Potential of the kNN Method for Estimation and Monitoring off-Reserve Forest Resources in Ghana

Thesis submitted in partial fulfilment of the requirements of
the degree Doctor rer. nat. of the
Faculty of Forest and Environmental Sciences,
Albert-Ludwigs-Universität
Freiburg im Breisgau, Germany

by

Christian Kutzer

Freiburg im Breisgau, Germany
2008

Name of Dean:	Prof. Dr. Heinz Rennenberg
Name of Supervisor:	Prof. Dr. Dr. h. c. Dieter R. Pelz
Name of 2nd Reviewer:	Prof. Dr. Barbara Koch
Date of thesis defence:	23rd June 2008

Acknowledgements

First I would like to thank my supervisor Prof. Dr. Dr. h. c. Dieter R. Pelz, director of the Department of Forest Biometry, for accepting me as his student, for his patience, and invaluable advice to complete this study. I am grateful to Tropenbos International for cooperating and supporting fieldwork and data acquisition. I also thank Prof. Dr. Barbara Koch, director of the Department of Remote Sensing and Landscape Information Systems, for taking up the role of co-referent of this study.

I am very much indebted to Dr. Francis Bih and his family, with whom I worked and lived during my stay in Ghana. For his extraordinary efforts during fieldwork, I thank Kwame Dankwa. Mr S. K. Nketiah and the entire staff of Tropenbos International-Ghana deserve my thanks for their warm reception, support and hospitality.

I would like to thank Dr. Patricia Clesly for proofreading my English. Thanks also go to Dr. Roberto Scoz, Dr. Lilian Soto, Dr. Weeraphart Khunrattanasiri, and to all my colleagues in the Department of Forest Biometry and the Department of Remote Sensing and Landscape Information Systems, who have contributed to the success of my work.

Finally I would like to thank Dr. Wolfgang Stümer, who developed and provided the kNN programme for the calculations of this study.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AOI	Area of Interest
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
CCD	Charge-Coupled Device
CNES	Centre National d´Etudes Spatiales
DEM	Digital Elevation Model
EDG	Earth Observing System Data Gateway
EOS	Earth Observing System
EPE	Estimated Position Error of the GPS receiver
ETM^+	Enhanced Thematic Mapper Plus
FAO	Food and Agriculture Organization
FSD	Forest Services Division
GIS	Geographic Information Science
GPS	Global Positioning System
IFOV	Instantaneous Field of View
IR	Infrared
kNN	k Nearest Neighbour
LUS	Land Use System
LUT	Land Use Type
MODIS	Moderate Resolution Imaging Spectrometer
MS-NFI	Finnish multi-source National Forest Inventory
MSS	Multispectral Scanner
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NTFP	Non-timber Forest Product
OA	Overall Accuracy
PCA	Principal Component Analysis
RS	Remote Sensing
SD	Standard Deviation
SWIR	Short Wave Infrared
TBI	Tropenbos International
TIR	Thermal Infrared
TM	Thematic Mapper

UTM	Universal Transverse Mercator
VNIR	Visible and Near Infrared
WGS 84	World Geodetic System 1984
WD	Wildlife Division

Abstract

A foreseeable exhaust of fossil fuels and the economic development of emerging nations accelerates the transformation of forest lands into monocultures, e.g. for bio fuel production. On this account, cost efficient methods to enable the monitoring of land resources has become a vital ambition. The application of remote sensing techniques has become an integral part of forest attribute estimation and mapping. The aim of this study was to evaluate the potentials of the kNN method by combining terrestrial with remotely sensed data for the development of a pixel-based monitoring system for the small scaled mosaic of different land use types of the off-reserve forests of the Goaso forest district in Ghana, West Africa. For this reason, occurrence and distribution of land use types like cocoa and non-timber forest resources, such as raphia palms, were estimated, applying the kNN method to ASTER data. Overall accuracies, ranging from 79 % for plantain, to 83 % for oil palms, were found for single-attribute classifications, whereas a multi-attribute approach showed overall accuracies of up to 70 %. Values of k between 3 and 6 seem appropriate for mapping bamboo and others. Optimisation of spectral bands improves results considerably.

Key Words: k nearest neighbour, combination, terrestrial, remote sensing, estimation, distribution, non-timber forest product

1 Introduction

Like most tropical countries, the forest resources of Ghana are decreasing at alarming figures. Agricultural expansion has been identified as the common factor in deforestation (FAO, 2003). Farming is the predominant occupation in the forest zone of Ghana (Abagale et al., 2003). In many respects, degradation of Ghanaian's off-reserve forests indirectly increases pressure on forest reserves. About one third to two-thirds of the timber harvested annually in Ghana comes from off-reserve forests (Mayers et al., 1996).

About twenty per cent of Ghana's high forest zone is under reservation and now represents the only permanent forest estate. It is mainly under the control of the Forest Services Division (FSD) and in parts under the Wildlife Division (WD). The remaining 80 per cent – the so-called off-reserve forests – have been chiefly converted to farmland over the past few decades (Mayers et al., 1996; Kotey et al., 1998). Nowadays these lands outside forest reserves exhibit a mosaic of agricultural field, fallow lands, secondary forest patches, and settlements. However, the farm and fallow areas host extensive forest resources. Despite the ruling farming activities in the high forest zone, trees on farms are very common. Utilizable tree species on farmland belong to communities, not to the landowner. Nevertheless, it is an offence for an individual or a community to cut or sell merchantable tree species without permission from the Forest Department. Loggers are required by law to compensate farmers for damage to food and cash crops resulting from logging operations on their land. Tragically, they rarely comply. Amongst other reasons, this frustration leads some farmers to illegally destroying valuable tree species (Owubah et al., 2001).

FAO is convinced that it is fundamental for forestry and agriculture to work hand in hand. The management of forests and trees, including the use of agroforestry and watershed management, is an integral part of the effort to reduce food insecurity, minimize poverty and improve environmental quality for the rural poor (FAO, 2003).

It is of great significance if the condition of forest land is able to provide the variety of goods and services required from them. However, wise and prospective planning and management decisions on the level of stakeholders, local authorities and policy-

makers can only be undertaken when adequate information on land resources are available.

Within the forest reserves, the Forest Services Division of The Forestry Commission has acquired information such as botanical composition and quality, as well as the extent of these resources (Hawthorn et al., 1993). Among frequent local forest inventories, a comprehensive national forest inventory was conducted between 1986 to 1992 and provided sufficient information to resource managers to utilise both, timber and non-timber forest products (Affum-Baffoe, 2001). Nevertheless, the inventory did not cover the off-reserve forests. While forest reserves are repeatedly inventoried and monitored, the only forest-/ tree-inventory in the off-reserve forests was carried out in 1995, simply for timber trees (Van Leeuwen, 2006). More supportive policy-making and adequate management of natural resources needs pertinent, timely and precisely quantitative and qualitative data on the condition of land and forest resources and their dynamics. One of the basic prerequisites for better use and sustainable development of land is information on existing land use patterns and changes in land use through time (Anderson et al., 1976).

1.1 Background

Forest inventories, particularly in the tropics, are a time and cost-intensive source of information for the management of natural resources. The expenses for the generation of detailed forest stand maps, simply by the use of terrestrial methods, is likely to exceed the budget of forest planning. The capabilities of remote sensing (RS) and geographic information science (GIS) for planning, mapping, and monitoring of land changes, are rising. Particularly with regard to extensive areas, the application of remote sensing for forest inventories will result in the reduction of costs. This technology has been used increasingly as a tool to measure spatial and temporal patterns of land cover in many regions of the world (e.g. Alo & Pontius, 2006; Ngamabou, 2006; Ratanopad & Kainz, 2006). Practical application of remote sensing in the last few decades is presented by Sohlberg & Sokolov (1986).

It is this point that this study is particularly focused upon. It is the intention to develop and adopt and assess a methodology to generate distribution maps of selected land use types by combining terrestrial with spectral data. The k-nearest neighbour algorithm (kNN) is a method for classifying objects based on closest training examples in the feature space. With the combination of terrestrial samples and satellite data through the application of the k-nearest neighbour method, forest inventories, detailed stand maps and image classifications have already been supported (Holmgren et al., 2000; Holmström, 2001; Holmström et al., 2001; Gustafsson, 2002; Stümer, 2004; Tomppo, 2005). Rajaniemi et al. (2003) studied the variation of rain forest vegetation by means of satellite images and field data.

1.2 Objectives

The general objective of the study is to evaluate the potential of the kNN method applied as regards the combination of terrestrial sampling and remote sensing to estimate non-timber forest products and land resources, to inventory off-reserve tree and forest resources, and to generate resource distribution maps.

The specific objective is to assess different parameters and their contribution to estimation accuracy. The desired results of this study are guidelines and recommendations for the development of a monitoring system for specific non-timber forest products, tree and forest resources, and other typical land uses in the off-reserve areas. Major land use types (e.g. cocoa & oil palm plantations), and locally important non-timber forest resources (e.g. bamboo & raphia palms) will be selected and their ability to be identified with the methodology will be verified and quantified. To achieve this target, a field data collection procedure has to be adjusted and the small-sized mosaic of different land use types must thus be registered and defined.

In this study, the kNN method is used for categorical data. For this purpose, the kNN programme developed by Stümer (2004) is applied. Data collection and sampling procedure will be conducted in collaboration with Tropenbos International-Ghana (TBI). The following results are expected:

- Assessment of non-timber forest resources and land use types that can be estimated with the methodology
- Assessment of the sampling design and procedure
- Assessment of band number versus band selection
- Identification of parameter values of high significance
- Benchmarks in respect to accuracy and overall agreement
- Guidelines and recommendations for the development of a monitoring design

With respect to the specific characteristics for conducing forest inventories in developing countries in the tropics, some conditions have to be met (FAO, 1976) and should enter into the inventory design. Thus particular attention as regards infrastructure, funding, facilities, labour, etc. to enable practicability, has to be taken into consideration.

1.3 Framework of the study

This study arose in connection with the Ghana Programme of Tropenbos International (cp. http://www.tropenbos.nl/sites/site_ghana.php). The mission of TBI-Ghana is to generate scientific input for the sustainable management of natural resources of Ghana's high forest zone. The Department of Forest Biometry of the Albert-Ludwigs-University Freiburg, is one of TBI-Ghana's partners, where this study together with a study dealing with assessment methods of non-timber forest products (NTFP) in off-reserve forests (Bih, 2006), was accompanied.

In the project section, geo-information applications of off-reserve tree management in the Goaso district, the main research objective is to improve forest management of off-reserve land resources in Goaso. Therefore new methodologies will be developed or existing ones will be adapted to assess availability and potential of off-reserve resources.

The expected outputs are:

To extend the existing forest management planning information system for Goaso district, by including currently missing, but essential, data on: a. Off-reserve tree resources. b. Off-reserve NTFPs. c. Off-reserve tree resource use and management by local communities and logging companies.

To develop several – alternative – scenarios for improved off-reserve tree resource management through: a. Matching of land/tree resource uses that are relevant to local stakeholders with the bio-physical potential of land/tree resources. b. Identification of incentive mechanisms to facilitate implementation of selected scenarios.

To increase the capacity of relevant institutions in Ghana in the application of geo-information for forest/tree resources management.

2 State of the Art

2.1 Remote Sensing and Survey Instruments

Remote sensing is the small or large-scale acquisition of information of an object or phenomenon, which is not in physical contact with the object (Campbell, 2002). The term was introduced in the United States of America, when the first images of the earth were taken from space in the 60^{ies} of the 20th century (Kraus & Schneider, 1988). Multispectral remote sensing is defined as the collection of reflected, emitted, or backscattered energy from an object or area of interest in multiple bands of the electromagnetic spectrum (Jensen, 2000). It includes data collection in hundreds of bands (Logicon, 1997).

With the launch of Landsat 1 in 1972, the civil use of satellite based remote sensing began (Kellenberger, 1996; Jensen, 2000). However, remote sensing had already been involved in military and civilian organisations before 1972. The broad interest of scientists, politicians and planners to survey, map and monitor the earth surface has led to the launch of a variety of satellites carrying different sensors. Besides cameras as the traditional aerial photographic survey system, digital line by line scanners have been developed. They are differentiated into passive systems, which survey reflected or emitted radiation in specific wavelengths from the earth surface, and active remote sensing systems, e. g. spaceborne radar systems, which illuminate or radiate objects artificially while their reflex reflection is detected (cp. Albertz, 1991; Löffler, 1994; Hildebrand 1996; Kraus & Schneider, 1998; Richards & Xiuping, 2005). There is a tremendous variety of digital multispectral and hyperspectral remote sensing systems. An overview of selected operational available satellites und the corresponding passive and active sensors, which are used for remote sensing of the earth surface, are summarised in table 2-1.

Table 2-1 Earth Observation Satellites. The following table includes those satellites which are currently operational. It is limited to sensors which are primarily intended for remote sensing of the Earth surface. (Source: after Environmental Remote Sensing Center, University of Wisconsin-Madison[1])

Satellite Name	Source	Launch	Sensors	Types	No. of Bands	Resolution [meters]
Landsat-5	US	1984	MSS	Multispectral	4	82
			TM	Multispectral	6	30
					1	120
SPOT-1	France	1986	HRV	Multispectral	3	20
				Panchromatic	1	10
SPOT-2	France	1990	HRV	Multispectral	3	20
				Panchromatic	1	10
IRS-1B	India	1991	LISS-I	Multispectral	4	72.5
			LISS-II	Multispectral	4	36.25
ERS-1	ESA	1991	AMI	Radar	1	26
			ATSR	Multispectral	4	1000
RESURS-O1-3	Russia	1994	MSU-SK	Multispectral	4	170
					1	600
NOAA-14	US	1994	AVHRR	Multispectral	5	1100
IRS-1C	India	1995	WiFS	Multispectral	2	188
			LISS-III	Multispectral	3	23
					1	70
			Pan	Panchromatic	1	5.8
ERS-2	ESA	1995	AMI	Radar	1	26
			ATSR	Multispectral	4	1000
RADARSAT 1	Canada	1995	SAR	Radar	1	9-100
OrbView-2 (SeaStar)	US/Orbimage	1997	SeaWiFS	Multispectral	8	1130
IRS-1D	India	1997	WiFS	Multispectral	2	188
			LISS-III	Multispectral	3	23
					1	70
			Pan	Panchromatic	1	5.8
SPOT-4	France	1998	VI	Multispectral	4	1150
			HRV	Multispectral	4	20
				Panchromatic	1	10
NOAA-15 (NOAA-K)	US	1998	AVHRR	Multispectral	5	1100
Landsat-7	US	1999	ETM+	Multispectral	6	30

