5 band combination usually ranks first using the determinant measure.

The VOD partially corresponds to the order generated by the APD (N=50, see table B.2). The combination 3-4-5 has the highest APD, followed in decreasing order by 2-4-5, 1-4-5 and 3-4-7 combinations, with the combination 4-5-7 ranking fifth. The other measures give lower performance, especially because of the low ranking of the 2-4-5 band combination. All measures rank the 1-4-5 band combination too high.

The reduced set of 44 spectral classes yields the following order for APT values: 2-4-5 and 3-4-5 show the highest scores for APD, followed by 4-5-7 and 1-4-5, and subsequently by combinations with markedly lower APDs. This corresponds to the four-band combinations indicated by the VOD measure, with some difference in ranking order. The VOD measure indicates the feature combinations which provide the highest average divergence between class pairs. The determinant and sum of eigenvalues are considered less accurate estimators because of the clear difference in ranking order.

The correspondence of the VOD measure with the APTD, reflecting probability of correct classification, is less than with the APD. Based on the APTD, the 2-4-5 combination is designated the best feature combination, followed with a slightly lower score by the combinations 3-4-5 and 2-4-7, and, in fourth position, by 3-4-7. The order of the combinations is the same, whether based on the set of 50 spectral classes and based on the reduced set of 44 classes.

Neither of the band combinations 3-4-7 or 2-4-7 are indicated by the VOD measure, nor by the other measures, as belonging to the four most relevant feature combinations. The VOD measure does not serve as a reliable estimator of the information content of band combinations as regards classification performance.

This is explained by the fact that the spectral variance, and as such the variance of difference, does not provide a reliable estimate of the single band discriminative power. This was discussed in the second section of this appendix. The contribution of TM band 7 is underestimated.

We investigated the use of the coefficient of variation as an alternative to the variance as a measure for the information content of single spectral bands. That required recalculation (scaling) of the covariance matrix and the determination determinant, eigenvalues, and VOD for the individual band combinations on the basis of the adjusted covariance matrix.

Using the scaled VOD values, higher priority is given to the combinations 2-4-7 and 3-4-7. In fact, this would give priority to these two combinations over the combinations 3-4-5 and 2-4-5. This priority reflects the overestimation of the informational value of TM band 7 and an underestimation of TM band 5, as consequence of the application of the coefficient of variation as single band information content. The highest ranking combination based on the scaled VOD measure proves to be 1-4-7, while ranking only sixth on the basis of the APTD score. Therefore, we may conclude that measures based on the coefficient of variation do not perform better in indicating band combinations with a high probability for correct classification.

4. CONCLUSIONS

1. With respect to the single bands, the variance serves as a measure for the average distance between land cover spectral classes. This does not apply to single band discriminative power.

   Use of the variance as a measure of information content over-estimates the importance of TM band 4. The contribution of TM band 5 and TM band 7 is underestimated, while in fact TM band 5 is most powerful in discriminating between the cover types. TM band 4 and TM band 7 are equally important. The importance we ascribe to TM band

5 is in agreement with the conclusions of Sheffield (1985).
The coefficient of variance does not provide an alternative to the variance measure, due to the overestimation of information content of TM band 7.

2. More or less the same applies with respect to the band combinations. The variance of difference measure serves to indicate band combinations showing high scores for the average pairwise divergence. The VOD performs better than the determinant or the eigenvalues in this case.
The VOD, determinant, and eigenvalues are less suitable as indicators for classification performance of band combinations (given by the average pairwise transformed divergence). They failed to indicate the combinations including TM band 7, while of the four spectral band combination with highest APTD values two inlcuded TM band 7. The four combinations with the highest APTD score were 2-4-5, 3-4-5, 2-4-7 and 3-4-7. The limited suitability of the VOD and other measures is explained by the low accuracy of the variance as estimator of single band information value. Two of the four relevant spectral band are indicated by the VOD measure.
The VOD measure, based on the scaled covariance matrix (i.e. adjusted for the average value level), overestimates the information value of band combinations which include TM band 7. On the other hand, band combinations including TM band 5 are under-estimated. Also the adjusted VOD did not correctly indicate the most relevant band combinations.

3. The fact that the same ranking applies to band combinations for the complete set and the reduced set of 44 cover classes, on the basis of their APTD value, suggests that the same band combinations serve to discriminate between general cover classes (like vegetated area, built-up area and bare soil, and waterbodies) as between vegetated cover classes.

It is concluded that exact measures of information content for specific applications cannot be inferred from general image variance characteristics.

APPENDIX C.

FARMING SYSTEM CLASSIFICATION.

For the classification of the farms two labels are attributed; One for the composition of the farm with respect to crops and land uses (the farm land use types), one for the scale of production and management level.

The following farm land use types are defined:

1. Mixed farms. As well crops as grassland is encountered on the farms.
   1A. Besides to animal husbandry, are the farms dedicated to the production of crops like maize, rice, yucca or other root and tuber crops.
   1B. The farm activity is directed to the cultivation of tree crops (Cacao, Citrus, Guanabana) and/or other perennial crops like plantain, papaya and to animal husbandry.
   1C. As well annual crops as perennial crops are cultivated and livestock is kept on the farm.
2. Farms dedicated to livestock production. More than 90 percent of the farm area is used for grassland.
   2A. Farms dedicated to animal husbandry, with less than 10 percent of the farm area is dedicated to the production of predominantly annual crops with some times a limited area with perennial crops.
   2AP. Cattle farms with less than 10 percent of the farm area used for perennial crops, predominantly tree crops.
   2B. Farms dedicated to animal husbandry only.
3. Farms dedicated to crop production. Mostly annual crops like maize, rice or root and tuber crops like yucca and chamol. Often perennial crops are found, sometimes for he larger part of the farm area (plantain, cacao, etc.).
4. Plantations and orchards.
   4A. Macadamia
   4B. Citrus and/or other fruit trees (Guanabana), sometimes parts of the farm are used for pastures.
   4C. Coffee or Cacao, in cases both. Sometimes annual crops are grown in smaller amounts.
   4D. Ornamental plants.
   4E. Combination of the different (semi-)perennial plantation crops like plantain, banana and yucca. Also sometimes annual crops are grown mostly for home consumption.
   4F. Banana.
   4G. Reforestation (plantings of f.e. Laurel and Eucalyptus) or forest reserves.
   4H. Palm heart plantations.
5. Non farm area.
   5A. Homestead gardens
   5B. Non cultivated pastures, neglected areas.
6. Farms with a considerable part under forest cover and with pastures. Income is obtained for a -sometimes important- part derived from the exploitation of the forest.
7. Farms with considerable parts under forest cover and a parts of the farm area dedicated to a variety of crops and pasture. Crop production is meant for home consumption.

8. With part of the area under forest cover and a part dedicated to plantation crops like macadamia, cardamon, vanilla and other.
9. Areas with parts dedicated to reforestation and parts for livestock production.

Codes referring to scale of production and management levels:

1. Commercial enterprises with a high level of capital investment (machinery and buildings) and high level of inputs (materials and labour). Scale of production is large. Production of commercial crops generally export oriented.
   (Banana plantation, larger ornamental crop plantation, larger macadamia plantations, etc.)
2. Intermediate level
   - Large scale, extensive stock breeding. Larger farms dedicated to livestock production. Generally a few persons are employed for the current activities. Level of mechanization is very low; the use of horses is common. A coral is present and often improved pastures are found.
   - Production on intermediate level with intensive use of the land. Intermediate level of capital investment and use of hired labour. Machinery is used for field preparation, irrigation and/or other farm activities. The production goods are generally owned. This type of management is applicable to enterprises, intermediate in size, which produce ornamental crops, the smaller macadamia plantations and also the farms with larger areas for the commercial production of maize other arable crops are attributed this management class.
3. Farms which largely depend on the family for the labour inputs, sometimes temporal workers are hired. A tractor is in cases hired for the ploughing of maize plots. Generally only traditional tools are used. The scale of production and level of capital investment is low.
   3B. Only part of the income is generated at the farm. The owner is employed off farm.
   3D. Owner does not live at the farm. The management of the farm is in hands of a hired person.
4. Areas without agricultural use or on marginal levels only.
   4A. With agricultural production on very limited scale and predominantly intended for home consumption; not to be considered as an economic activity. The areas concern homestead gardens or other plots of very limited dimensions.
   4B. Abandoned or severely neglected area. Generally pastures.

APPENDIX D.

LAND COVER CLASSIFICATION OF THE GRS AREA

For the land cover classification of the Guacimo-Rio Jiménez-Siquirres area see the attached poster.

KEY TO LAND COVER CLASSES.

1. FOREST AREA
2. SWAMP AND PEAT VEGETATION. (Consisting mainly of Yolillo (*Raphia Taedigera*))
3. SECONDARY FOREST
4. SECONDARY VEGETATION. (Comprising dense woody regrowth as well as scrubby vegetation. Young tree planting and e.g. cassave might also be classified as secondary vegetation.)
5. PASTURE.
6. (SEMI-) NATURAL GRASSLANDS AND DEGRADED GRASSLANDS. (Also comprising Guinea pastures (*Panicum maximum*), not to be considered a degraded grassland).
7. WOODED AREA. (Comprising cacao lots, orchards, macadamia plantations, wooded pastures, reforested area, wooded river banks and other densily wooded areas).
8. PLANTAIN
9. BANANA
10. BAMBOO
11. ORNAMENTAL CROPS. (Especially referring to Yuca (*Dracaene massangeana*)).
12. PEJIVALLE (*Bactris gasipaës*), PALMITO (*Bactris gasipaës*, harvested as young plants) or COCONUT PLANTATION.
13. BARE SOIL AND BUILT-UP AREA.
14. RIVERBEDDING, SAND & BOULDERS.
15. RIVER.
16. LAGOONS, SHADOW OF CLOUDS.
17. CLOUDS.

APPENDIX E

LAND USE PATTERNS IN THE GRS AREA

KEY TO LAND USE PATTERNS

1. FOREST LAND

01/F1 Lowland and promontory humid tropical forest.
02/F2 Forested areas within cultivated lands.
03/f3 Disturbed forest. Small forested parts within cultivated lands

2. AREA OF AGRICULTURAL PENETRATION

11/AP1 Forest covers > 30 % of the area. Land use: timber wood extraction and cattle farming.
12/AP2 Forest cover > 50 % of the area. Use: Timber wood extraction and cattle farming.

3. AGRICULTURAL LANDS, FARM LAND

SMALL FARMS

Area with mixed land cover
31/M1 Area with residential and agricultural function. Very small properties, homestead gardens, annual cropping and grazing. Production for home consumption.
32/M2 Very small mostly mixed farms. Most important: annual crop production, perennial crops and livestock less important.
33/M3 Very small to small farms. Dominant use: annual cropping, perennial cropping and grazing are less important
34/M4 Very small to small farms, dedicated to annual crop (maize,cassava) livestock production. Fruit trees are generally found within the pastures.
35/M5 Very small to small farms. Use: annual and perennial crops.

Mixed land cover and forest
36/F1 Are partly covered with forest. Very small to small farms. Annual, perennial and tree crops.
37/F1 Small farms. Land is used for grazing sometimes in combination with perennial crops

Area with predominant grassland cover
38/P1 Very small farms. Cattle farming and perennial crop production (e.g. palmito)
39/P2 Very small to small farms. Land used for grazing and tree crops (e.g. fruit, macadamia)

Area with Wooded Area as dominant cover
40/W1 Very small to small farms. Cacao plantations and pastures (mostly with trees).

Area with Bare Soil as dominant cover
45/B1 Very small farms. Land use: annual cropping (e.g. maize, cassava)
46/M1 Land cover of bare soil and forest. Recently cleared area

MEDIUM SIZED FARMS

Are with mixed land cover
51/M1 Land use for Grazing or grazing in combination with crop production
52/M2 Farms dedicated to annual cropping or combination of annual and perennial cropping and grazing.
53/M3 Small to medium sized farms, land use directed to perennial and annual crop production.

Area with grassland cover
56/P1 Cattle farming
57/P2 Parts of the area is classified as degraded grassland. Land use for grazing, to lesser extend for annual and perennial cropping.

LARGE AND VERY LARGE FARMS

Area with grassland cover
61/P1 Livestock production. Often improved grass species is used. No trees scattered over the fields.
62/P2 Parts classified as degraded pastures. Cattle farming.
63/P3 Land cover for a large part composed of degraded grasslands. Cattle farming
64/P4 Land cover consists of Wooded Area, cultivated pasture, degraded grasslands and forest or secondary vegetation. Large farms dedicated to livestock breeding and reforestation.

Area with Bare Soil as dominant cover

66/B1 Land cover classified as bare soil, wooded area and secondary vegetation. Large commercial farms dedicated to (arable) crop production for export.

67/B2 Part of the area is classified as ornamental crop. Commercial enterprise dedicated to production of ornamental crops.

68/B3 Land cover consists of bare soil and pasture. Land use in process of change.

69/B4 Bare soil as result of clearing.

Area with forest cover

70/F1 Part of large farm with forest cover.

4. PLANTATIONS

Banana plantations

81/B1 Banana plantation

82/B2 40 to 60 % of the area is classified as banana. Parts of the plantation area not covered with banana (low plant densities, open spots due to unsuitable soils and roads traversing the plantation.

Dominant land cover : Wooded Area

85/W1 Macadamia plantation

Land cover partly consisting of bamboo

87/BM1 Land cover consists of bamboo, forest, secondary vegetation. Area used for production of bamboo.

Area with mixed land cover

86/M1 Land cover: Wooded Area, Secondary vegetation, grassland and/or forest. Different kinds of crops: E.g. coffee, cardamon. In cases parts forested.

5. TOWNS AND VILLAGES.

91/B1 Area classified as non vegetated: Residential area.

6. WETLAND

04/F4 Dominant land cover: Forest. Parts classified as swamp vegetation (*Yolillo*).

18/W4 Land cover consists of Wooded Area, non - cultivated grassland and bare soil. Areas with semi-natural grass and woody vegetation, for part of the year inundated.

7. OTHER, WASTE LANDS

With forest as dominant cover.