					1	60
				Panchromatic	1	15
ROCSAT-1	Taiwan	1999	n/a	Panchromatic	n/a	2
				Multispectral	n/a	8
IRS-P4 (Oceansat)	India	1999	OCM	Multispectral	8	360
IKONOS	Space Imaging	1999	IKONOS	Multispectral	4	4
				Panchromatic	1	1
Terra (EOS AM-1)	US	1999	ASTER	Multispectral	14	15,30,90
			MISR	Multispectral	4	275
			MODIS	Multispectral	36	250,500,1000
NOAA-L	US	2000	AVHRR	Multispectral	5	1100
EO-1	US	2000	ALI	Panchromatic	1	10
				Multispectral	9	30
			Hyperion	Hyperspectral	220	30
			LAC	Hyperspectral	256	250
EROS-A1	ImageSat International	2000	Panchromatic	Panchromatic	1	1.5
QuickBird	DigitalGlobe	2001	Multispectral	Multispectral	4	2.44
			Panchromatic	Panchromatic	1	0.61
MTI	US	2001	MTI	Multispectral	15	5
Envisat-1	ESA	2002	ASAR	Radar	1	30, 150
			MERIS	Multispectral	15	300
Aqua (EOS PM-1)	US	2002	MODIS	Multispectral	36	300, 1200
SPOT-5	France	2002	HRV	Multispectral	3	10
					1	20
				Panchromatic	1	2.5, 5
NOAA-M	US	2002	AVHRR	Multispectral	5	1100
ICESat	US	2003	GLAS	Lidar	2	70
CBERS 2	China/Brazil	2003	HRCC	Multispectral	4	20
				Panchromatic	1	20
			IRMSS	Panchromatic	1	80
				SWIR	2	80
				TIR	1	160

			WFI	Wide Field Imager	2	260
OrbView-3	Orbimage	2003	OrbView	Multispectral	4	4
				Panchromatic	1	1
ROCSAT-2	Taiwan	2004	PAN	Panchromatic	1	2
			MS	Multispectral	4	8
IRS-P6 (ResourceSat-1)	India	2004	LISS 3/4	Multispectral	7	5.8, 23.5
			AWiFS	Multispectral	3	80

[1]http://www.ersc.wisc.edu/resources/EOSC.php, updated: 02/18/2005

While three-band multispectral data from aerial photography are sufficient for many applications, more spectral bands postitioned at optimum locations throughout the electromagnetic spectrum might be much better for specific applications. The enhancement of the survey systems tends to result in a continuous increase of the spectral resolution and spectral coverage up to hyperspectral remote sensing. Systems, which show a high spectral resolution with scores of channels, as well as an almost continuous spectral coverage in the optical spectral range, are called hyperspectral sensors (cp. Goetz et al., 1985; Goetz, 1992; Ball, 1995; Clark, 1999). Image sequences with a higher number of spectral bands than number of spectral distinctive classes are called hyperspectral images.

Digital remote sensor data are usually stored as a matrix of numbers. Each digital value is located at a specific row and column in a matrix. The smallest non-divisional element of a digital image is called *pixel*. The pixel features a brightness value bound with each row and column in the image. The dataset consists of a defined number of individual multispectral bands. These bands are all geometrically registered to one another. Remote sensor data quantized to 8 bits have brightness values that range from 0 to 255.

Landsat 1 to 3 were launched into circular orbits around Earth at a nominal altitude of 919 km. The satellites' orbital inclination was 99° and made them nearly polar. They orbited the Earth once every 103 minutes, resulting in 14 orbits per day. The Landsat

Multispectral Scanner (MSS) was placed on Landsat satellites 1 through 5. Sensors such as the Landsat MSS and TM are optical-mechanical systems in which a mirror scans the terrain perpendicular to the flight direction. Six parallel detectors sensitive to four spectral channels (bands) in the electromagnetic spectrum viewed the ground simultaneously: 0.5 to 0.6 µm (green), 0.6 to 0.7 µm (red), 0.7 to 0.8 µm (reflective infrared), and 0.8 to 1.1 µm (reflective infrared). These bands were originally numbered 4, 5, 6, and 7. The instantaneous field of view (IFOV) of each detector was square and resulted in a ground resolution element of approximately 79 x 79 m (Jensen, 2000).

The Landsat Thematic Mapper (TM) is a scanning optical-mechanical sensor that records energy in visible, reflective-infrared, middle infrared, and thermal infrared regions of the electromagnetic spectrum. It collects multispectral imagery that has higher spatial, spectral, temporal, and radiometric resolution than Landsat MSS (SBRC, 1994).

Landsat 7 Enhanced Thematic Mapper Plus (ETM^+) was officially integrated into NASA's Earth Observing System (EOS) and was designed to work in harmony with NASA´s EOS Terra satellite which uses – except for ASTER – a suite of relatively coarse spatial resolution. The ETM^+ bands 1-5 and 7 are identical to those found on Landsat 4 and 5 and have the same 30 x 30 m spatial resolution. The thermal infrared band 6 now has 60 x 60 m spatial resolution instead of 120 x 120 m. Perhaps most notable is the new 15 x 15 m panchromatic band (0.52-0.90 µm).

The first SPOT satellite was developed by the French Centre National d´Etudes Spatiales (CNES) in cooperation with Belgium and Sweden. It was launched in 1986 and has a spatial resolution of 10 x 10 m (panchromatic mode) and 20 x 20 m (multispectral mode).

The survey systems MODIS and ASTER were launched with the American Terra-satellite EOS AM-1 in 1999. MODIS-data possess 36 spectral channels and a maximum resolution of 250 m, while ASTER-data achieve a resolution of 15 m with 14 channels (NASA, 2007). The ASTER satellite is a cooperative effort between

NASA and Japan's Ministry of International Trade and Industry. It obtains detailed information on surface temperature, emissivity, reflectance, and elevation, and is the only high-spatial resolution instrument on the Terra satellite. Further specifications of the Terra Spacecraft are shown in table 2-2.

Table 2-2 Specifications of the Terra Spacecraft (NASA, 2004).

Launch date:	December 1999
Orbit:	705 km altitude, sun-synchronous, so that at any given latitude it crosses directly overhead at the same time each day.
Orbit inclination:	98.3 degrees from the Equator
Orbit period:	98.88 minutes
Equator crossing:	10.30 a.m. (north to south)
Ground track repeat cycle:	16 days, i.e. every 16 days (or 233 orbits) the pattern of orbits repeats itself
Builder:	Lockheed Martin

The ASTER sensor consists of three separate instrument subsystems. Individual bandwidths and subsystem characteristics are summarized in table 2-3. ASTER's three subsystems are: the Visible and Near Infrared (VNIR), the Shortwave Infrared (SWIR), and the Thermal Infrared (TIR).

Table 2-3 ASTER Instrument Characteristics (NASA, 2004).

Characteristic	VNIR	SWIR	TIR
Spectral Range	Band 1: 0.52-0.60 µm Nadir looking	Band 4: 1.600-1.700 µm	Band 10: 8.125-8.475 µm
	Band 2: 0.63-0.69 µm Nadir looking	Band 5: 2.145-2.185 µm	Band 11: 8.475-8.825 µm
	Band 3: 0.76-0.86 µm Nadir looking	Band 6: 2.185-2.225 µm	Band 12: 8.925-9.275 µm
	Band 3: 0.76-0.86 µm Backward looking	Band 7: 2.235-2.285 µm	Band 13: 10.25-10.95 µm
		Band 8: 2.295-2.365 µm	Band 14: 10.95-11.65 µm
		Band 9: 2.360-2.430 µm	
Ground Resolution	15 m	30 m	90 m
Swath Width [km]	60	60	60
Detector Type	Si	PtSi-Si	HgCdTe
Quantization (bits)	8	8	12

The VNIR subsystem operates in three spectral bands at visible and near-IR wavelengths, with a resolution of 15 m (figure 2-1). It consists of two telescopes, one nadir-looking with a three-spectral-band detector, and the other backward-looking with a single-band detector. The backward-looking telescope provides a second view of the target area in band 3 for stereo observations. Thermal control of the charge-coupled device (CCD) detectors is provided by a platform-provided cold plate. Cross-track pointing to 24 degrees on either side of the track is accomplished by rotating the entire telescope assembly. Band separation is through a combination of dichromic elements and interference filters that allow all three bands to view the same ground area simultaneously. The data rate is 62 Mbps when all four bands are operating. Two on-board halogen lamps are used for calibration of the nadir-looking detectors. This calibration source is always in the optical path.

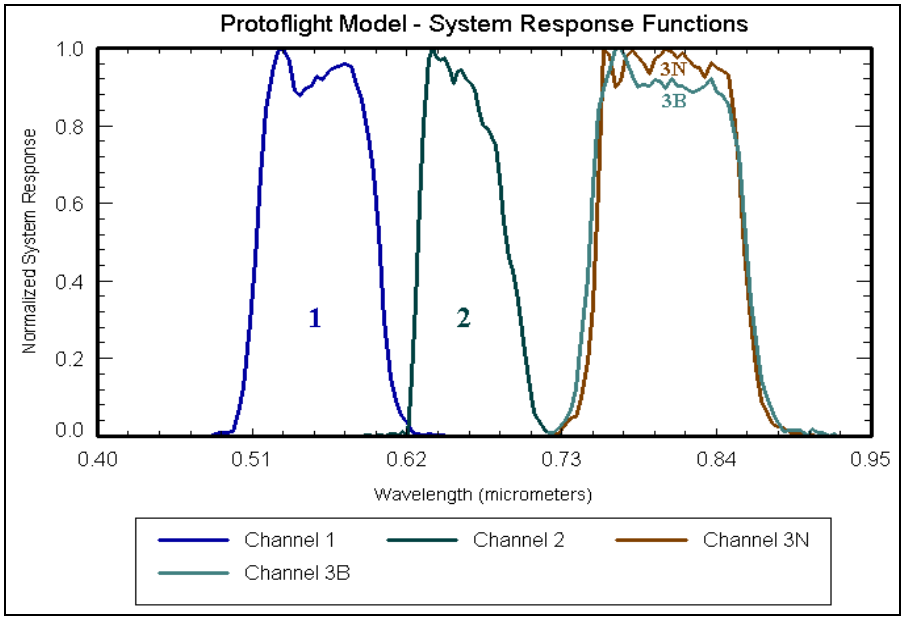


Figure 2-1 ASTER VNIR Chart (NASA, 2004).

The SWIR subsystem operates in six spectral bands in the near-IR region through a single, nadir-pointing telescope that provides 30 m resolution (figure 2-2). Cross-track pointing (± 8.550) is accomplished by a pointing mirror. Because of the size of the detector/filter combination, the detectors must be widely spaced, causing a parallax error of about 0.5 pixels per 900 m of elevation. This error is correctable if elevation data, such as a Digital Elevation Model (DEM), are available. Two on-board halogen lamps are used for calibration in a manner similar to that used for the VNIR

subsystem, however, the pointing mirror must turn to see the calibration source. The maximum data rate is 23 Mbps.

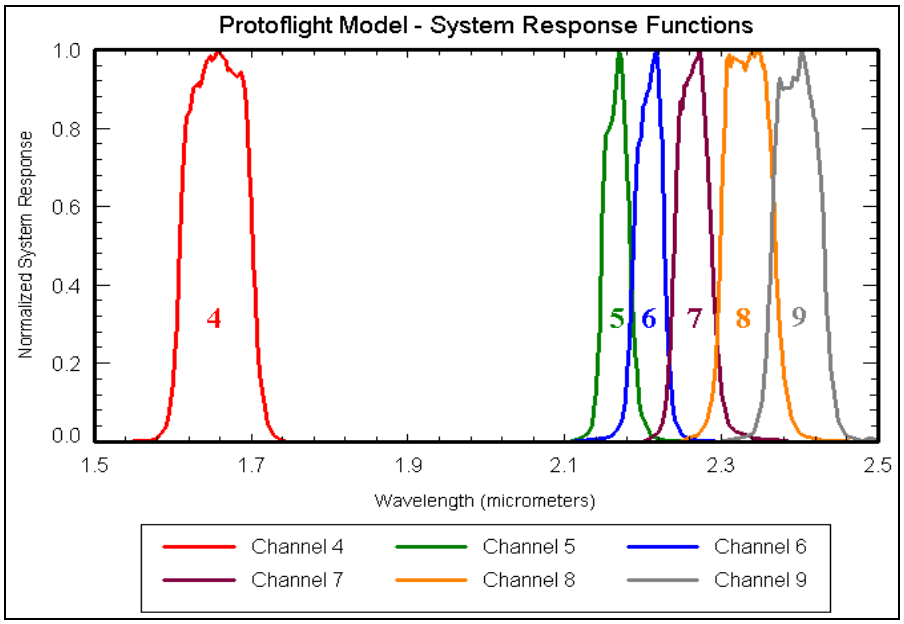


Figure 2-2 ASTER SWIR Chart (NASA, 2004).

The TIR subsystem operates in five bands in the thermal infrared region using a single, fixed-position, nadir-looking telescope with a resolution of 90 m. Unlike the other instrument subsystems, it has a "whiskbroom" scanning mirror. Each band uses 10 detectors in a staggered array with optical bandpass filters over each detector element. The maximum data rate is 4.2 Mbps. The scanning mirror functions both for scanning and cross-track pointing (to ± 8.55 degrees). In the scanning mode, the mirror oscillates at about 7 Hz and, during oscillation, data are collected in one direction only. During calibration, the scanning mirror rotates 90 degrees from the nadir position to view an internal black body. Because of the instrument's high data rate, restrictions have been imposed so that the average data rate is manageable by the spacecraft data management system. This restriction is a one-orbit maximum average rate of 16.6 Mbps and a two-orbit maximum average rate of 8.3 Mbps, which results in an approximately 9.3 % duty cycle.

ASTER products alone, or in combination with products of other sensors on the EOS-AM1 satellite, make a wide variety of research and applications possible (figure 2-3). Some examples of research and applications using ASTER products are vegetation

monitoring in tropical rain forests, classification of trees and pastures, measurement of planting and forest areas, monitoring of forest growth, investigation of damages caused by forest fire, making and updating maps, classification of land usage, investigation of human influence on the environment, et cetera (ASTER GDS, 2007).