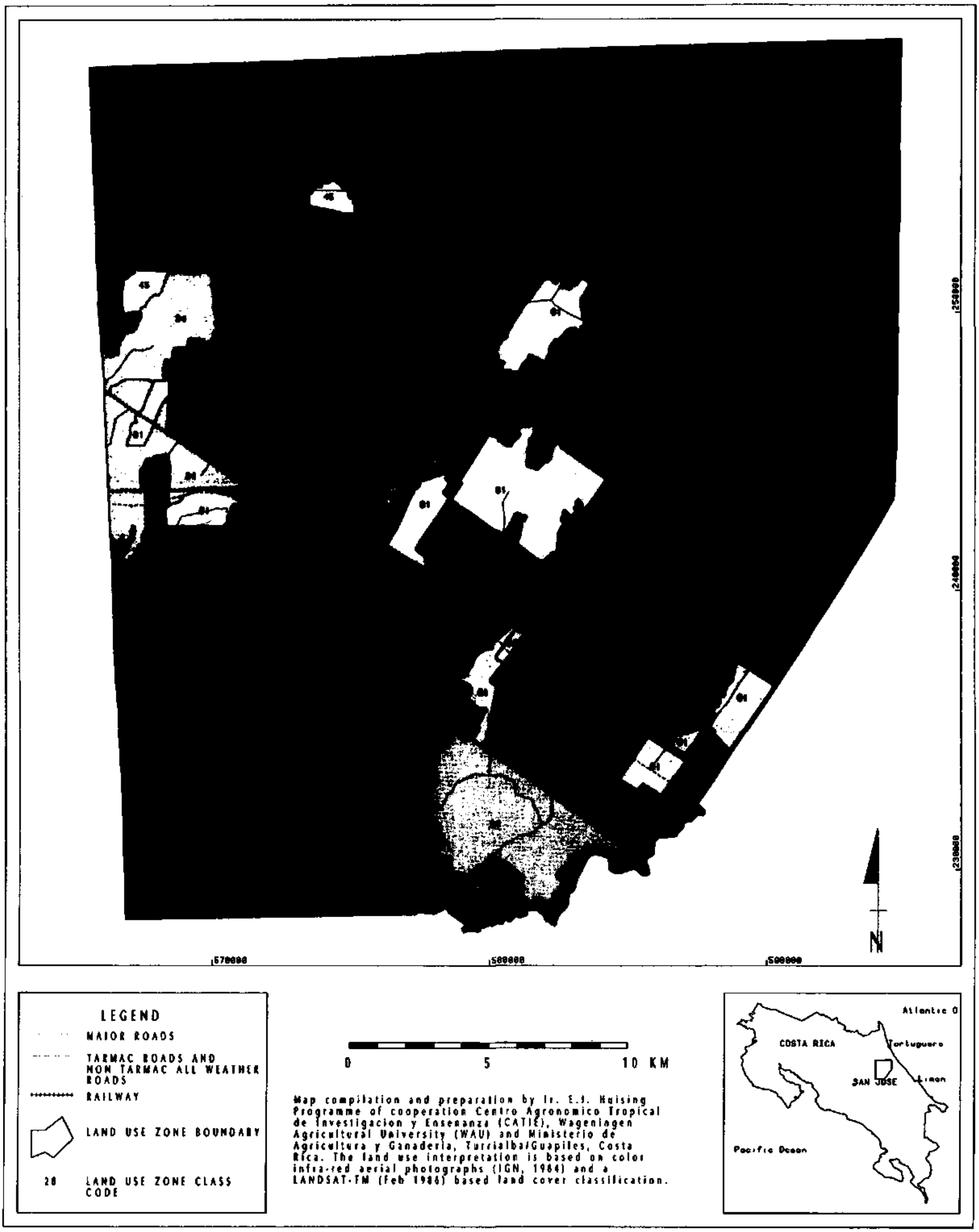
13/F3 Land cover consists of Forest and Wooded Area. Reforested area, parts used for extensive grazing.

Area with mixed land cover.

16/P1 Land cover: Wooded Area, degraded grasslands and pastures. Medium to large farms, land use for very extensive grazing.

17/W2 Land cover: Wooded Area, Forest and (to lesser extend) bare soil and secondary vegetation. Large farms. Land not used or very extensively.

19/W5 Land cover: Wooded Area an Forest and (sub dominant) cultivated and non-cultivated grasslands. Small farms, land partly used, partly waste land.


N
LEGEND
MAIOR ROADS
TARMAC ROADS AND NON TARMAC ALL WEATHER ROADS
RAILWAY
LAND USE ZONE BOUNDARY
28 LAND USE ZONE CLASS CODE
0 5 10 KM
Map compilation and preparation by Ir. E.J. Huising Programme of cooperation Centro Agronomico Tropical de Investigacion y Ensenanza (CATIE), Wageningen Agricultural University (WAU) and Ministerio de Agricultura y Ganaderia, Turrialba/Guapiles, Costa Rica. The land use interpretation is based on color infra-red aerial photographs (IGN, 1984) and a LANDSAT-TM (Feb 1986) based land cover classification.
Atlantic O
COSTA RICA
Tortuguero
SAN JOSE
Limon
Pacific Ocean

REFERENCES

Ahmad, W., 1986. Land use/cover mapping using remotely sensed data with special reference on applications to forestry. A review of literature: 1973-1984. Divisional report 86/1, CSIRO, Institute of biological resources, Division of water and land resources, Canberra A.C.T.

Alberts, J. and W. Kreiling, 1989. Photogrammetric guide. Herbert Wichmann Verlag GmbH, Karlsruhe.

American Society of Photogrammetry, 1960. Manual of photographic interpretation, A.S.P., Washington, D.C.

Anderson, J.R. et al., 1976. A land use and land cover classification system for use with remote sensor data, USGS Professional paper 964, U.S. Gov. Printing Office, Washington, D.C.

Argialas, D.P. and C.A. Harlow, 1990. Computational image interpretation models: An overview and perspective. *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, pp. 871-886.

Aronoff, S., 1984. An Approach to Optimized Labelling of Image Classes. *Photogrammetric Engineering and Remote Sensing*, Vol. 50, No. 6, pp. 719-727.

Bomberger, E.H. and H.W. Dill Jr., 1960. Photo interpretation in agriculture. In: Colwell, R.N. (Ed), Manual of photographic interpretation. American Society of Photogrammetry (ASP), Washington, USA.

Bouma, J., H.A.J. van Lanen, A. Breeuwsma, H.J.M. Wösten and M.J. Kooistra, 1986. Soil survey data needs when studying modern land use problems. *Soil Use and Management*, Vol. 2, No. 4, 125-130.

Bouma, J., 1989. Using soil survey data for quantitative land evaluation. Advances in Soil Science, Vol. 9, Springer-Verlag New York Inc., pp. 177-213.

Bregt, A.K., J. Bouma and M. Jellinek, 1987. Comparison of thematic maps derived from a soil map and from kriging of point data. *Geoderma* 39: pp. 281-291.

Buchheim, M.P. and T.M. Lillesand, 1989. Semi-Automated Training Field Extraction and Analysis for Efficient Digital Image Classification. Photogrammetric Engineering and Remote Sensing, Vol. 55, No. 9, pp. 1347-1355.

Buiten, H.J., 1988. Matching and mapping of remote sensing images. Paper presented at: Symposium of the Int. Society for Photogrammetry and Remote Sensing, 1988, Kyoto, Japan.

Bunschoten, L., 1989. Remote sensing. Mogelijkheden en toepassingen binnen het CBS. Mndstat landb, 1989/12, Centraal Bureau for Statistiek, 's-Gravenhage, The Netherlands.

Burley, T.M. 1961. Land use or land utilization? *Professional Geographer* 13 (6):18-20.

Burrough, P.A., 1986. Principles of geographical information systems for land resources assessment. Monographs on soil and resources survey no. 12, Oxford science publications, Clarendon press, Oxford, UK.