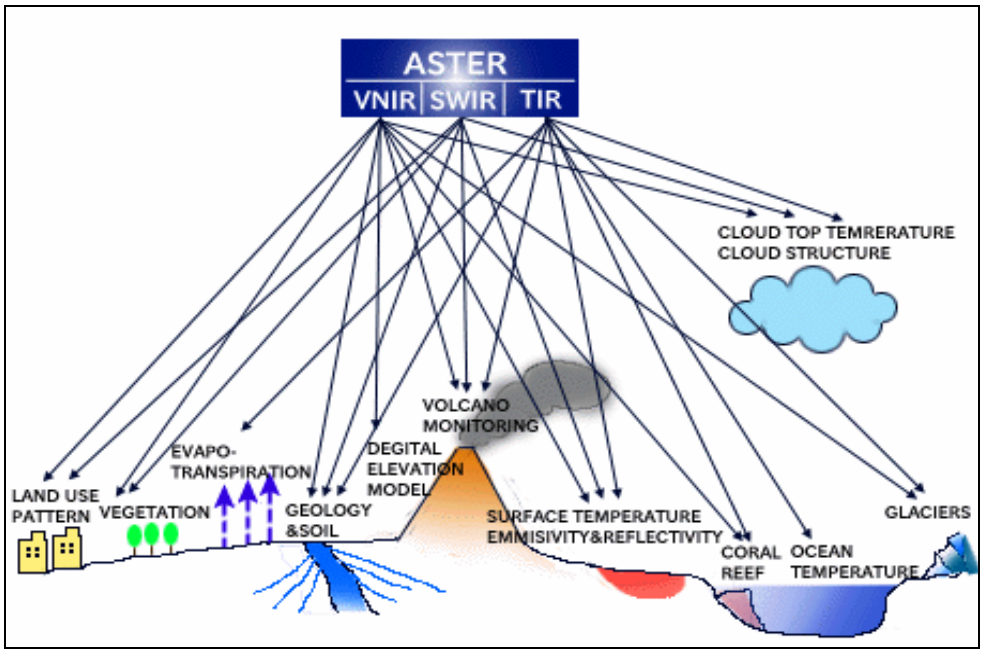


Figure 2-3 ASTER and its contribution to profound understanding of local and regional scale phenomena on and around the land surface. (Source: ASTER GDS, accessed on 08/22/2007 at http://www.gds.aster.ersdac.or.jp/gds_www2002/index_e.html)

2.2 Overview of the kNN Method

2.2.1 History of the kNN Method

The k-nearest neighbour method is a non-parametric classification algorithm and one of the oldest and simplest methods of pattern recognition in machine learning (Niemann, 1983; Altmann, 1992). It is described as an operation of the non-parametric discriminant analysis (Atkeson et al., 1997). K-nearest neighbour classification developed from the need to perform discriminant analysis when reliable parametric estimates of Bayes probabilities densities are unknown or difficult to determine for a given classification problem. It was Fix & Hodges (1951) who introduced a non-parametric method for pattern classification that has since become

known the k-nearest neighbour rule. The formal properties of the kNN rule were worked out by Cover & Hart (1967). Investigations on distance weighted approaches (Dudani, 1976; Bailey & Jain, 1978), and fuzzy methods (Jozwik, 1983; Keller et al., 1985) followed.

The kNN algorithm is a method for classifying objects based on the closest or most similar training samples in the feature space. It is a form of instance-based learning, also called lazy learning. An object is classified by a majority vote of its neighbours. This so-called nearest neighbour is determined by the use of distance functions, normally the Euclidian distance, though other distance measures, such as the Manhattan distance or Mahalanobis distance could, in principle, be used instead. Eventually, the unknown object is assigned to the class most similar amongst its k-nearest neighbours (figure 2-4).

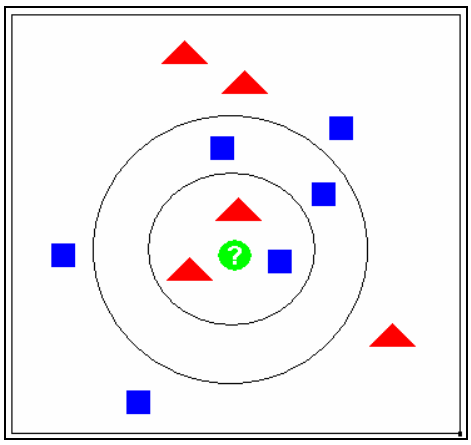


Figure 2-4 Example of kNN classification. The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If k = 3, it is classified to the second class, because there are 2 triangles and only 1 square inside the inner circle. If k = 5, it is classified to the first class (3 squares vs. 2 triangles inside the outer circle).

By simply assigning the property value for the object to be the average of the values of its *k-nearest* neighbours, the method can be used for regression. The neighbours are taken from a set of objects for which the correct classification (or property value) is known. This set can be defined as the training collective for the algorithm.

For the appraisal of the satellite data by means of the kNN method, reference pixel will be selected which correspond geographically to the terrestrial samples. For unknown pixel, i.e. pixel with no corresponding terrestrial sample information, the k-nearest neighbours in the spectral feature space will be identified by means of the Euclidian distance. For the selected k reference pixel, the corresponding values of the terrestrially recorded attributes are weighted with distance to the spectral feature space and assigned to the pixel. This procedure will be repeated for every pixel, till every pixel of the satellite image is designated by a value of the attribute.

The classification of image data in remote sensing serves only as one example of the application of the kNN method. In medical science, for example, kNN classifications are used for the analysis of MRI-data (magnetic resonance imaging data) and biochemical agents (Warfield, 1996; Qi, 2002). In chemical science, chemical compounds are analysed by means of the kNN method (Downs & Bernard, 2001) and in computer science, database queries, such as the search of photos, videos, and documents are realised (cf. de Vries et al., 2002, Tuncel & Rose, 2002). A rather new field of application represents artificial intelligence, where instance-based learning is put into effect with the kNN method (cf. Duch & Grudzinski, 1999). Furthermore, text classification, as it is used for spam filtering, author identification, or topic identification has to be mentioned (Nakov & Dobrikov, 2004; Kerwin, 2006).

The advantages of the kNN algorithm are its easy understandable basic concept, which has an excellent capacity to be generalized for several real problems. Available sample information are all directly taken into account for the estimations of the missing feature, and the learning process is carried out quickly. Adverse is the extremely high effort for computing and capacious disc space requirements. The resulting models are not easily interpreted by users and a strong sensitivity respective irrelevant or noise-induced input parameters and a lack of extrapolation capacity have to be mentioned (Hessenmöller & Elsenhans, 2002; Haendel, 2003).

2.2.2 Applications of the kNN Method in Forestry

Forest applications, basically estimation of specific forest stand parameters, can be found in several studies. K-nearest neighbour estimations on basal area and diameter distribution were examined by Haara et al. (1997), Maltamo & Kangas (1998), Niggemeyer (1999), and Niggemeyer & Schmitd (1999). Tommola et al. (1999) developed the kNN method as a wood procurement planning tool. A comparison of parametric methods and kNN was undertaken by Hessenmöller & Elsenhans (2002). Further examples of kNN estimations of stand parameters or alternatives for parametric growth models are described by Malinen (2003a, 2003b) or Nieschulze et al. (2005).

Sironen et al. (2001) estimated individual tree growth with the k-nearest neighbour and k-most similar neighbour methods. Other valuations on the approximation of single tree variables were realised by Holm et al. (1997) or Korhonen & Kangas (1997). A comparison of different non-parametric approaches for singles tree variables estimation can be found by Malinen et al. (2003). Among another non-parametric method, the kNN estimation is used to predict the internal quality and value of Norway spruce trees. Lemm et al. (2005) thought about basic principles and investigated the kNN method for estimating the productivity of timber harvest.

The Pteridophyte and Melastomataceae species richness were estimated by Rajaniemi et al. (2003) in the Amazonian rainforests. Alternative methods for the estimation of biomass on tree level through consideration of the kNN method were investigated by Fehrmann (2006). Beyond this, a wide range of applications can be observed in the classification of raster data of digital aerial photographs or satellite images, with respect to multi-source inventory sampling procedures. Examples are the Finnish multi-source National Forest Inventory (MS-NFI) (e.g. Kilkki & Päivinen, 1987; Moer, 1987; Tomppo, 1991; Moer & Stage, 1995; Moer & Hershey, 1999; Anttila, 2002; Temesgen, 2003; LeMay & Temesgen, 2004) or the US-American National Forest Inventory (q.v. McRoberts et. al., 2002). In this connection, the kNN method is used to enable the integration of remotely sensed data and field reference plots via the alignment of the spectral signatures of remote sensing sensors and the

terrestrial data (Holmström et al., 2001; Stümer, 2004). Furthermore, the kNN method was tested in line with the state forest inventory of North Rhine-Westphalia in 1997 (Tomppo & Pekkarinen, 1997). An application of remote sensing with kNN for communal forest inventory was described by Muinonen & Tokola (1990).

Originally, the method was developed for the derivation of maps for metrical data. Due to the fact that categorical data disallow average determination, the kNN method was extended to the application of categorical data (Köhl et al., 2000). The application of the kNN method for categorical data in the form of estimation of deadwood occurrence was studied by Stümer (2004). Point accuracy of a non-parametric method in the estimation of forest characteristics was described by Tokola et al. (1996). Instructions for kNN applications in forest resources description and estimation were given by Haapanen & Ek (2001). Further studies on adapting kNN methods for forest attribute estimation, image classification, and mapping were undertaken by Hapaanen et al. (2002) and Haapanen et al. (2004). Cabaravdic (2007) studied the efficient estimation of forest attributes with kNN.

3 The Study Area

3.1 Geography, Topography, Climate

The West African country, Ghana, adjoins the Gulf of Guinea in the south and is surrounded by the republics Ivory Coast in the west, Togo in the east and Burkina Faso in the north. The study area comprises the off-reserve forests of the forest district Goaso in Ghana, excluding all forest reserves and shelterbelts, and was pre-selected by the Tropenbos Ghana Project, into which this study is incorporated into (Tropenbos, 2007). The study area includes forest and agricultural sites of the Asunafo Administrative District of the Brong Ahafo Region of Ghana. It covers a total area of 2187.5 km^2, including 6 forest reserves with a total area of 779.4 km^2 (Tropenbos, 2007). The geographical position lies between latitudes 6° 27' North and 7° 00' North and longitudes 20° 23' West and 2° 52' West (figure 3-1).

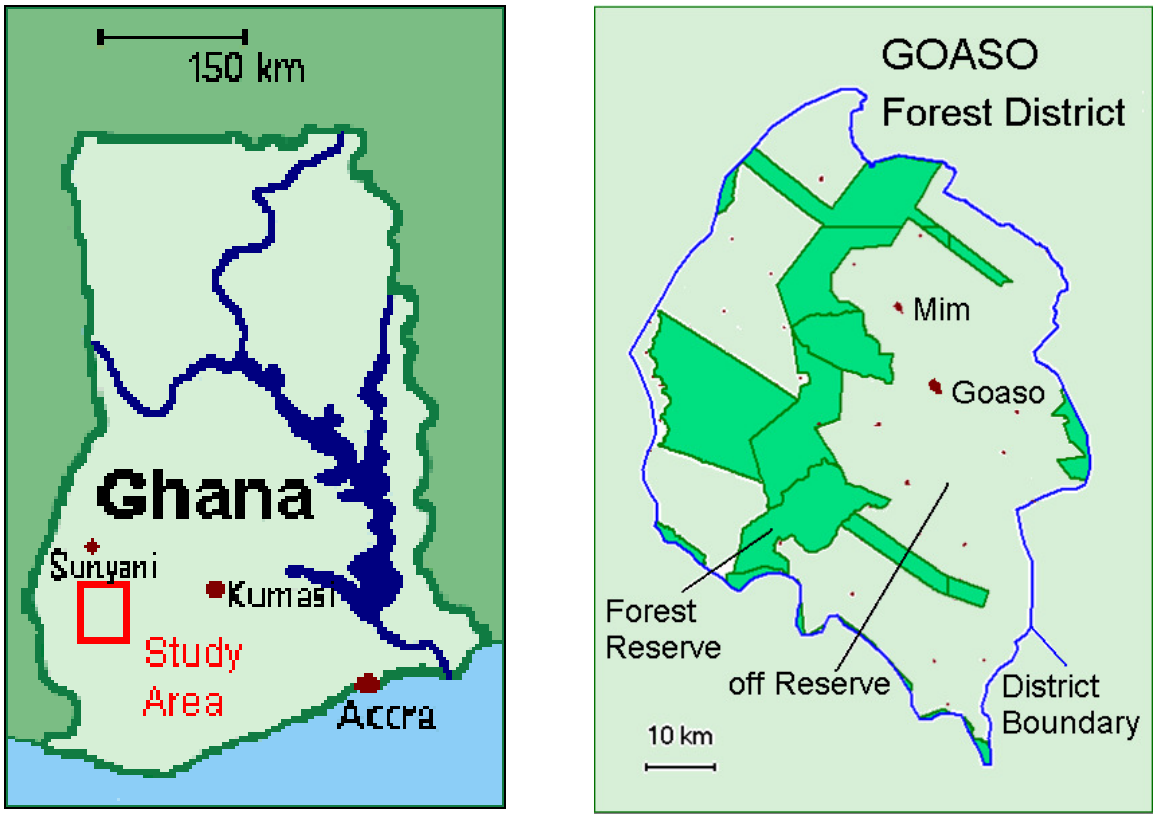


Figure 3-1 The research site are the off-reserve forests of the Goaso forest district in the southwest of Ghana.

The topography is generally low-lying, ranging from 150 m to 300 m above sea level. The research site is located in an ecological zone described as Moist Semi-deciduous Forest. The rainfall pattern is bimodal, with a major rainy season lasting from April to July and a short dry period in August. The minor rainy season begins in September and ends in October, followed by a long dry season. Temperatures normally vary between 26°C and 29°C, with a mean annual rainfall of about 1500 mm to 1750 mm (A.D.A., 2002).

3.2 Soil and Vegetation

The soils of the study area are dominated by the soil type locally classified as forest ochrosols, which are very common within the forest zone of Ghana. In the upper horizons, they are usually red and well-drained, whereas the middle horizons turn into brown, being only moderately well-drained. Food and tree crops like cocoa are supported by these slightly acid to neutral forest soils. Unlike ochrosols, which tend to be fertile and supportive for food and tree crops like cocoa, the yellowish forest oxysols are very acid, leached, and nutrient poor.

However, most nutrients of the ochrosols are concentrated within the upper 30 - 40 cm, where organic matter has built up over years through decomposition of plant biomass. Direct exposure to wind, rain, and sun leads to erosion and leaching, and, therefore, to degradation of soil fertility (Hall & Swaine, 1981; Benneh & Agyapong, 1990).

The vegetation of the research area falls within the category of tropical semi-deciduous forest with a more or less uneven tree canopy of 10 to 40 m, but also emergent trees of more than 60 m total height. In the dry season, some tree species are deciduous, though shrubs and tree species of the understorey are evergreen (Hall, 1976). Goaso Forest District is divided into off-forest reserve and, alongside, 6 permanent forest reserves and their corresponding shelter belts. The forest reserves are managed by district forest managers of the Forestry Services Division of Ghana and are officially not accessible to local people. Large proportions of the forest

reserves feature a dense continuous canopy, hosting a large number of timber tree species, such as *Ceiba pentandra*, *Milicia excels*a, *Triplochiton scleroxylon*, *Terminalia superba*, *Entandrophragma species*, and others (A.D.A, 2002). Outside the forest reserves, farming activities have formed a polymorphic mosaic of different land use types, such as cocoa farms as the major cash crop, arable land, and patches of secondary forest or fallow land, dominated by the so called “Siam Weed” *Chromolaena odorata* (Rouw, 1991).

3.3 Land Use

In general, land use “refers to management activities, conduced by man directed at a tract of land” with the intention to obtain products and/or benefits through using land resources (de Bie, 2000). On the other hand, land cover refers to the vegetation, structures, or other features that cover the land. Satellites give a physical description of the earth's surface, while most maps describe the area functionally. Both descriptions are referred to as land cover and land use, respectively, and are often confused (Barnsley et al., 2001; Fisher et al., 2005).

The major occupation in the off-forest reserves is farming, consisting of cash cropping and subsistence cropping (Dias, 2003). Besides mono cropping, monoculture and land rotation, the traditional systems of farming are still practiced (Asamoah-Boaten, 2003). They have been referred to as shifting cultivation or rotational bush fallowing system. Cultivated land is abandoned for a few years in order that exhausted soils can recover and nutrients are able to accumulate again. These land parcels are termed fallow and shape the typical patch mosaic of the off-reserve forests with its different succession stages and occurrence. The minimum forest fallow period for full recovery of soil fertility lasts 15 years; however, population pressure and resultant demand for more farming land, has reduced this period to an alarming 4-6 years (Wills, 1962). Land rotation is a modified form of shifting cultivation and means that a farmer clears and cultivates a piece of land until it becomes infertile (QWOD, 2001). Afterwards, he leaves it for another piece, only to return to the original site after some years.

Mixed cropping is usually practised. However, cash crops like cocoa and oil palm, are cultivated on a larger scale in the form of monocropping. However, they mostly start as mixed cropping, then being interplanted with staple crops like plantain cocoyam, cassava or maize, in the first few years. Food crops have shorter growth cycles, mature faster and may be harvested several times per year. With the passing of time and when replacing food crops are not substituted, additional space is provided for the remaining and long-lasting cocoa trees or oil palms, meaning that food crops will gradually disappear. This results in all imaginable proportions of intercropping between cash and food crops.

As regards observations made, several authors (Affum-Baffore, 2001; Asamoa-Boateng, 2003; Dias, 2003; Voado, 2004; Alo & Pontius, 2006; Bih, 2006) allude to different land uses in the off-reserve areas of the Goaso forest district. The following list summarises the major land uses, adopting the term land use systems (LUS):

3.3.1 Annual Cropping System

These lands are basically under the cultivation of non-woody crops, which are harvestable within a one-year period. However, crops like cassava (*Manihot esculenta*) or plantain (*Musa paradisiaca*) can remain on the land for further years, so crops are not strictly annual. Other crops comprise vegetables, maize (*Zea mays*), cocoyam (*Xanthosoma sagittifolium*), and wild yam (*Dioscorea spp.).* Mixed cropping is quite common here, but they can also be grown as mono crops and are mostly planted for subsistence use and not sold externally. As the cultivated crops do not cope well with shade, trees are undesirable and only odd tree individuals, which have a particular value for the farmers, are tolerated (Amanor, 1996).

3.3.2 Perennial Cropping System

This land cover is characterised by woody species like cocoa (*Theobroma cacao*), oil palm (*Elaeis guineensis*), and occasionally teak (*Tectona spp.*), which take several

years to reach maturity for harvesting. During seedlings stage, oil palm and, in particular, cocoa require shade from excessive sun. Fruit trees, such as *Citrus sinensis*, *Persia americana*, and *Magnifera indica,* are often found in cocoa farms, forming a minimum of two storeys: cocoa and fruit tree at the lower level and big shade trees in the upper storey. The products are the typical cash crops of the region and are cultivated on a large scale in the form of monocultural plantations.

3.3.3 Young Fallowing System

Young fallows may also be labelled shrub or bush fallows. They are the first successive stages after agricultural activities have been halted. Farmers leave their agricultural fields when soils are exhausted and they intend to come back when fertility has been regained. Their absence usually lasts for about 4 years, with a rotation cycle of two years (Abbiw et al., 2002). The abandoned land is usually taken over by grassy, herbaceous and shrubby vegetation, and, occasionally, some trees. With the presence of *Chromolaena odorata*, it is quickly covered by this weed, which dominates and suppresses the herbal layer and even climbs and overgrows shrubs and bushes.

3.3.4 Old Fallowing System

Another term for this land use system is tree fallow. It develops from former young fallows, which have not been transformed into agricultural fields, or it emerges from lands that were intensively logged. These lands are untended for 5 or more years (Abbiw et al., 2002). Often these lands are left to collect forest and wood products or may become forestland in transition. These areas consist of a dense vegetation of young regenerated trees and, perhaps, older trees, but could also refer to secondary forest of patches of forest lands.

3.3.5 Grass Fallowing System

Grass fallows normally refer to areas densely covered by tall growing grasses like elephant grass (*Pennisetum purpureum*). The vegetation reaches 3 to 4 meters and shows scarcely any trees. These lands are usually found in very low-lying areas which are seasonally flooded. The cultivation of rice has caused high levels of degradation on these sites.

3.4 Land Use Types

For this study, ten land uses were identified and described. They are basically adopted and deduced from the major land use types described before, but do also include land cover to some extent, where clear functional aspects were ambiguous. With respect to appearance and growth pattern or proportions and composition, which would affect the feasibility of a further classification with the methods of this study, a selection was made and adjusted. To distinguish the land uses defined in similar studies and described above, the term land use type will, therefore, be adopted for this study. They are:

3.4.1 Bamboo

The bamboo species *Bambusa vulgaris*, as a major non-timber forest resource, is widely spread in the study area, though it does not contribute high proportions of overall surface cover. Bamboo clumps can be found individually dispersed, consisting of less than hundred canes, covering about 100 m^2. More frequently observable are clusters of several clumps, mostly along or close to streams. Oftentimes, they aggregate to vast numbers of clusters, covering an area of more than one hectare, with individual clumps consisting of several hundred canes. Where bamboo clusters emerge, they are the dominant species, hardly hosting any sort of vegetation underneath. Nevertheless, single overshadowing trees may happen to appear. The use of bamboo is manifold and ranges from articles of daily use and in

the household, over furniture up to construction. Examples of bamboo occurrences in the study area are shown in figure 3-2.

Figure 3-2 The NTFP bamboo is basically found along small streams (left) or in areas with good water supply. Clumps may heavily be exploited (right), which causes drastic changes in reflection.

3.4.2 Banana/Plantain Plantation

Banana and plantain are of the genus *Musa*. In the tropical regions of the world, plantain (*Musa paradisiaca*) is a staple consistent crop used as a basic food, in contrast to the soft sweet banana, which grows wild and is used as fruit in the area. In this study, both expressions are used synonymously and refer to the plantain. Land use types in this category have a diverse appearance, depending on whether they are young or mature plantations. Older plantations (figure 3-3) consist of more biomass and tend to cover more portions of bare soil, while young plantations host smaller individuals, display more portions of bare soil and happen to be intercropped by annual crops. For sampling, a proportion of maximal 20 % surface covered with trees or scrubs is tolerated.

Figure 3-3 Mature banana plantations are likely to reach a surface cover of almost 100 per cent, but shade trees within these lands influence reflection (left). Young plantings of plantain are often intercropped with vegetables, exposed to seasonal weeding activities (right).

3.4.3 Bush Fallow

For this study, bush fallows are defined by a vegetation height of 3 to 15 metres total height and a minimum surface cover of 20 % of bush, shrub or tree species (figure 3-4).

Figure 3-4 Young bush fallow with considerable proportions of herbaceous plants (left) and old, dense stands (right).

3.4.4 Cocoa Plantation

Usually, cocoa plantations (*Theobroma cacao*), are intensively maintained. They have developed with a closed canopy of shading trees, dispersed tree individuals or even without any cover beyond it (figure 3-5).

Figure 3-5 Widely homogeneous cocoa plantation in the form of a monoculture (left). In contrast, cocoa trees occur planted to a considerable proportion underneath an open canopy of shade trees, in the form of a multi land use mixture, e.g. with plantains (right).

3.4.5 Elephant Grass

Elephant grass (*Pennisetum purpureum*) usually populates former rice fields when they are poor of nutrients and abandoned by farmers (figure 3-6). The areas are very low-lying and seasonally flooded. This land use type is equivalent to the formerly described land use “grass fallowing system”. Tree and shrub species are very rare in these areas, but appear to cumulate in dryer zones. The total height may reach up to 4 metres.

Figure 3-6 Abandoned rice fields on lowlands are likely to turn into densely vegetated areas of elephant grass (background). Lush growth of cocoyam (foreground) and raphia palms (centre) point to a stream.

3.4.6 Grassy Vegetation

The grassy vegetation type is defined as exhibiting a vegetation height of 0 to 3 metres total height, with the limitation that the vegetation is dominated by grasses, including pastures and grass species like *Panicum maximum*, *Cynodon nlemfuensis*, *Chloris gayana*, *Andropogon gayanus*, *Bracharia ruziziensis, Tripsacum luxum, Setaria sphacelata,* and *Cenchrus ciliaris.* Often, these areas are seasonally flooded, which has reduced the propagation of competitive herbaceous vegetation (figures 3-7 & 3-10).

Figure 3-7 Small patch of grassy vegetation surrounded by bush fallow (left). A large area covered with grassy vegetation is shown in the right photo, with off-reserve forest in the background.

3.4.7 Herbaceous vegetation

Similar to that previously mentioned, this vegetation type covers all areas with a vegetation height of 0 to 3 metres total height. In contrast to the grassy vegetation type, this land use type consists mainly of herbs and herbaceous vegetation. In this category, most annual crops and some perennials are included, i.e. cassava (*Manihot esculenta*), maize (*Zea Mays*), cocoyam (*Xanthosoma Sagittifolium*), wild yam (*Dioscorea spp.*), or vegetables (figure 3-8). Recently abandoned agricultural fields, successively covered by herbaceous vegetation, particularly areas overgrown with *Chromolaena odorata,* also fall to this category.

Figure 3-8 Herbaceous vegetation in this study includes areas planted with annual crops like maize and perennials like cassava and cocoyam (left), or abandoned cropland, successively populated by weedy herbaceous vegetation (right).

3.4.8 Oil Palm Plantation

These areas exhibit regular lines, planted with oil palm (*Elaeis guineensis*). Similar to the cash crop cocoa, these plantations are intensively maintained. However, mature oil palm plantations do not go together with any tall competitive vegetation or even trees. A photo of the oil palm compared with the raphia palm is shown in figure 3-9.

Figure 3-9 Besides cocoa, oil palms (left) are an important cash crop in the region, not to be mistaken with the raphia palm (right), which is a non-timber forest product, used for bindings or housing.

3.4.9 Raphia Palm

Normally, raphia palms (*Raphia hookeri)* occur sporadically or in loose groups along slow-moving streams or in swampy areas. Nevertheless, they can also be found in dense “raphia palm forests”, if the required wet conditions are found in large areas (figures 3-9 & 3-10).

Figure 3-10 Raphia palms (left) require wet conditions for germination and growth. In marshy site conditions, large proportions of raphia palms may occur, whereas grassy vegetation only colonises the marsh margins, with the ability to resist only seasonal flooding (right).

3.4.10 Trees/Forest

This land use type refers to all areas with a vegetation type of more than 15 metres of total height. Small forest patches or groups of trees are included in this land use type, if they cover at least 20 % of the surface, except for apparent cocoa plantations. Single tree giants, like *Ceiba pentandra (figure 3-11),* are also included in this category.

Figure 3-11 Single giant trees (left) and loose stands of forest (right) are found all over the study area. Even though standing on a farmer's property, timber belongs to the community, not the landowner.

4 Remote Sensing Data

ASTER is the sensor mounted on EOS- (Earth Observing System) AM1 (renamed TERRA), the first platform of the Earth observation programme from space (EOS programme) which the NASA promotes. TERRA is launched in a sun-synchronous orbit, and passes the equator from north to south at 10:30 in the morning, local time (ASTER GDS, 2007). The data are available by searching and ordering those data from the Earth Observing System Data Gateway (EDG).

The ASTER products cover the spectral range of 0.52 to 11.65 microns with 14 bands. 15 m, 30 m, and 90 m spatial resolutions are offered in the visible and near infrared spectral region, the shortwave infrared spectral region, and the thermal infrared spectral region, respectively. For band 3 (0.76 microns to 0.86 microns), both the usual nadir-looking telescope and a backward-looking telescope are used to produce stereoscopic images acquired in the same orbit.

4.1 ASTER Image

In this study an ungeoreferenced ASTER standard product level 1B was used (figure 4-1). The image was taken on the 26th February, 2003. Out of 14 bands of the ASTER image, bands 1, 2, 3N, and 3B with a resolution of 15 m, and bands 4, 5, 6, 7, 8, and 9 with a resolution of 30 m were taken.

4.1.1 Image Geometric Correction

For the geo-referencing of the satellite image, existing models based on precise topographic maps were not applicable and official basing points were not available. Present maps were more than 30 years old, suffering from constantly moving rivers, settlements and lines of communication. Therefore, separate basing points were collected and applied to the image geometric correction function “Rubber Sheeting Model” of the software ERDAS IMAGINE.

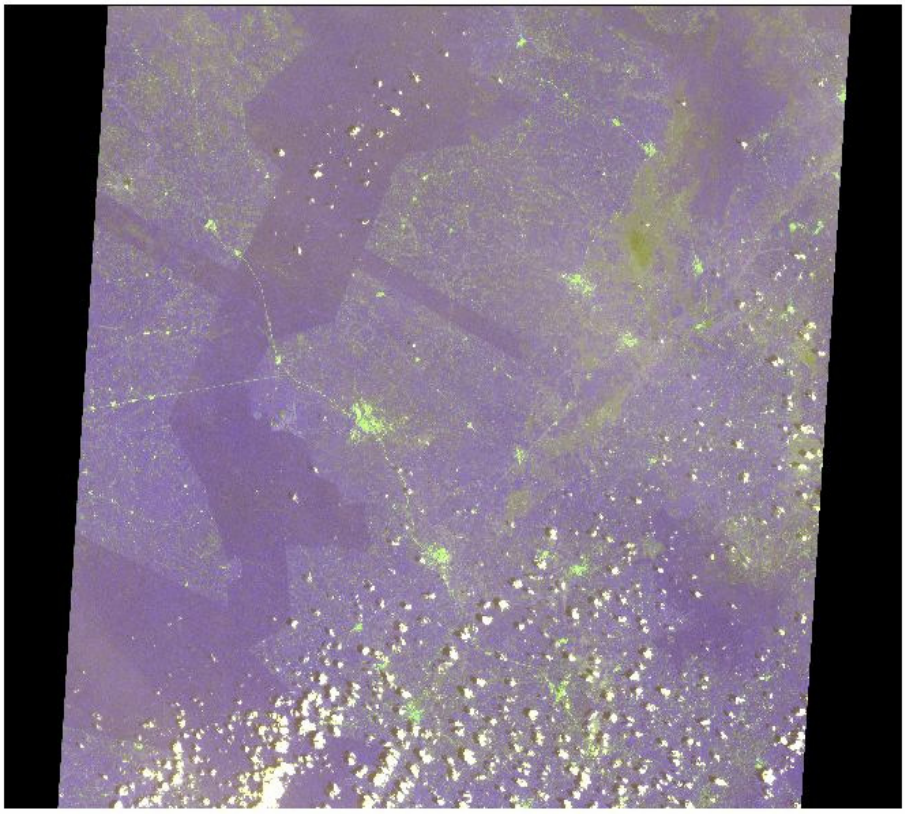


Figure 4-1 The ungeoreferenced ASTER 1B image of the study area taken on the 26th February, 2003. The well-defined dark polygon area middle left shows the protected forest reserves and shelter belts of the Goaso forest district.

4.1.1.1 Selection of Basing Points

On the satellite image of the area, small-scaled, well-defined topographic features, such as crossroads and intersections, artificial water bodies and reservoirs, isolated farmhouses and roofs of schools are excellent in serving as basing points to geo-reference the satellite image (figure 4-2). Such basing points were identified on the image and in the field, and their geographic coordinates were recorded.