Butera, M.K., 1986. A correlation and regression analysis of percent canopy closure versus TMS spectral response for selected forest sites in the San Juan National Forest, Colorado. *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-24, no.1, pp 122-129.

Campbell, J.B., 1981. Spatial Correlation Effects upon Accuracy of Supervised Classification of Land Cover. *Photogrammetric engineering and Remote Sensing*, Vol. 47, No. 3, pp. 355-363.

CATIE, 1986. Algunas consideraciones sobre la producción de ganado de doble proposito en el istmo centroamericano. Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica.

Chrisman, N.R., 1989. The error component in spatial data. In: M. Goodchild and S. Gopal (Eds), Accuracy of spatial data bases. Taylor & Francis, London.

Christensen, E.J., J.R. Jensen, E.W. Ramsey and H.E. Mackey, 1988. Aircraft MSS data registration and vegetation classification for wetland change detection. *Int. J. Remote Sensing*, Vol. 9, no .1, pp. 33-38.

Chuvieco E. and Congalton R.G., 1988. Using cluster analysis to improve the selection of training statistics in classifying remotely sensed data. *Phot. Engineering and Remote Sensing*, Vol 54, No. 9, p. 1275-1281.

Cihlar J., M.C. Dobson, T. Schmugge, P. Hoogeboom, A.R.P. Janse, F. Baret, J. Guyot, T. le Toan and P. Pampaloni, 1987. Procedures for the description of agricultural crops and soils in optical and microwave remote sensing studies. Review article. *Int. Journal of Remote Sensing*, Vol. 8, No. 3, p 427-439.

Clawson, M and C.L. Stewart, 1965. Land use information. A critical survey of US statistics including possibilities for greater uniformity. The Johms Hopkins press for resources for the future, Inc., Baltimore.

Clevers, J.G.P.W., 1986. The derivation of a simplified reflectance model for the estimation of LAI. In: Remote Sensing for Resources Development and Environmental Management. Proceedings off the 7th international symposium on. ISPRS Commission VII, Enschede, 1986.

Clevers, J.G.P.W., 1986. The application of a vegetation index in correcting the infrared reflectance for soil background. In: Remote Sensing for Resources Development and Environmental Management. Proceedings of the 7th international symposium on. ISPRS Commission VII, Enschede, 1986.

Clevers, J.G.P.W., 1992. Influence of leaf properties on the relationship between WDVI and LAI: a sensitivity analysis with the Sail and Prospect model. Report LUW-LMK-199202, Department of Landsurveying and Remote Sensing, Wageningen Agricultural University.

Corine, 1989. Corine programme land cover nomenclature, Directorate-General, Environment, Consumer protection and nuclear safety, European Commission, Brussels

Crist, E.P. and R.C. Cicone, 1984. A physically-Based Transformation of Thematic Mapper Data - The TM Tasselled Cap. *IEEE Transactions on Geoscience and Remote Sensing*. Vol. GE-22, No. 3, pp. 256-263.

Crist, E.P. and R.C. Cicone, 1984. Application of the Tasselled Cap Concept to Simulated Thematic Mapper Data. *Phot. Engineering and Remote Sensing*, Vol. 50, No. 3, pp. 343-352.

Cross, A. and D.C. Mason, 1985. Segmentation of remotely-sensed images by a split and merge process. Paper presented at the international conference on Advanced technology for monitoring and processing global environmental data, University of London, 10 -12 Sept 1985. Remote Sensing Society and CERMA, London, UK.

Curran, P.J., 1983. Multispectral remote sensing for the estimation of green leaf area index. *Phil. Trans. R. Soc. Lond.* A 309. pp 257-270.

Curran, P.J. and H.D. Williamson, 1987. GLAI estimation using measurements of red, near infrared, and middle infrared radiance. *Phot. Eng. and Remote Sensing*, Vol. 53, No. 2, pp 181-186, 1987.

Desachy, J., P. Debord and S. Castan, 1988. An expert system for satellite image interpretation and GIS based problem solving. Proceedings ISPRS congress, Kyoto, Japan, vol. 27, p. 518-528.

Duda, R.O. and P.E. Hart, 1973. Pattern classification and scene analysis. John Wiley $ Sons, Inc., New York, 482 p.

Ellis, F., 1983. Las transnacionales del banano en centroamerica. Editorial Universitaria Centroamericana (EDUCA), San José, Costa Rica.

Erdas Inc. 1987. User's Guide. Erdas Inc., Atlanta.

Essink, L. and H. Romkema, 1989. Ontwerpen van informatiesystemen. Academic Service, Amsterdam.

FAO, 1976. A framework for land evaluation. FAO Soils bulletin 32. Food and Agriculture Organization of the United Nations, Rome.

FAO, 1984. Guidelines: land evaluation for rainfed agriculture. FAO Soils bulletin 52. Food and Agriculture Organization of the United Nations, Rome.

Foody, G.M., 1988. Incorporating remotely sensed data into a GIS: The problem of classification evaluation. *Geocarto international* (3) 1988.

Fox, J. and P. Suharsono, 1986. Land units, land dissection and land cover in east Java. *ITC-Journal* 1986-2. International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede.

Franklin, J., 1986. Thematic mapper analysis of coniferous forest structure and composition. *International Journal of Remote Sensing*, vol.7, no. 10, pp 1287-1301.

Fresco, L., H. Huizing, H. van Keulen, H. Luning and R. Schipper, 1990. Land evaluation and farming systems analyses for land use planning. FAO Guidelines: Working document. FAO, Rome.

Fresco, L.O. and E. Westphal, 1988. A hierarchical classification of farm systems. *Experimental Agriculture*, Vol. 24, pp. 399-419.

Fung T. and E. LeDrew, 1987. Application of Principle Components analyses to change detection. *Phot. Eng. and Remote Sensing*, Vol.53, No.12, December 1987, pp 1649-1658.

Gardner, B.R., B.L. Blad, D.R. Thompson and K.E. Henderson, 1985. Evaluation and interpretation of of Thematic Mapper Ratios in Equations for estimating Corn growth Parameters. Remote Sensing Environ., 18: 225-234.

Gils, H. van, S. Groten, H. Huizing, W. van Wijngaarden and D. van der Zee, 1987. Land ecology- and land use survey. ITC textbook, Lecture Series N9. ITC. Enschede, The Netherlands.

Goel, N.S and T. Grier, 1988. Estimation of Canopy Parameters for Inhomogeneous Vegetation Canopies from Reflectance Data: III. TRIM: A model for Radiative Transfer in Heterogeneous Three-Dimensional Canopies. *Remote Sensing of Environment* 25: 255-293.

Goodenough, D.G., M. Goldberg, Plunkett, G. and J. Zelek, 1987. An expert system for remote sensing. *IEEE transactions on geoscience and remote sensing*, Vol. GE-25, No. 3, may 1987.

Gong P. and P. Howarth, 1990. The use of structural information for improving land - cover classification accuracies at the rural-urban fringe. *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 1, p. 67-73.

Gurney, C.M., 1983. The use of contextual information in the classification of remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, Vol. 49, No. 1, p. 55-64.

Haack B.N., 1984. Multisensor Data Analysis of Urban Environments. *Photogrammetric Engineering and Remote Sensing*, Vol.50, No.10, 1984, pp 1471-1477.

Haas, R.H., 1985. Using satellite data for mapping and monitoring vegetation resources. In: Remote sensing applications for consumptive use (evapotranspiration). A.I. Johnson & A. Rango (Eds), AWRA monograph series no. 6. American Water Resources Association, Maryland, 1985.

Hadipriono, F.C., J.G. Lyon and W.H. Thomas Li, 1991. Expert opinion in satellite data interpretation. *Photogrammetric Engineering and Remote Sensing*, Vol. 57, No. 1, pp. 75-78.

Hall, C., 1984. Costa Rica, una interpretación geográfica con perpectiva histórica. Editorial Costa Rica (ECR), San José Costa Rica.

Hall-Könyves, K., 1990. Crop monitoring in Sweden. *Int.J.Remote Sensing*, Vol. 11, No. 3, p. 461-484.

Heuvelink, G.B.M, P.A. Burrough and A. Stein, 1989. Propagation of errors in spatial data modelling with GIS. *Int.J. Geographical Information Systems*, Vol. 3, no. 4, pp 303-322.

Hopkins, P.F., A.L. Maclean and T.M. Lillesand, 1988. Assessment of Thematic Mapper Imagery for Forestry Applications under Lake State Conditions. *Photogrammetric Engineering and Remote Sensing*, vol.54, no.1, pp.61-68.

Horler, D.N.H. and F.J. Ahern, 1986. Forestry Information Content of Thematic Mapper Data. *International Journal of Remote Sensing*, vol.7, no.3, pp.405-428.

Huising, E.J. and M.A. Mulders, 1991. The use of TM data for identification of land cover and for detection of land use changes in the Atlantic Zone of Costa Rica. BCRS report no. 91-36. Netherlands Remote Sensing Board, The Netherlands.

Husson, A., 1989. Remote sensing and agriculture, agricultural statistics project for the european community. EARSeL News, Jan.- June 1989, no. 38.

Hutchinson, C.F., 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. *Phot. Eng. and Remote Sensing*, Vol. 48, No. 1, Jan 1982, p.123-130.

ICOMAND (International Committee on the classification of Andisols). Circular lettre No. 10, 29 February, 1988.

Ingebritsen S.E. and R.J.P.Lyon, 1985. Principle component analyses of multitemporal image pairs. *Int.J.Remote Sensing*, Vol. 6, No. 5, 687-696.

Ioka, M. and M. Koda, 1986. Performance of Landsat-5 TM data in land-cover classification. *Int.J.Remote Sensing*, 1986, vol. 7, no. 12, p. 1715-1728.

Janssen, L.L.F., 1990. GIS supported land cover classification of satellite images. Paper presented at the First European Conference on Geographical Information Systems, Amsterdam, the Netherlands, April 10 - 13, 1990.

Janssen, L.L.F and J.D. van Amsterdam, 1991. An object based approach to the classification of remotely sensed images. Proceedings of the 1991 Intenational Geoscience and Remote Sensing Symposium : IGARSS'91, Remote Sensing: global monitoring for earth management, June 3-6, Vol. IV, pp. 2191-2195.

Janssen, L.L.F., M.N. Jaarsma and E.T.M. van der Linden, 1990. Integrating topographic data with remote sensing for land cover classification. *PE&RS*, Vol. 56, No. 11, p.1503-1506.

Jaramillo, R., 1979. Las principales caracteristicas morfológicas del fruto de banano; variedad Cavendish gigante (musa AAA) en Costa Rica. UPEB, Panama.

Jensen R.J., 1983. Urban/suburban land use analyses. In : Manual of Remote Sensing, Vol II, p. 1571-1666.

Jensen, R.J., 1986. Introductory Digital Image Processing - A Remote Sensing Perspective. Prentice-Hall, Englewood Cliffs, New Jersey.

Jupp, D.L.B., A.H. Strahler and C.E. Woodcock, 1988. Autocorrelation and Regularization in Digital Images. I. Basic Theory. *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 26, No. 4, pp. 463-473.

Jupp, D.L.B., J. Walker and L.K Penridge, 1986. Interpretation of Vegetation Structure in Landsat-MSS Imagery: a Case Study in Distributed Semi-arid Eucalypt Woodlands. Part 2. Model-based Analysis. *Journal of Environmental Management* 23, 35-57.

Kauth, R.J. and G.S. Thomas, 1976. The Tasselled Cap- a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. Proceedings of the symposium on machine processing of remotely sensed data, Purdue University, West Lafayette, IN, pp. 4B41-4B51.

Kenk, E., M. Sondheim and B. Yee, 1988. Methods for improving accuracy of thematic mapper ground cover classifications. *Canadian Journal of Remote Sensing*, vol. 14, no. 1, may 1988.

Klir, G.J. and T.A. Folger, 1988. Fuzzy sets, uncertainty and information. Prentice-Hall International, U.S.A.

Labovitz, M.L., 1986. Issues Arising from Sampling Designs and Band Selection in Discriminating Ground Reference Attributes Using Remotely Sensed Data. *PE&RS*, Vol. 52, No. 2, pp. 201-211.

Lee, T. and J.A. Richards, 1985. A low cost classifier for multitemporal applications. *Int.J. Remote Sensing*, 1985, Vol. 6, No. 8, 1405-1417.

Lee, T.L., A. Richards and P.H. Swain, 1987. Probabilistic and evidential approaches for multi-source data analyses, *IEEE transactions on geoscience and remote sensing*, Vol. GE-25, No. 3, pp. 283-293.

Leprieur C.E., J.M. Durand and J.L. Peyron, 1988. Influence of topography on forest reflectance using landsat thematic mapper and digital terrain data. *Photogrammetric engineering and remote sensing*, vol. 54, no. 4, pp. 491-496.

Li, X and A.H. Strahler, 1985. Geometric Modelling of a Conifer Forest Canopy. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-23, No. 5, pp. 705-721.

Lillesand, T.M. and R.W. Kiefer, 1987. Remote Sensing and Image Interpretation (2nd edition). John Wiley & Sons, New York, 721 p.

Lindgren, D.T., 1985. Land use planning and remote sensing. Martinus Nijhoff publishers, Boston, USA.

Lintz Jr., J. and D.S. Simonett (Eds), 1976. *Remote sensing of environment*. Wesley publishing company, Reading, Masachusets.

Lipschutz, S., 1987. Theory and problems of linear algebra. SI (metric) edition. Schaum's outline series. McGraw-Hill Book company.

Lucas, P.J.F. and L.C. van der Gaag, 1988. Principes van expertsystemen. Academic Service. Schoonhoven, The Netherlands.

Manore, M.J. and R.J. Brown, 1990. Remote sensing/GIS integration in the Canadian crop information System. *Geocarto International*, (1).

Markham, B.L. and J.L.Barker, 1987. Thematic Mapper bandpass solar exoatmospheric irradiances. *Int J. Remote Sensing*, 8: 517-523.

Markham, B.L. and J.L.Barker. Landsat MSS and TM post-calibration dynamic ranges, exo-atmospheric reflectances and at-satellite temperatures.

Marsman, B.A. and J.J. de Gruijter, 1986. Quality of soil maps: A comparison of survey methods in a sandy area. Soil survey papers. No. 15, Soil Survey Institute, Wageningen.

Meijerink, A.M.J., 1985. Geo information systems for land use zoning and watershed management. *ITC Journal* 1985-4. International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede.

Meijerink, A.M., Valenzuela, C.R. and A. Stewart, 1988. The integrated land and watershed management information system. ITC publication number 7, Enschede, The Netherlands.

Middelkoop, H., J. Miltenburg and N. Mulder, 1989. Knowledge engineering for image interpretation and classification: a trial run. *ITC Journal*, 1989-1, pp. 27-32.