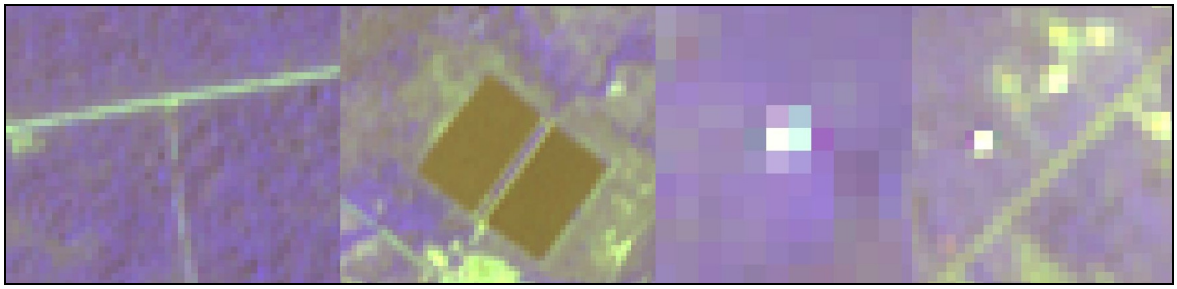


Figure 4-2 Topographic features which were used to create basing points for the geometric image correction. From the left: road intersection, artificial drinking water reservoirs, isolated farm surrounded by dense vegetation, single house with brand-new aluminium roof.

4.1.1.2 Error of the Geometric Image Correction

A total of 300 points were taken to identify and describe several locations to serve as potential basing points, or ground control points for the determination of the expected errors of the image geometric correction. This high number of points was ispositional in that various locations (e.g. intersections, houses, etc.) were defined by more than one point (figure 4-3). This enabled a clear identification of a recorded feature in the image and visualized obvious GPS errors and deviations. It turned out that the best results for the geometric correction were achieved by only using four basing points at the corners of the study area.

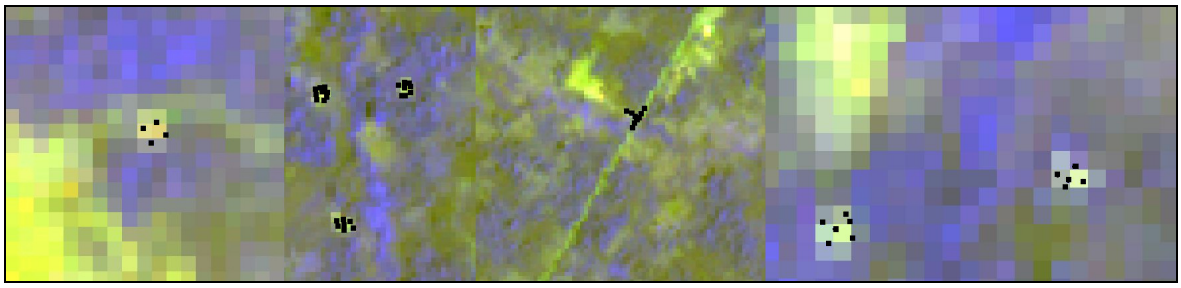


Figure 4-3 The expected error of the image geometric correction was determined via control points. These control points were overlaid with the satellite image, making deviations visible.

Within this tetragon, 33 locations remained to serve as ground control points (figure 4-4). After the image geometric correction process, the control points served to evaluate the accuracy of the geo-referenced image.

The evaluation resulted from a visible comparison of control points of a specific location (e.g. farmhouse, intersection, etc.) and their corresponding location on the satellite image. Due to the image resampling procedure “nearest neighbour”, which was chosen to avoid spatial averaging of pixel reflection values, the deviations of the corrected image can only be estimated with respect to the image resolution of 15 metres.

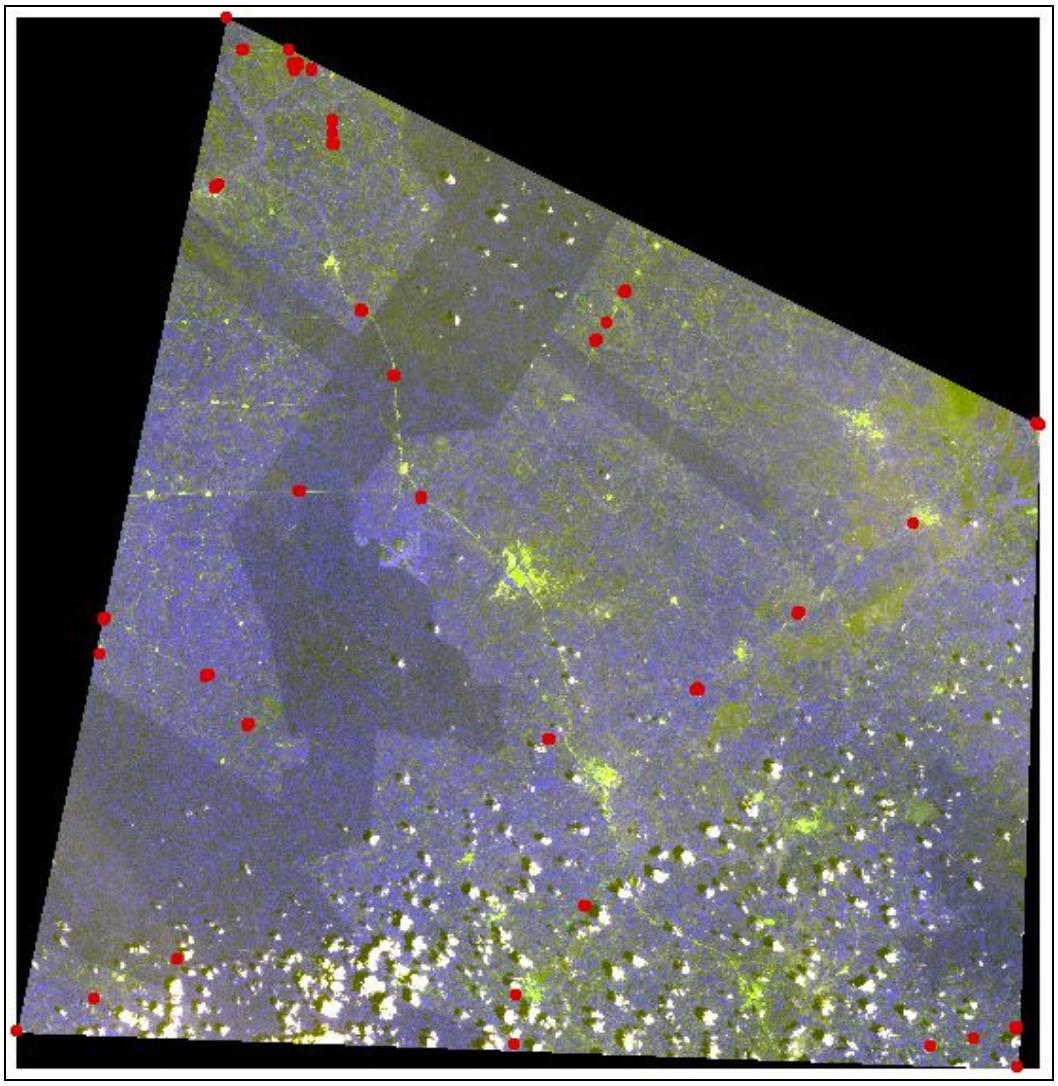


Figure 4-4 Distribution of the basing and control points in the study area of the Goaso forest district.

The four categories of deviation were defined as: ‘below 5 m’, ‘5 - 10 m’, ’10 - 15 m’, and ‘above 15 m’. Out of the 33 test points, 26 points showed a deviation below 5 m, 5 points had a deviation of 5 to 10 m, and two points were found with a deviation of 10 to 15 m (table 4-1). There were no points found with a deviation of more than 15

metres (pixel size). All points with a deviation of more than 5 metres were found very marginally on the satellite image. This leads to the conclusion that the geometrically corrected image features a deviation error of up to 5 metres in most parts and slightly higher in a minor proportion.

Table 4-1 Deviation errors of control points after the geometric image correction.

Deviation	Number	Proportions
< 5 m	26	79 %
5 - 10 m	5	15 %
10 - 15 m	2	6 %
> 15 m	0	0 %
total	33	100 %

4.1.2 Generation of Extra Bands

The ASTER image used in this study possesses the 10 original bands 1, 2, 3N, 3B, 4, 5, 6, 7, 8, and 9. Preliminary tests indicated that the classification accuracy of the kNN estimations increases with a higher number of applied bands, or, at least, by specific selection of band combinations. Stümer (2004) maintained that to acquire the highest number of correct classified pixel it was necessary to use all available bands, including different scenes, and bands which were generated by mathematical combination of the original bands (e.g. NDVI). For this study only one satellite image was available, featuring heavy cloud cover in the southern parts of the image, but at acceptable intensity in the areas of interest. Therefore, several spectral enhancement functions – offered by the software ERDAS IMAGINE – were applied and added to the original bands.

The Principal Component Analysis (PCA) is a procedure used to reduce multidimensional data sets to lower dimensions for analysis. It compresses redundant data values into fewer bands, which are often more interpretable than the source data. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system, such that the greatest variance

by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

PCA can be used for dimensionality reduction in a data set by retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" aspects of the data. But this is not necessarily the case, depending on the application. Further information on PCA can be found with Köhl & Lautner (2001) and Lautner (2001).

Indices are used to generate images by mathematical combination of the pixel values of several bands. Often they are used for vegetation analysis and to minimize shade effects of satellite images. A traditional index to characterize vegetation in remote sensing data is the Normalized Difference Vegetation Index (NDVI) (Hildebrandt 1996; Richards & Xiuping, 1999). The NDVI is a simple numerical indicator that can be used to analyze remote sensing measurements, and assess whether the target being observed contains live green vegetation or not. The NDVI uses characteristic reflection differences between visible light and near infrared (Equation 4.1).

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{4.1}$$

RED and NIR stand for the spectral reflectance measurements acquired in the red and near-infrared regions, respectively. The spectral reflectance measurements are themselves ratios of the reflected over the incoming radiation in each spectral band individually, hence they take on values between 0.0 and 1.0. By design, the NDVI itself thus varies between -1.0 and +1.0. Subsequent work has shown that the NDVI is directly related to the photosynthetic capacity and hence energy absorption of plant canopies (Sellers, 1985; Myneni et al., 1995).

Further image enhancements offered by the software ERDAS IMAGINE (version 8.5), were applied to generate further bands. In some cases, the functions were

optimized for application on Landsat images. As it was the aim to only generate additional bands by mathematical functions, they were applied and modified for the use on the present ASTER image. A complete list of all bands and their origin is shown in table 4-2.

Table 4-2 List of all used bands and their origin. The band numbers in the first column refer to the newly generated image file; the names in the fourth column refer to the original ASTER file.

Band No.	Spectral Enhancement	Band Name/Function	Prior Appl. of Haze/ Noise Red.	Name of Output file
1	original band	1	no	nac_123n3b456789_i_iz.img
2	original band	2	no	nac_123n3b456789_i_iz.img
3	original band	3n	no	nac_123n3b456789_i_iz.img
4	original band	3b	no	nac_123n3b456789_i_iz.img
5	original band	4	no	nac_123n3b456789_i_iz.img
6	original band	5	no	nac_123n3b456789_i_iz.img
7	original band	6	no	nac_123n3b456789_i_iz.img
8	original band	7	no	nac_123n3b456789_i_iz.img
9	original band	8	no	nac_123n3b456789_i_iz.img
10	original band	9	no	nac_123n3b456789_i_iz.img
11	IR/R	2/1	no	irr_2d1.img
12	IR/R	3n/2	no	irr_3d2.img
13	IR/R	3b/2	no	irr_4d2.img
14	IR/R	3b/3n	no	irr_4d3.img
15	NDVI	2-1/2+1	no	ndvi_2-1d2+1.img
16	NDVI	3n-2/3n+2	no	ndvi_3-2d3+2.img
17	NDVI	3b-2/3b+2	no	ndvi_4-2d4+2.img
18	NDVI	3b-3n/3b+3n	no	ndvi_4-3d4+3.img
19	SQRT(IR/R)	sqrt(2/1)	no	sqrt_2d1.img
20	SQRT(IR/R)	sqrt(3n/2)	no	sqrt_3d2.img
21	SQRT(IR/R)	sqrt(3b/2)	no	sqrt_4d2.img
22	SQRT(IR/R)	sqrt(3b/3n)	no	sqrt_4d3.img
23	TNDVI	sqrt(2-1/2+1)+0.5	no	tndvi_sqrt2-1d2+1+05.img
24	TNDVI	sqrt(3n-1/3n+1)+0.5	no	tndvi_sqrt3.2d3+2+05.img
25	TNDVI	sqrt(3b-1/3b+1)+0.5	no	tndvi_sqrt4-2d4+2+05.img
26	TNDVI	sqrt(3b-	no	tndvi_sqrt4-3d4+3+05.img
27	Veg.Index	2-1	no	veg_2-1.img
28	Veg.Index	3n-2	no	veg_3-2.img
29	Veg.Index	3b-2	no	veg_4-2.img
30	Veg.Index	3b-3n	no	veg_4-3.img
31	clay	5/7	no	clay_5d7.img
32	ferro	5/4	no	fer_5d4.img
33	hyd	5/7,3n/1,3b/3n	no	hyd_5d7_3d1_4d3.img
34	hyd	5/7,3n/1,3b/3n	no	hyd_5d7_3d1_4d3.img
35	hyd	5/7,3n/1,3b/3n	no	hyd_5d7_3d1_4d3.img
36	iron	3n/1	no	iro_3d1.img
37	min	3/7,5/3b,3n/1	no	min_5d7_5d4_3d1.img
38	min	3/7,5/3b,3n/2	no	min_5d7_5d4_3d1.img
39	min	3/7,5/3b,3n/3	no	min_5d7_5d4_3d1.img
40	r31	3n/1	no	r31_3d1.img
41	Principal	1	no	prn_cmp_iz_10_01.img
42	Principal	2	no	prn_cmp_iz_10_01.img
43	Principal	3n	no	prn_cmp_iz_10_01.img
44	Principal	3b	no	prn_cmp_iz_10_01.img
45	Principal	4	no	prn_cmp_iz_10_01.img
46	Principal	5	no	prn_cmp_iz_10_01.img
47	Principal	6	no	prn_cmp_iz_10_01.img

48	Principal	7	no	prn_cmp_iz_10_01.img
49	Principal	8	no	prn_cmp_iz_10_01.img
50	Principal	9	no	prn_cmp_iz_10_01.img
51	Tasseled Cap	1	no	tas_cap_123n3b_iz.img
52	Tasseled Cap	2	no	tas_cap_123n3b_iz.img
53	Tasseled Cap	3	no	tas_cap_123n3b_iz.img
54	Tasseled Cap	4	no	tas_cap_456789_iz.img
55	Tasseled Cap	5	no	tas_cap_456789_iz.img
56	Tasseled Cap	6	no	tas_cap_456789_iz.img
57	Tasseled Cap	7	no	tas_cap_456789_iz.img
58	Tasseled Cap	8	no	tas_cap_456789_iz.img
59	Tasseled Cap	9	no	tas_cap_456789_iz.img
60	Destripe	1	no	des_tm_r_iz.img
61	Destripe	2	no	des_tm_r_iz.img
62	Destripe	3n	no	des_tm_r_iz.img
63	Destripe	3b	no	des_tm_r_iz.img
64	Destripe	4	no	des_tm_r_iz.img
65	Destripe	5	no	des_tm_r_iz.img
66	Destripe	6	no	des_tm_r_iz.img
67	Destripe	7	no	des_tm_r_iz.img
68	Destripe	8	no	des_tm_r_iz.img
69	Destripe	9	no	des_tm_r_iz.img
70	Haze Reduction	1	no	haz_red_h_iz.img
71	Haze Reduction	2	no	haz_red_h_iz.img
72	Haze Reduction	3n	no	haz_red_h_iz.img
73	Haze Reduction	3b	no	haz_red_h_iz.img
74	Haze Reduction	4	no	haz_red_h_iz.img
75	Haze Reduction	5	no	haz_red_h_iz.img
76	Haze Reduction	6	no	haz_red_h_iz.img
77	Haze Reduction	7	no	haz_red_h_iz.img
78	Haze Reduction	8	no	haz_red_h_iz.img
79	Haze Reduction	9	no	haz_red_h_iz.img
80	Noise Reduction	1	no	nos_red.img
81	Noise Reduction	2	no	nos_red.img
82	Noise Reduction	3n	no	nos_red.img
83	Noise Reduction	3b	no	nos_red.img
84	Noise Reduction	4	no	nos_red.img
85	Noise Reduction	5	no	nos_red.img
86	Noise Reduction	6	no	nos_red.img
87	Noise Reduction	7	no	nos_red.img
88	Noise Reduction	8	no	nos_red.img
89	Noise Reduction	9	no	nos_red.img
90	ori geo	1	yes	hze_nse_red_123n3b456789_iz.img
91	ori geo	2	yes	hze_nse_red_123n3b456789_iz.img
92	ori geo	3n	yes	hze_nse_red_123n3b456789_iz.img
93	ori geo	3b	yes	hze_nse_red_123n3b456789_iz.img
94	ori geo	4	yes	hze_nse_red_123n3b456789_iz.img
95	ori geo	5	yes	hze_nse_red_123n3b456789_iz.img
96	ori geo	6	yes	hze_nse_red_123n3b456789_iz.img
97	ori geo	7	yes	hze_nse_red_123n3b456789_iz.img
98	ori geo	8	yes	hze_nse_red_123n3b456789_iz.img
99	ori geo	9	yes	hze_nse_red_123n3b456789_iz.img
100	IR/R	2/1	yes	atc_irr_2d1.img
101	IR/R	3n/2	yes	atc_irr_3d2.img
102	IR/R	3b/2	yes	atc_irr_4d2.img
103	IR/R	3b/3n	yes	atc_irr_4d3.img
104	NDVI	2-1/2+1	yes	atc_ndvi_2-1d2+1.img
105	NDVI	3n-2/3n+2	yes	atc_ndvi_3-2d3+2.img
106	NDVI	3b-2/3b+2	yes	atc_ndvi_4-2d4+2.img
107	NDVI	3b-3n/3b+3n	yes	atc_ndvi_4-3d4+3.img

108	SQRT(IR/R)	sqrt(2/1)	yes	atc_sqrt_2d1.img
109	SQRT(IR/R)	sqrt(3n/2)	yes	atc_sqrt_3d2.img
110	SQRT(IR/R)	sqrt(3b/2)	yes	atc_sqrt_4d2.img
111	SQRT(IR/R)	sqrt(3b/3n)	yes	atc_sqrt_4d3.img
112	TNDVI	sqrt(2-1/2+1)+0.5	yes	atc_tndvi_2-1d2+1+05.img
113	TNDVI	sqrt(3n-1/3n+1)+0.5	yes	atc_tndvi_3-2d3+2+05.img
114	TNDVI	sqrt(3b-1/3b+1)+0.5	yes	atc_tndvi_4-2d4+2+05.img
115	TNDVI	sqrt(3b-	yes	atc_tndvi_4-3d4+3+05.img
116	Veg.Index	2-1	yes	atc_vi_2-1.img
117	Veg.Index	3n-2	yes	atc_vi_3-2.img
118	Veg.Index	3b-2	yes	atc_vi_4-2.img
119	Veg.Index	3b-3n	yes	atc_vi_4-3.img
120	clay	5/7	yes	atc__cm_5d7.img
121	ferro	5/4	yes	atc__fm_5d4.img
122	hyd	5/7,3n/1,3b/3n	yes	atc__hc_5d7_3d1_4d3.img
123	hyd	5/7,3n/1,3b/3n	yes	atc__hc_5d7_3d1_4d3.img
124	hyd	5/7,3n/1,3b/3n	yes	atc__hc_5d7_3d1_4d3.img
125	iron	3n/1	yes	atc__io_3d1.img
126	min	3/7,5/3b,3n/1	yes	atc__mc_5d7_5d4_3d1.img
127	min	3/7,5/3b,3n/2	yes	atc__mc_5d7_5d4_3d1.img
128	min	3/7,5/3b,3n/3	yes	atc__mc_5d7_5d4_3d1.img
129	r31	3n/1	yes	atc__r31_3d1.img
130	Principal	1	yes	atc_prn_cmp_1-10_iz.img
131	Principal	2	yes	atc_prn_cmp_1-10_iz.img
132	Principal	3n	yes	atc_prn_cmp_1-10_iz.img
133	Principal	3b	yes	atc_prn_cmp_1-10_iz.img
134	Principal	4	yes	atc_prn_cmp_1-10_iz.img
135	Principal	5	yes	atc_prn_cmp_1-10_iz.img
136	Principal	6	yes	atc_prn_cmp_1-10_iz.img
137	Principal	7	yes	atc_prn_cmp_1-10_iz.img
138	Principal	8	yes	atc_prn_cmp_1-10_iz.img
139	Principal	9	yes	atc_prn_cmp_1-10_iz.img
140	Tasseled Cap	1	yes	atc_tas_cap_1-10_r_iz.img
141	Tasseled Cap	2	yes	atc_tas_cap_1-10_r_iz.img
142	Tasseled Cap	3	yes	atc_tas_cap_1-10_r_iz.img
143	Tasseled Cap	4	yes	atc_tas_cap_1-10_r_iz.img
144	Tasseled Cap	5	yes	atc_tas_cap_1-10_r_iz.img
145	Tasseled Cap	6	yes	atc_tas_cap_1-10_r_iz.img
146	Tasseled Cap	7	yes	atc_tas_cap_1-10_r_iz.img
147	Tasseled Cap	8	yes	atc_tas_cap_1-10_r_iz.img
148	Tasseled Cap	9	yes	atc_tas_cap_1-10_r_iz.img
149	Destripe	1	yes	dst_1-10_r_iz.img
150	Destripe	2	yes	dst_1-10_r_iz.img
151	Destripe	3n	yes	dst_1-10_r_iz.img
152	Destripe	3b	yes	dst_1-10_r_iz.img
153	Destripe	4	yes	dst_1-10_r_iz.img
154	Destripe	5	yes	dst_1-10_r_iz.img
155	Destripe	6	yes	dst_1-10_r_iz.img
156	Destripe	7	yes	dst_1-10_r_iz.img
157	Destripe	8	yes	dst_1-10_r_iz.img
158	Destripe	9	yes	dst_1-10_r_iz.img

5 Methods

5.1 Inventory Design/Data Acquisition

The collection of terrestrial data was predominantly limited by the sampling design of the study of Bih (2006), as the terrestrial data collections of the two studies were executed cooperatively. Bih was using two types of sampling designs. The first was named “adaptive cluster sampling with a systematic base”, which was carried out with random starting points along two per cent randomly selected rivers and streams of the study area. The second was called “systematic cluster sampling design” based on a systematic grid of 7 by 7 km. Plots in each category described a square of 50 metres.

For some reasons, the corners of these systematic square plots were not sufficient to function as the only coordinates for an adequate sample design for this study. On the one hand, the study area consisted of an inhomogeneous mosaic of different land uses, randomly covered by trees, and strongly varying in size and shape. This made it necessary to locate the defined land use types selectively, to enable minimum plot size for definite identification on the satellite image, to ensure adequate sample size for each land use type and to avoid negative reflection influences of shading trees or land use boundaries. On the other hand, some land use types occur rarely or solitary (e.g. raphia palm), while others (e.g. oil palm, cocoa) happen to appear excessively and frequently.

In the surroundings of the 50 m sample squares, either along rivers, or based on the 7 by 7 km grid, typical samples of the described and demanded land use types were selectively located, the geographic coordinates of the variable sample circle centres were acquired using a GPS receiver (description chapter 5.2), and further information and characteristics on the vegetation were recorded:

- main land use type
- proportions of other land use types, including shade trees
- average vegetation height
- radius of the variable circle plot (respectively crown diameter for individual giant trees)
- date of record
- extent of exploitation (only for bamboo)

Only sample plots with a circle radius of at least 15 m were selected. With the software Arc View, a buffer around each plot was generated dependent on the circle radius (figures 5-1 & 5-2). Afterwards, each encircled pixel was exported with its geographic coordinates and the code for the registered main land use type (table 5-1).

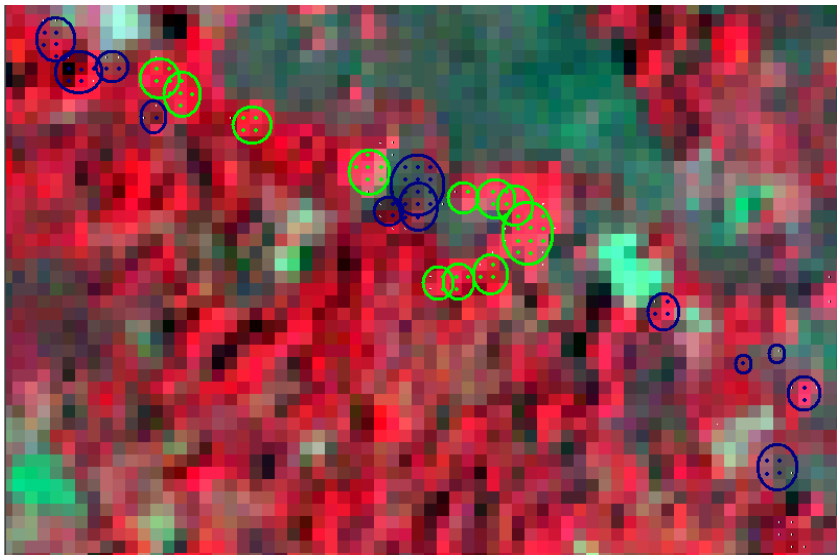


Figure 5-1 'Bamboo' and 'non-bamboo' plots are shown along a stream (invisible). The circles correspond with the recorded circular plots of the recorded land use types. Completely encircled pixels were identified, labelled, and entered the kNN data base.

Table 5-1 Cypher code of the land use types and sample size for each category.

Land Use Type	Code	Sample Size
Bamboo	1	309
Banana/Plantain	2	325
Bush Fallow	3	600
Cocoa Plantation	4	573
Elephant Grass	5	111
Grassy Vegetation	6	171
Herbaceous vegetation	7	319
Oil Palm Plantation	8	317
Raphia Palms	9	118
Trees/Forest	10	517
total		3360

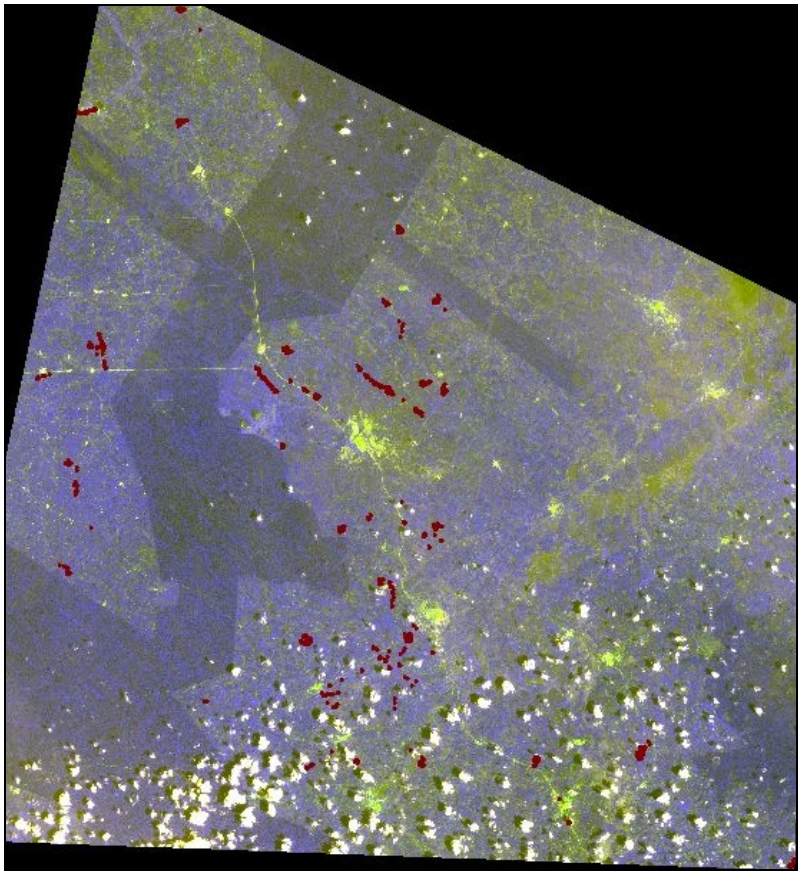


Figure 5-2 Distribution of the terrestrial sample plots in the study area.

5.2 GPS Receiver Specifications and Position Accuracy

The acquisition of geographic coordinates were conducted using the GPS receiver 'Garmin GPS II plus' (Garmin, 1997). The settings of the receiver offer a variety of position formats and map dates. For this study, the position format UTM (Universal Transverse Mercator), zone 30N, and the map datum World Geodetic System 1984 (WGS 84) were chosen.

The receiver's position accuracy is indicated to be within 15 metres (Garmin, 1997). Position accuracies can be improved with the receiver's average function, which acquires the current position every few seconds and calculates an average of the accumulated position values. For position acquisition, this function was applied for a few minutes during site assessment. It was desired that the estimated position error (EPE) of the receiver was lower than five metres before registering, though this level was not reached in dense forest patches. EPEs below five metres imply a fairly good position accuracy, whereas an independent verification of one hundred records of a specific location over several days showed, that even with the EPE-average function, the receiver's position accuracy exhibits errors of up to 9 metres (table 5-2), even though this occurred only in 1 % of the records. An actual error below five metres was found for 95 % of the records. The geographical spread is illustrated in figure 5-3.

Table 5-2 Distribution of the receiver's geographic error on the several accuracy classes.