Molenaar M., 1989a. Single Valued Vector Maps - A Concept in Geographic Information Systems. *GIS*, Vol. 2, No. 1, p.18-26.

Molenaar M., 1989b. Status and problems of Geographical Information Systems, A conceptual approach. Proceedings of the 42nd photogrammetry week at Stuttgart University, September 11-16, Stuttgart, pp. 7-23.

Molenaar M., 1991a. Terrain objects, data structures and query spaces. In: M. Schilcher (Ed), Geo-Informatik: Beitrage zum internationalen anwenderforum 1991 Geo Informationssysteme und umwelt informatik, Duisburg 20-21 Feb. 1991. Siemens Nixdorf Informationssysteme AG, Berlin, FRD, 13 p.

Molenaar, M., 1991b. Formal Data Structures, Object Dynamics and Consistency Rules. In: H. Ebner, D. Fritsch and C. Heipke (Eds.). Digital Photogrammetric Systems. Wichmann Verlag, Karlsruhe.

Molenaar, M., 1991c. Object-hierarchies, why is data standardization so difficult. In: Kockelhoven, *et al.* (Eds.), Kadaster in perspectief, Dienst van het Kadaster en de Openbare Registers, Apeldoorn.

Molenaar, M., 1991d. Object hierarchies and uncertainty in GIS or why is standardisation so difficult? Centre for Geo-information processing; Dept. for land surveying and remote sensing, Wageningen.

Molenaar, M. and L.L.F. Janssen, 1991. Integrated processing of remotely sensed and geographic data for land inventory purposes. In prep.

Molenaar, M. and J. Stuiver, 1987. A PC digital monoplotting system for map updating. *ITC Journal*, 1987-4. Enschede, The Netherlands.

Mulder, N.J., 1988. Digital image processing, computer-aided classification and mapping. In: A.W. Küchler and I.S. Zonneveld (eds.), Vegetation mapping. Kluwer academic publishers. Dordrecht. The Netherlands.

Mulders, M.A., S. de Bruin and B.P. Schuiling, 1992. Structured approach to land-cover mapping of the Atlantic Zone of Costa Rica using single date TM data. *Int. J. Remote Sensing*, Vol. 13, No. 16, pp 3017-3033.

Myeni, R.B., G. Asrar, R.B. Burnett and E.T. Kanemasu, 1987. Radiative transfer in an anisotropically scattering vegetative medium. *Agriculture and Forest Meteorology*, 41, 97-121.

Myers, V.I, 1970. Soil, water and plant relations. In: Remote Sensing with Special Reference to Agriculture and Forestry. Nationa; acedemy of Sciences, Washington, D.C., pp. 253-323.

Myers, V.I. (Ed.), 1983. Remote sensing applications in agriculture. In: J.E. Estes and G.A. Thorley (Eds), Manual of remote sensing. A.S.P. The Sheridon Press. U.S.A.

Nunnally, N.R., 1974. Interpreting land use from remote sensor imagery. In: J.E. Estes and L.W. Senger (EDS), Remote Sensing. Techniques for environmental analyses.

Oñoro, M.T. de (Ed.), 1990. El asentamiento Nequev. Interacción de campesinos y estado en el aprovechamiento de los recursos naturales. Serie técnica. Informe técnico no. 162. Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica.

Oosterom, A.P., H.J. Stuiver, W.K. Krabbe and R.M. Hootsmans, 1992. Geographical information techniques and photogrammetry in soil and landscape inventories of the Atlantic Zone in Costa Rica. p. 31-36. In: Wielemaker, W.G.and A.P Oosterom.

Openshaw, S., 1989. Learning to live with errors in spatial data bases. In: M. Goodchild and S. Gopal (Eds), Accuracy of spatial data bases. Taylor & Francis, London.

Peplies, R.W., 1976. Cultural and landscape interpretation. In: J. Lintz, Jr. and D.S. Simonett (Eds), Remote sensing of environment, Addison-Wesley publishing company, Inc., Reading, Massachusetts.

Peterson, D.L.; M.A. Spanner; S.W. Running and T.B. Teuber, 1987. Relationship of thematic mapper simulator data to leaf area index of temperate coniferous forest. *Remote Sensing of Environment*, vol. 22, no. 3, pp. 323-341.

Peuquet, D.J., 1984. A conceptual framework and comparison of spatial data models. *Cartographica*. Vol. 21, No. 4, pp. 66-113.

Plummer, S.E., 1988. Exploring the relationship between leaf nitrogen content, biomass and near-infrared/red reflectance ratio. *Int. J. Remote Sensing*, 1988, vol. 9, no. 1, pp 177-183.

Pouget M. and M.A. Mulders, 1988. Description of the land surface for correlation with remote sensing data. Paper presented on the 5th Int. Symposium of Working Group Remote Sensing of ISSS, Budapest, 1988.

Rhind, D.W. and R. Hudson, 1980. Land use. University paperbacks. Methuen & Co., New York, USA.

Richards, J.A. 1986. Remote sensing digital image analysis, an introduction. Springer Verlag, Berlin-Heidelberg-New York.

Richardson, A.J., J.H. Everitt and H.W. Gausman, 1983. Radiometric estimation of biomass and nitrogen content of Alicia grass. Remote Sensing Environ., 13: 179-184.

Robinson, V.B. and A. U Frank, 1987. Expert systems for geographic information systems. *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 10, pp. 1435-1441.

Rossiter, D.G., 1990. ALES: a framework for land evaluation using a microcomputer. *Soil Use and Management*, Vol. 6, No. 1, pp. 7-20.

Sadowski, F.G., J.A. Sturdevant, W.H. Andetson, P.H. Seevers, J.W. Feuquay, L.K. Balick, F.A. Wlatz and D.T. Lauer, 1985. Early results of Investigations of Landsat 4 Thematic Mapper and Multispectral Scanner Applications. In: Proceedings of Landsat-4 Science Characterization Early Results Symposium. February 22-24, 1983. Greenbelt, Maryland (NASA Goddard Space Flight Centre), NASA publ. 2355, pp.281-297.

Schneider, S., 1974. Luftbild und luftbildinterpretation. Walter de Gruyter. Berlin-New York, 1974.

Scheffé, H., 1959. The analysis of variance. John Wiley & Sons Inc., New York, USA.

Sheffield, C., 1985. Selecting Band Combinations from Multispectral Data. *Photogrammetric Engineering and Remote Sensing*, Vol. 51, No. 6, pp. 681-687.

Silva, E.S., 1990. Geografía de Costa Rica. Segunda edición corrigida y aumentada. Universidad Estatal A Distancia, San José, Costa Rica.

Singh A., 1984. Some clarifications about the pairwise divergence measure in remote sensing. *Int. J. Remote Sensing*, Vol.5, no.3, pp. 623-627.

Singh A., 1987. Spectral separability of tropical forest cover classes. *Int. J. Remote Sensing*, 1987, vol.8., no.7, pp.971-979.

Skidmore A.K., 1989. An expert system classifies eucalypt forest types using thematic data and a digital terrain model. *Photogrammetric engineering and remote sensing*, vol. 55, no. 10, pp.1449-1464.

Skidmore, A.K. and B.J. Turner, 1988. Forest Mapping Accuracies Are Improved Using a Supervised Nonparametric Classifier with SPOT Data. *Phot. Engineering and Remote Sensing*, Vol. 54, No. 10, pp. 1415-1421.