Error [m]	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	Σ
No.	29	38	16	6	6	3	1	0	1	100

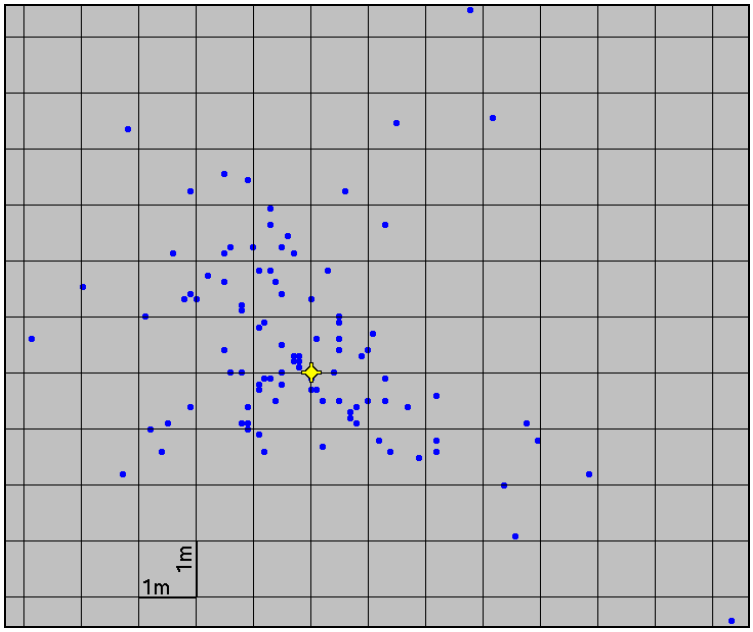


Figure 5-3 Geographical spread of the receiver´s position error of a hundred records on a specific location. The individual records are uncovered by random disarrangement of 0.1 to 0.5 metres, as the minimum unit measurement of the GPS receiver is one metre. The marked point defines the averaged position.

5.3 kNN Method

This study analyses the capacity to classify and produce distribution maps through the combination of terrestrial samples and remote sensing data with the application of the kNN method. Based on pixels, whose geographic position corresponds with sample points of the field surveys, all pixels are classified. Even pixels, which do not match with a terrestrial sample, are estimated to belong to a specific class.

5.3.1 The kNN Method for Metric Data

With the kNN method, stand variables (e.g. stand volume, stem number or biomass) result from the average value of the k neighbour samples. These neighbours are weighted by the distance value, which describes the spectral similarity (figure 5-4).

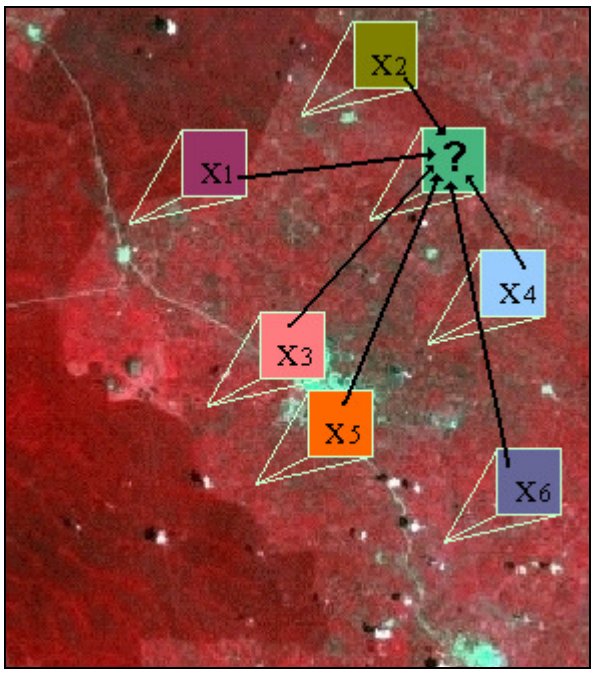


Figure 5-4 The kNN method – With the already known characteristics of pixels (x_1 to x_6), the unknown pixel is classified through an estimator.

Each pixel contains spectral information for each channel in the form of a digital value. The spectral difference between two pixels is defined with the use of a metrics. A common distance value is the Euclidean distance, $d_{(i)p}$, which has to be calculated from the target pixel p, to every sample pixel i, for which a terrestrial observation is available. If x_1 and x_2 are the characteristic vectors of two pixels whose similarity have to be tested, the Euclidean distance $d(x_1, x_2)$ between them is

$$d_{(x_1,x_2)} = \left[\sum_{j=1}^{N} \left(x_{1j} - x_{2j} \right)^2 \right]^{\frac{1}{2}} \qquad (5.1)$$

where N is the number of the spectral components (e.g. used channels).

A Euclidean distance value, where the correlative relation between characteristics remains unconsidered, is the Mahalanobis distance (Bortz, 1993).

$$d'_{(x_1,x_2)} = \left(\sum_{j=1}^{N} \sum_{k=1}^{N} c^{(x_1,x_2)} * (x_{1j} - x_{2j}) * (x_{1k} - x_{2k}) \right)^{\frac{1}{2}} \qquad (5.2)$$

with $c^{(x_1,x_2)}$ = element (x_1,x_2) from the inverses of the variance-covariance-matrix.

A generalised Euclidean distance is described as the Minkowski-r-distance (Bortz, 1993). By replacing the exponent 2 (respectively ½) with r (respectively 1/r), a generalisation of the Euclidean distance equation (5.1) is the result:

$$d''_{(x1,x2)} = \left[\sum_{j=1}^{N} (x_{1j} - x_{2j})^r \right]^{1/r} \qquad (5.3)$$

From r = 1, the City-Block-Metric results (Bortz, 1993), of which the distance of two points results from the sum of the attribute differences.

By variation of the metric coefficient, attribute differences become weighted differently. With r = 1, all attribute differences are weighted equally, irrespectively of their term. With r = 2, bigger differences get a stronger loading compared to smaller differences. With r = ∞, the biggest attribute difference is weighted 1, all others become weighted 0.

Objects exhibit a different influence on the intensity, spatial variability, and the spectral range of the reflection. The spectral range with a wide variability in the reflex reflection is often more practical for the differentiation of attribute classes of an attribute. The variability of the spectral information varies within the channels. To consider channels with a wide variability in the reflex reflection and to weight their influence on the differentiation of attribute classes of an attribute, the parameter a_j for the weighting of the channels has been introduced (Franco-Lopez et al., 2001).

$$d_{(x_1,x_2)} = \left[\sum_{j=1}^{N} a^r{}_j \left(x_{1j} - x_{2j} \right)^r \right]^{\frac{1}{r}} \qquad (5.4)$$

If the parameter a_j for j = 1,…N is chosen equal to 1, all channels have the same weight when calculating the distance. However, each channel can adequately be linked by a weighting a_j.

The k = 1 to k = 50 nearest spectral neighbours, i.e. pixel with corresponding terrestrial observations are used for the following analyses. Pixel which meet the stipulation

$$d_{(1),p} \leq d_{(2),p} \leq \ldots \leq d_{(k),p} \leq \ldots \leq d_{(n),p} \quad (5.5)$$

in the spectral feature space, where $d_{(k),p}$ is the distance of the *k-nearest* neighbours and *n* is the number of available pixels with corresponding terrestrial data. All pixels with distances in the spectral feature space greater than $d_{(k),p}$ of the observed pixel *p* are ignored. With *k* = 1, only the pixel with the lowest spectral difference is been used for the further calculations. The higher the values for *k* that are used, the more pixels with corresponding terrestrial records affect the estimation of the target value, which contains no terrestrial records. The use of *k* samples means that the random scatter, caused by signal errors, may be narrowed.

The distance values do only represent the differences between the spectral information of two pixels. To integrate the attribute values from the terrestrial observations, which are allocated to the *k-nearest* pixels, for further calculations, they have to be weighted according to their spectral distance. Hence a weighting $w_{(i),p}$ is calculated for each extracted pixel.

$$w_{(i),p} = \frac{\frac{1}{d^2_{(i),p}}}{\sum_{i=1}^{k} \frac{1}{d^2_{(i),p}}} \quad (5.6)$$

The more similar the spectral information is, the higher is the weighting and, therefore, the influence on the attribute value, which has to be calculated.

Maltamo & Kangas (1998) have modified (5.5) to determine the pixel weighting

$$w'_{(i),p} = \frac{\left(\frac{1}{1+d_{(i),p}}\right)^t}{\sum_{i=1}^{k}\left(\frac{1}{1+d_{(i),p}}\right)^t} \tag{5.7}$$

where *k* describes the number of nearest neighbours and *t* influences the weighting of the distance. The bigger *t* is chosen, the bigger is the weighting of a pixel with a narrow spectral distance. The sum of all weightings $w'_{(i),p}$ is always 1.

The attribute value of a pixel sought-after is calculated with help of the attribute values, derived from terrestrial sampling, brought into correlation with the corresponding weighted spectral data of the *k-nearest* pixels.

$$\hat{m}_p = \sum_{i=1}^{k} w_{(i),p} m_{(i),p} \tag{5.8}$$

where $m_{(i),p}$ are the terrestrially recorded values of $i = 1,\ldots k$ pixels, which are located nearest to pixel *p* in the spectral space. The process is repeated for every pixel and results in intensive computations, depending on the resolution of the sensor and the size of the inventory area (Stümer, 2004).

By variation of the variables *k*, *r*, *t*, and a_j, influence on the estimator may be exerted. Calculations with varying values for the parameters *k*, *r*, and *t* are made. For an easier comparability, the value 1 was chosen for the variables a_1, a_2, …, a_j. To determine the optimal settings for the variables, the overall accuracies of the kNN estimations were calculated and alternatives with higher accuracies were identified.

5.3.2 The kNN method for Categorical Data

The estimation procedures shown above are only defined for interval- and absolutely scaled data. Under the assumption of the permutation variance, they can not be transfused on rank- or scaled data. An approach to the solution of the problem is to forego a weighted averaging to derive the referable attribute class through the probability of the observed pixel.

Initially, as in the metric model, the distance from the *k-nearest* neighbours in the spectral feature space to pixel *p* is defined and the weightings $w_{(i),p}$ are derived (equation 5.1 to 5.6). An attribute *j* with *t* attribute classes is dedicated to each of the *k* reference pixel. For each of the *t* attribute classes of *j*, the sum of the weights $w_{(i),p}$, $w_{p,j}$, are calculated.

$$w_{(i),p,j} = \begin{cases} w_{(i),p,} & \text{if pixel } p \text{ is dedicated to attribute } j \\ 0, & \text{in all other cases} \end{cases} \qquad (5.9)$$

$$w_{p,j} = \sum_{i=1}^{k} w_{(i),p,j} \text{ mit } p = 1,...,N; \quad i = 1,...,k; \qquad (5.10)$$

From the set of the *t* weights $w_{p,j}$, the maximum is defined and the corresponding attribute *j* with the corresponding attribute class *t* is allocated to pixel *p* (Stümer, 2004).

5.3.3 Operation of the kNN Method

For the application of the kNN programme of Stümer (2004), two input files are necessary, an ‘image file’ and a ‘field sample file’. The required aerial data (‘image file’), which are necessary for the kNN calculations, are converted from the geo-referenced ASTER image into an ASCII-file. For this purpose, the utility “pixel to

ASCII" of the software ERDAS IMAGINE (ERDAS Inc.), which reads and exports the "*.img" image file, was applied. Since only small portions of the image are of interest, only pixels with corresponding terrestrial records were selected by defining an AOI-layer (Area of Interest). This fraction enabled an efficient computation, by discarding dispensable data. The resulting file comprised the geographical coordinates (x, y) and the grey scale values of all bands of each pixel.

The compilation of the "Field Sample File" results analogue to the preparation of the "Image File". Firstly, the geographical coordinates of all samples are saved in an ASCII data file. With the utility "Pixel to ASCII" of the software ERDAS IMAGINE, the grey scale values of the particular sample coordinates are exported. The last column of the ASCII data file is complemented with the cypher code of the particular land use types. For nominal scaled attributes, the kNN programme additionally outputs the occurrence probability 'p'.

The output file of the kNN programme comprises the x and y coordinates and the cypher code of the calculated land use type. The format is an ASCII data file and was easily imported into GIS software, like ERDAS IMAGINE or ARC VIEW.

Besides the visualisation of the results, the accuracy of the classification was calculated. To determine the overall accuracy of the kNN estimations and to identify the optimal input settings, band selection, and parameter adjustments and combinations, the pool of sample pixels for each land use type was divided into two collectives with the use of random numbers. The first collective served as input or training sample (ground truth) for the kNN programme, the second collective served as control or test unit (verification) and was only used to compare the congruence of the kNN estimation with the terrestrial surveyed reference points. Thus the best version with the highest accuracy was determined.

5.4 Error Analysis

For the evaluation of the reliability of the results, two types of exactness can be distinguished:

- accuracy
- precision

The results of calculations or a measurement can be accurate but not precise; precise but not accurate; neither; or both (figure 5-5). A result is called *valid* if it is both *accurate* and *precise*. The related terms in surveying are *error* (random variability in research) and *bias* (non-random or directed effects caused by a factor or factors unrelated by the independent variable).

5.4.1 Precision

Precision is also called reproducibility or repeatability, the degree to which further measurements or calculations show the same or similar results. It refers to the degree of deviation of an estimated mean, $\hat{\mu}$, from the actual population mean, μ, if the sample was repeated. In contrast to the accuracy, the selection of the sample size affects the precision (Cochran, 1977). Precision is usually characterised in terms of the standard deviation of the measurements.

In probability and statistics, the standard deviation of a probability distribution, random variable, or population is a measure of the spread of its values. It is defined as the square root of the variance. The standard deviation is the most common measure of statistical dispersion, measuring how widely spread the values in a data set is.

The variance of a random variable, probability distribution, or sample is a measure of statistical dispersion, averaging the squared distance of its possible values from the expected value. The variance is a way to capture the scale or degree of a distribution being spread out. It is the average of the squared differences between data points and the mean.

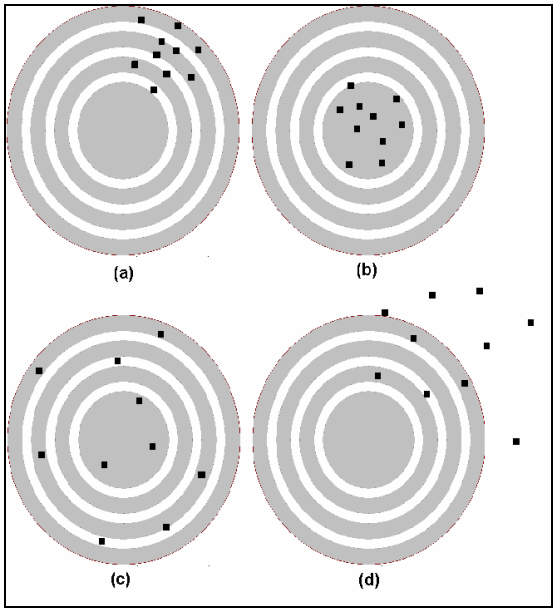
5.4.2 Accuracy Assessment

Accuracy is the degree of conformity of a measured or calculated quantity to its true value. It refers to the deviation of the actual mean, μ. The accuracy is not affected by the sample size.

Accuracy is the degree of veracity while precision is the degree of reproducibility. The analogy used here to explain the difference between accuracy and precision is the target comparison (figure 5-5). Repeated measurements are compared to arrows fired at a target. Accuracy describes the closeness of arrows to the bullseye at the target centre. Arrows that strike closer to the bullseye are considered more accurate. The closer a system's measurements to the accepted value, the more accurate the system is considered to be.