SPSS, 1983. SPSS^x User's Guide. SPSS Inc., McGraw Hill Book Company, U.S.A.

Srinivasan, A. and J.A. Richards, 1990. Knowledge-based techniques for multi-source classification. *Int.J.Remote Sensing*, 1990, Vol. 11, No. 3, p 505-525.

Stein, A., M. Hoogerwerf and J. Bouma, 1988. Use of soil map delineations to improve (co-)kriging of point data on moisture deficits. *Geoderma*, 43: 163-177.

Strahler, A.H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote sensing of environment* 10: 135-163 (1980).

Suits, G.H., 1972. The Cause of Azimuthal Variations in Directional Reflectance of Vegetative Canopies. *Remote Sensing of Environment* 2, 175-182.

Swain, P.H., 1978. Bayesian Classification in a time-varying environment. *IEEE transactions on systems, man and cybernetics*, vol. smc-8, no. 12, dec 1978.

Swain, P.H., and S.M. Davis (eds.), 1978. 'Remote Sensing. The Quantitative Approach'. Mc Graw-Hill, New York, 396 p.

Swain, P.H. and H. Hauska, 1977. The decision tree classifier: Design and Potential. *IEEE Transactions on geoscience electronics*, vol. GE-15, no. 3, July 1977.

Swain, P. H., Robertson, T. V., and Wacker, A. G., 1971. Comparison of the divergence and B-distance in feature selection. LARS Information Note 020871, Laboratory for the Application of Remote Sensing. Purdue University, West Lafayette, Indiana, U.S.A.

Thaman, R., 1974. Remote sensing of agricultural resources. In: J.E. Estes and L.W. Senger (EDS), Remote Sensing. Techniques for environmental analyses.

The Tico Times, 1991, the june 7 issue; English language weekly, San José, Costa Rica.

Ton, J., J. Sticklen and A.K. Jain, 1991. Knowledge-based segmentation of landsat images. *IEEE transactions on geoscience and remote sensing*, vol. 29, no. 2, pp. 222-232.

Tucker, C.J., 1977. Spectral estimation of grass canaopy variables. *Remote Sensing Environ.*, 6: 11-26.

Tucker, C.J., 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment* 8: 127-150.

Tucker, C.J., 1980. Remote Sensing of Leaf Water Content in the Near Infrared. *Remote Sensing of Environment* 10: 23-32.

Tucker, C.J. and M.W. Garratt, 1977. Leaf optical system modeled as a stochastic process. *Applied Optics*, Vol. 16, No. 3, pp. 635-642.

USDA, 1975. Soil Taxonomy. A basic system for soil classification for making and interpreting soil surveys. Agricultural handbook no. 6, SCS, U.S. Governmental printing office, Washington.

Veldkamp E., E.J. Huising, A. Stein and J. Bouma, 1990. Variability of measured banana yields in a Costarican plantation as expressed by soil survey and thematic mapper data. *Geoderma* 47, pp. 337-348.

Vogelmann, J.E., 1988. Detection of forest change in the green mountains of vermont using multispectral scanner data. *Int. J. Remote Sensing*, Vol. 9, no. 7, pp. 1187-1200.

Vrana, R., 1989. Historical data as an explicit component of land information systems. *Int. J. Geographical Information Systems*, Vol. 3, No. 1, pp. 33-49.

Waaijenberg, H. (Ed.), 1990. Rio Jiménez, ejemplo de la problemática agraria en la Zona Atlántica de Costa Rica. Un análisis con enfoque histórico. Serie técnica, Informe técnico No. 160, Centro Agronómico Tropical de Investigación y Enseñanza, CATIE, Turrialba, Costa Rica.

Wielemaker, W. (Ed.), 1990. Colonización de las lomas de cocori. Deforestación y utilización de los recursos de tierra en la Zona Atlantica de Costa Rica. Serie técnica, Informe técnico No. 157, Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Costa Rica.

Wielemaker, W.G. and A.P. Oosterom, 1992. Un sistema de información para paisajes y suelos. In W.G. Wielemaker and S.B. Kroonenberg (ed.) Generación y applicatión de la información de suelos de la Zone Atlantica de Costa Rica. Actas del taller información de suelos. Guapiles 2-4 Oct. 1990. Serie Técnica, informa técnico no. 170.

Wijngaarden, W. van and A. Kooiman. 1988. CUMU: the land cover and land use database. *ITC-Journal* 1988-1. International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede.

Wilkinson, G.G. and J. Mégier, 1990. A belief function approach for using GIS derived spatial context in satellite image understanding. Proceedings European conference on geographical information systems, Amsterdam, The Netherlands, April 10-13. EGIS Foundation, Utrecht, The Netherlands.

Williamson, H.D. 1988. Evaluation of middle and thermal infrared radiance in indices used to estimate GLAI. *Int. J. Remote Sensing*, 1988, vol. 9, no. 2, pp. 275-283.

Wu, J., D. Cheng, W. Wang and D. Cai, 1988. Model-based remotely-sensed imagery interpretation. *Int.J.Remote Sensing*, 1988, Vol. 9, No. 8, p.1347-1356.

Young, A., 1986. Land evaluation and agroforestry diagnosis and design: Towards a reconciliation of procedures. *Soil Survey and Land Evaluation*, Vol. 5, pp. 61-76.

Zuviría, M. de, 1992. Mapping agrotopoclimates by integrating topographic, meteorological and land ecological data in a geographic information system. A case study of the Lom Sak area, north central Thailand. ITC publication number 14. International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede.

Samenvatting

Deze dissertatie beschrijft de landgebruiksinventarisatie op sub-regionale nivo in de Atlantische Zone van Costa Rica. Hiertoe wordt het concept van de "landgebruikszone" (LGZ) geïntroduceerd. De LGZ speelt een centrale rol in de definitie van een methodologie ter waarneming van landgebruik en landgebruiksverandering. Het landgebruik wordt beschreven a.d.h.v. landgebruikspatronen (LGP). Het LGP geeft aan welke bedrijfssystemen en landgebruikstypen voorkomen in een LGZ.

Een "object-georiënteerde" en "gegevens-gedreven" benadering wordt gehanteerd voor de landgebruiksinventarisatie. "Object-georiënteerd" betekent in dit geval dat het landgebruik wordt beschreven als een verzameling objecten (de LGZ) met hun specifieke geometrische en thematische karakteristieken. Een klassificatiesysteem wordt ontwikkeld, zodat elke klasse LGZ's bevat met een kenmerkende thematische inhoud, geometrie, aggregatie structuur en dynamiek. De verwerking van dergelijke complexe object informatie vereist aandacht voor de definitie van een gegevens model.

Satelliet beelden en luchtfoto's worden aangewend om informatie over het landgebruik te verkrijgen. Dit gebeurt d.m.v. patroon herkenning. De "gegevens-gedreven" benadering houdt in dit geval in dat de patronen (of klassen) niet a priori geven zijn, maar inductief bepaald worden. D.w.z. ze vloeien voort uit het inventarisatieproces. De gegevens-gedreven benadering is een strategie om inzicht te krijgen in het landgebruik op sub-regionaal nivo. Het inventarisatieproces is opgedeeld in een aantal sequentieel geordende stappen, die in afzonderlijke hoofdstukken beschreven worden.

Het eerste gedeelte van deze dissertatie gaat in op de LGZ als middel voor de inventarisatie van landgebruik en landgebruiksverandering. De vergelijking van luchtfoto's van de periode 1948-1952 met die van 1984, voor het Guacimo-Rio Jiménez-Siquirres (GRS) studiegebied, toont dat de LGZ stabiele grenzen hebben, mits deze tot de landbouwgebieden behoren. Dit impliceert dat we dezelfde LGZ's op de verschillende tijdsmomenten kunnen herkennen aan hun geometrie en waarmee registratie van landgebruiksverandering d.m.v. de thematische kenmerken mogelijk wordt gemaakt.

Informatie omtrent bedrijfsgrootte en bedrijfssystemen is verkregen d.m.v. een enquête onder boeren. De gegevens laten significante verschillen in samenstelling van de LGZ's. Op grond van deze verschillen zijn de landgebruikspatronen gedefinieerd. Dat de LGZ's geassocieerd zijn met een verscheidenheid aan landgebruikspatronen geeft aan dat de LGZs relevante ruimtelijke eenheden vertegenwoordigen voor de beschrijving van het landgebruik op sub-regionale schaal. De verschillen in landgebruikspatroon blijken gerelateerd te zijn aan verschillen in samenstelling van de landbedekking en aan de veldgrootte verdeling, zoals deze verkregen zijn d.m.v. classificatie en interpretatie van de satellietbeelden en luchtfoto's. Deze materialen geven dus informatie over het landgebruik, mits mits de geëigende interpretatieregels gehanteerd worden.

Bovenstaande impliceert eveneens dat verandering in het landgebruikspatroon afgeleid kan worden uit verandering in landbedkkingssamenstelling, onder de voorwaarde dat de geometrie van de LGZ niet verandert, en mits ook hier de geëigende interpretatieregels worden toegepast. Veranderingen in landbedekkingssamenstelling van de LGZ's tussen 1986 en 1990 zijn onderzocht m.b.v. satellietbeelden. Duidelijke trends in verandering van landgebruik werden waargenomen. Deze trends waren: een afname in maïs- en graslandareaal, een toename in het areaal voor de produktie van banaan en macadamia en toename van land bestemd voor herbebossing. Naast toe- of afname in areaal van een bepaald gewas konden bijvoorbeeld ook verandering van type grasland en verbetering van bananaplantages waargenomen worden.

Het tweede gedeelte beschrijft identificatie en klassificatie van de LGZ d.m.v. patroonherkenning. Allereerst wordt de stratificatie van het GRS studiegebied beschreven. Sub-regio's worden geïdentificeerd waarbij het ruimtelijk patroon, bepaald door de grootte, vorm en ordening van de velden, het belangrijkste criterium is.
Wanneer de LGZ's eenmaal vastgesteld zijn, kunnen de veldgrootte en landbedekkingskarateristieken bepaald worden. Een procedure voor de per pixel landbedekkingsklassificatie wordt besproken. Deze procedure begeleidt de beeldanalist in de complexe taak een verzameling trainingsklassen voor de 'maximum likelihood' klassificatie te definiëren. Nadruk wordt gelegd op de 'training' fase. De methode maakt gebruik van 'gecontroleerde' (supervised) en ongecontroleerde (unsupervised) klassificatie methoden.
De classificatie van de LGZ's wordt in aparte hoofdstukken behandeld. Op grond van statistische evaluatie van verschillen in gemiddelde veldgrootte zijn 5 relevante klassen gedefinieerd. De klassen om verschillen in landbedekkingssamenstelling te beschrijven worden gedefinieerd a.d.h.v. een hiërarchische clusteranalyse. Hierbij wordt een kritische afstand bepaald, die aangeeft of twee LGZ's (of groepen van LGZ's) significant van elkaar verschillen. Deze kritische afstand weerspiegelt de betrouwbaarheid van de gegeven landbedekkingssamenstelling, die op zijn beurt bepaald wordt door de nauwkeurigheid van de landbedekkingsklassificatie (verwijzend naar zowel thematische als geometrische nauwkeurigheid).

De veldgrootteklassen en de klassen van landbedekkingssamenstelling worden gegevensklassen genoemd. Uit deze klassen wordt informatie over het landgebruik afgeleid d.m.v. projectie van de gegevensklassen op informatiecategorieën. Dit gebeurt a.d.h.v beslisregels, die een conditioneel klasselabel toe kennen aan een LGZ. De conditie refereert aan een bepaalde context. Inzicht in de ingewikkelde beslisstructuur wordt verkregen door ordening van de regels in een hiërarchische structuur, hetgeen resulteert in een beslisboom. De beslisboom geeft een stapsgewijze specificatie van het landgebruikspatroon van een LGZ.

In het derde gedeelte wordt het landgebruik geëvalueerd m.b.t. het bio-fysische potentieel van het land. Voor het studiegebied zijn hiertoe de landgebruikskaart en de bodemkaart over elkaar gelegd. Het blijkt dat de grenzen van de landgebruikseenheden en de fysiografische-bodemeenheden in grote mate overeenkomen. Dit betekent niet dat het landgebruik in overeenstemming is met de bio-fysische capaciteit van de bodem. Resultaten geven aan dat in 18% van het gebied een gevaar voor degradatie van het land bestaat. In 51% van het gebied bestaan mogelijkheden voor meer intensief gebruik van de bodem. Gezien het samengestelde karakter van de eenheden kunnen combinaties van landgebruikstype en bodemtype niet met zekerheid vastgesteld worden. De studie heeft als zodanig een verkennend karakter.

In het laatste hoofdstuk wordt de opbrengstvariatie van banaan in één bepaalde plantage onderzocht. De plantage vertegenwoordigt een LGZ. Gegevens van de bodeminventarisatie verklaarden 67% van de variatie in opbrengst. Het gebruik van Landsat-TM gegevens als een additionele verklarende factor resulteerde niet een beter resultaat.

Curriculum Vitae

Jeroen Huising, born in Eibergen, The Netherlands at september 25, 1958.

In 1976 I finished secondary education in Enschede and started my study in soil science at the Wageningen Agricultural University. I graduated in 1984 at this university, majoring in tropical soil science and clay mineralogy. During this period I spend considerable time in Africa doing a reconnaissance soil survey and making preparations for research on red and yellow tropical soils, their physical behavior and clay mineralogy. Part of my study I did at the University of Amsterdam. This concerned the subjects international (economic) affairs and development economics. I worked for several months at the institute for research on multinational enterprises, investigating macro trends in dutch foreign investments as part of a research on dutch capital export.

From 1985 to 1987 I was employed at the ITC in Enschede (International Institute for Aerospace Survey and Earth Sciences). I was involved in the development of a PC-based GIS, which came to be known as the Integrated Land and Watershed management Information System (ILWIS). I also was involved with the GIS-LIS course, giving workshops on data base design for the rural and cadastral applications.

In 1987 I started my PhD research on the use of remote sensing and GIS for land use inventory in the Atlantic Zone of Costa Rica. I was seconded to the Atlantic Zone Programme, a programme of cooperation between the 'Centro Agronómico Tropical de Investigación y Enseñanza' (CATIE), the Wageningen Agricultural University (WAU) and Ministerio de Agricultura y Ganaderia (MAG). I returned to the Netherlands in 1991 to finalize my thesis. In 1992 I went back to Costa Rica for several months to do a consultancy for the Worldbank concerning the use of remote sensing and GIS for the assessment of national and regional deforestation trends.

For 1993 I am granted a post-doctoral fellowship at the International Training Centre for Post-Graduate Soil Scientists. The fellowship is funded by the European Economic Community.