To continue the analogy, if a large number of arrows are fired, precision would be the size of the arrow cluster. (When only one arrow is fired, precision is the size of the cluster one would expect, if this were repeated many times under the same conditions.) When all arrows are grouped tightly together, the cluster is considered precise since they all struck close to the same spot, if not necessarily near the bullseye. The measurements are precise, though not necessarily accurate.

Ideally, a measurement device is both accurate and precise, with measurements all close to and tightly clustered around the known value. The accuracy and precision of a measurement process is usually established by repeatedly measuring some traceable reference standard.



a) high precision, low accuracy

b) high precision, high accuracy

c) low precision, high accuracy

d) low precision, low accuracy

Figure 5-5 Accuracy and precision. (Source: modified after Häussler et al., 2000)

5.4.2.1 Confusion Matrix

For users, it is important to know how accurately classified maps represent reality. The accuracy is calculated for the determination of the error of the kNN estimations. Nevertheless, specification of only an overall accuracy is problematic, because misclassification might have different importance for several users.

Basic prerequisite for any kind of quantitative accuracy assessment is the availability of areas with a known class affiliation of a hundred per cent. These areas are called control areas. For computer based classifications, the unit of pixels is used for ground truthing, as satellite images are already sectioned into this unit.

For this study, control pixels are extracted from the pool of sample pixels, collected during the terrestrial fieldwork. Control pixels should not be identical to training pixels, because they affect the results and give the false impression of higher accuracy estimations. Therefore, two separate collectives were defined for the kNN

estimations and the subsequent accuracy assessment. For each individual kNN estimation run, a pixel serves either as training pixel, or as control pixel. For testing the stability as regards varying sample data, reapplications of the kNN calculations with varying collectives of training (input) and control pixels were made.

A common method for the accuracy assessment is the display of the classification results in the form of a confusion matrix, also known as error matrix (Kohavi & Provost, 1998). It is a visualization tool typically used in supervised learning. A confusion matrix lists the values of known cover types of the reference data in the columns and of the classified data in the rows. The main diagonal of the matrix lists the correctly classified pixels. One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes (i.e. commonly mislabelling one as another).

A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.
The entries in the confusion matrix (see table 5-3) have the following meaning in the context of the study:

a is the number of **correct** predictions that a case is **bamboo**,
b is the number of **incorrect** predictions that an instance is **non-bamboo**,
c is the number of **incorrect** predictions that an instance is **bamboo**, and
d is the number of **correct** predictions that an instance is **non-bamboo**.

Table 5-3 Example of a confusion matrix with the classes 'bamboo' and 'non-bamboo'.

		reference data			
		bamboo	non-bamboo	$\sum_{Line}$	user accuracy
classified data	bamboo	a	b	a+b	$\frac{a}{a+b}$
	non-bamboo	c	d	c+d	$\frac{d}{c+d}$
	$\sum_{Column}$	a+c	b+d	a+b+c+d	
	producer accuracy	$\frac{a}{a+c}$	$\frac{b}{b+d}$	overall acc.	$\frac{a+d}{a+b+c+d}$

The attained accuracy of a classification is characterized by the confusion matrix. Different measures and statistics can be derived from the values in this error matrix. where the following accuracy aspects and standard terms can be calculated:

The overall accuracy (OA) is the proportion of the total number of predictions that were correct and is calculated by dividing the correctly classified pixels (sum of the values in the main diagonal) by the total number of pixels checked.

Besides the overall accuracy, classification accuracy of individual classes is calculated in a similar manner. Two approaches are possible (Congalton, 1991):

- user's accuracy
- producer's accuracy

The producer's accuracy is calculated by dividing the number of correct pixels in one class, divided by the total number of pixels as derived from the reference data in this class. The producer's accuracy measures how well a certain area has been

classified. It includes the error of omission, which refers to the proportion of observed features on the ground that is not classified in the map. The more errors of omission exist, the lower the producer's accuracy (Banko, 1998).

$$\text{producer's accuracy [\%]} = 100\ \% - \text{error of omission [\%]} \quad (5.11)$$

The user's accuracy is defined by the measure of the correct classified pixels in a class divided by the total number of pixels that were classified in that class. The user's accuracy is, therefore, a measure of the reliability of the map. It informs the user how well the map represents what is really on the ground. It includes the error of commission, which refers to the proportion of the predicted features of the classification on the map that are not observed on the ground. The more errors of commission exist, the lower the user's accuracy.

$$\text{user's accuracy [\%]} = 100\ \% - \text{error of commission [\%]} \quad (5.12)$$

5.4.2.2 Kappa Coefficient

The kappa coefficient (K) is a measure of overall agreement of a matrix. In contrast to the overall accuracy, the kappa coefficient takes non-diagonal elements into account (Rosenfield & Fitzpatrick-Lins, 1986).

Cohen's kappa coefficient is a statistical measure of inter-rater reliability (Cohen, 1960). It measures the proportion of agreement after chance agreements have been removed from considerations. It is generally thought to be a more reliable measure than simple per cent agreement calculation, since K takes into account the agreement occurring by chance. Cohen's kappa measures the agreement between two raters that each classify *N* items into *C* mutually exclusive categories. A kappa of zero occurs when the agreement between classified data and verification data equals chance agreement (Fenstermaker, 1991). In 1983, the kappa coefficient was

introduced to remote sensing (Congalton & Mea, 1983; Congalton et al., 1983). The equation for Κ is given (Bishop et al., 1975):

$$\hat{K} = \frac{N\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r} X_{i+} X_{+i}}{N^2 - \sum_{i=1}^{r} X_{i+} X_{+i}} \tag{5.13}$$

where

r	=	number of rows and columns in error matrix,
N	=	total number of observations,
X_{ii}	=	observation in row i and column i,
X_{i+}	=	marginal total of row i, and
X_{+i}	=	marginal total of column i.

To interpret the formula of the kappa coefficient, the following formula is given:

$$\hat{K} = \frac{p_o - p_e}{1 - p_e} \tag{5.14}$$

where

p_o = accuracy of observed agreement, $\frac{\sum X_{ii}}{N}$,

p_e = estimate of chance agreement, $\frac{\sum X_{i+} X_{+i}}{N^2}$.

A table for interpreting Κ values is given by Landis & Koch (1977). Although it is based on personal opinion and by no means universally accepted, it is presented here (table 5-4), as many studies refer to it (Altmann, 1991; Kulbach, 1997; Ortiz et al., 1997; Komagata, 2002; Oehmichen, 2007).

Table 5-4 Table for the interpretation of kappa values after Landis & Koch (1977).

Kappa value [%]	*Interpretation*
< 0	*poor agreement*
0 to 20	*slight agreement*
20 to 40	*fair agreement*
40 to 60	*moderate agreement*
60 to 80	*substantial agreement*
80 to 100	*almost perfect agreement*

5.4.3 Bias

In estimation, the bias refers to the value of a parameter of a probability distribution, the difference between the expected value of the estimator and the true value of the parameter. Bias is a term which refers to how far the average statistic lies from the parameter it is estimating, that is, the error which arises when estimating a quantity. Errors from chance will cancel each other out in the long run, those from bias will not. Bias occurs if those estimates for the statistic are systematically lower or systematically higher than the parameter value.

A biased sample is a statistical sample of a population where some members of the population are less likely to be included than others. Bias, B, is directly associated with the accuracy, since

$$B=\hat{\mu} - \mu \qquad (5.15)$$

6 Analyses and Results

6.1 Sample Size of the Terrestrial Plots

As is mentioned in chapter 5.1, a total of 3360 sample pixels distributed over ten different land use types encroached upon the kNN estimations. For the accuracy analyses, the samples are divided into two collectives. The training pixels are used as input data (ground truth) for the kNN programme, whereas the accuracy of the kNN estimations is assessed with the control pixels and do have no influence on the current kNN estimation. Conspicuous is the varying number of sample pixels amongst the different land use types (table 6-1). The resources bamboo (309), oil palm (317) herbaceous vegetation (319), and bananas (325) occupy a medium position, concerning the number of sample pixels, whereas fallow lands (600), cocoa plantations (573), and trees/forest (517) exhibit almost twice as many sample pixels. A comparatively low sample size is indicated for grassy vegetation (171), raphia palms (118), and elephant grass (111).

Table 6-1 For the verification of the particular kNN estimations, the terrestrial samples are divided into two collectives, the input or training pixels and the test or control pixels. The two collectives comprise 'existent' and 'non-existent' pixels.

Land Use Type	No. Training Pixels	No. Control Pixels
Bamboo	209 + 209 (418)	100
Banana/Plantain	225 + 225 (450)	100
Bush Fallow	500 + 499 (999)	100
Cocoa Plantation	473 + 473 (946)	100
Elephant Grass	86 + 86 (172)	50
Grassy Vegetation	146 + 146 (292)	50
Herbaceous vegetation	219 + 219 (438)	100
Oil Palm Plantation	217 + 217 (434)	100
Raphia Palm	93 + 93 (186)	50
Trees/Forest	417 + 417 (834)	100

The number of pixels found for each category does not generally represent the allocation of the land uses, as the sample sites were recorded subjectively. It was the objective to sample an equal number of plots for each kind of resource. It turned out that bamboo clusters and groups of raphia palms were only found infrequently and almost only along streams. This fact led to the situation that every adequate bamboo and raphia source which emerged during fieldwork was sampled. Hence bamboo and raphia defined the minimum sample size for all resources. In this connection, it should be mentioned that the original plan was also to sample rattan – as an important non-timber forest product. Due to only marginal appearances during the whole fieldwork, and consequently inadequate sample quantities and qualities, the survey of this resource had to be neglected.

The low number of samples of elephant grass is due to its growth and characteristic occurrence. In the study area, elephant grass was only found extensively scaled in the wet lowland areas, covering almost 100 % of this habitat's surface. So the variability within this land use type, respectively, mixture with other resources, was very low.

The relatively high number of sample pixels for bush fallows, cocoa plantations, and trees/forest is due to the fact that these land resources are supposed to cover the highest proportions of the study area and thus are most likely to be sampled by a random sampling technique. Besides that, the number of approximately 500 sample pixels for an estimation run serves as the optimal number for the kNN field input. This is because the used kNN programme (Stümer, 2004) does only handle 999 input pixels for the calculations, which enables the kNN simulation of one attribute with two attribute classes, e.g. 499 pixels 'bush fallow' and 499 pixels 'non-bush fallow'.

However, the sample sizes are rated to be sufficient to cover the variability of the existing characteristics of each land use type, as well as balancing the disadvantages of selective sampling through a random sampling technique.

6.2 Analyses of Optimization Options

The generation of maps with help of the kNN method requires information concerning the minimum requirements of the output results. The requirements depend on the particular problem. However, overall accuracy plays a decisive role for the usability for specific applications.

To calculate an overall accuracy, a specific number of control pixels are randomly selected out of the pool of the registered sample pixels. These control pixels are not used for the kNN estimation of this particular run. They are compared with the kNN output pixels and give the degree of corresponding pixels, the overall accuracy.

The optimisation of several parameters to increase the overall accuracy of the kNN estimations is demonstrated in the subsequent observations.

6.2.1 Band Number

For the response to the question how the number of bands affect the accuracy of the kNN method, the overall accuracy versus the number of applied bands is illustrated in figure 6-1 for bamboo. For the calculations, the standard settings of the kNN programme were used (k = 5, r = 2, t = 2) and the bands were added in the order as they were listed in chapter 4.1.2 (table 4-2). Two test runs with randomly chosen pixels were shown. The strongest increase and the highest overall accuracy with 81,0 % was obtained with 8 and 10 bands, respectively. Additional bands lead to a dramatic decrease up to the use of 42 bands, whence the curve climbed again, but did not reach the previous peak. The two curves partly take course parallel as well as in the opposite direction. The highest divergence was found with 7 %, which leads to the presumption, that pixel selection for specific band combinations has a decisive influence on the expected results.

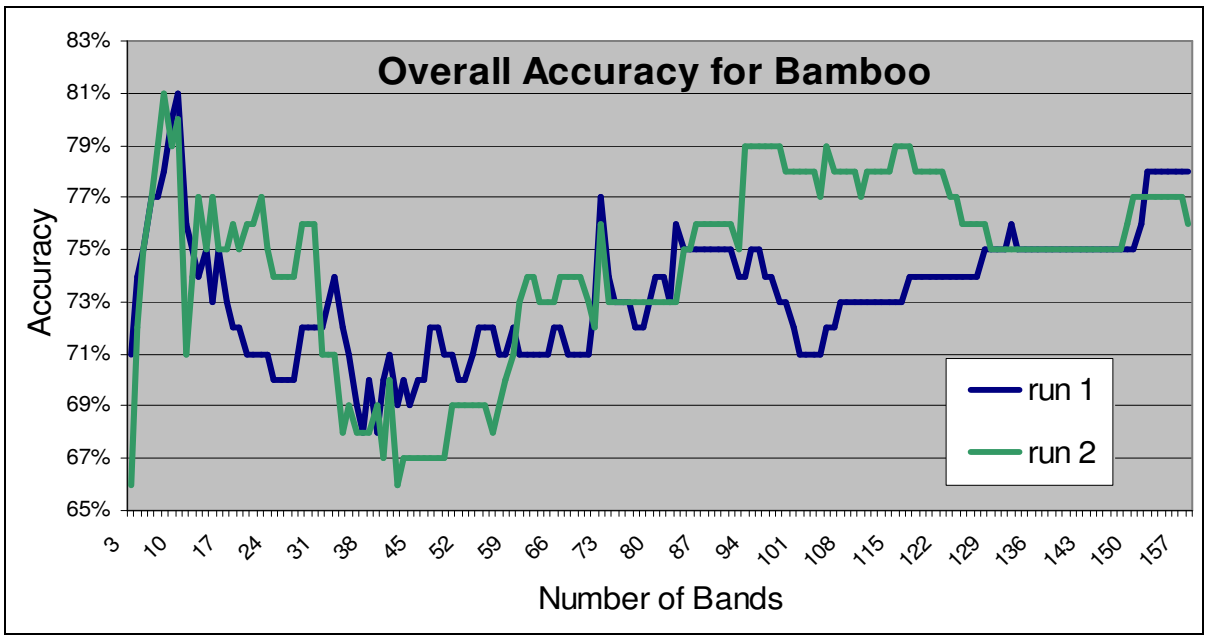


Figure 6-1 Overall accuracy vs. number of bands for two test runs with different selections of input pixels for bamboo (k = 5, r = 2, t = 2).

To minimize the stochastic influence of a low number of reruns, an average of the overall accuracy of the ten land use types with four test runs for each type is shown in figure 6-2. With respect to calculating time and effort, the number of bands accumulates in steps of ten additional bands at once, starting from the use of ten bands. Similar to bamboo alone, the overall accuracy starts from 66.3 % for the use of bands 1-3 and climaxes at the use of bands 1-10 with an accuracy of 74.5 %. Subsequently, the accuracy drops until the use of band class 1-40, from this point onwards, it climbs again, to an accuracy of 73.5 % at the use of bands 1-158 applied all together.

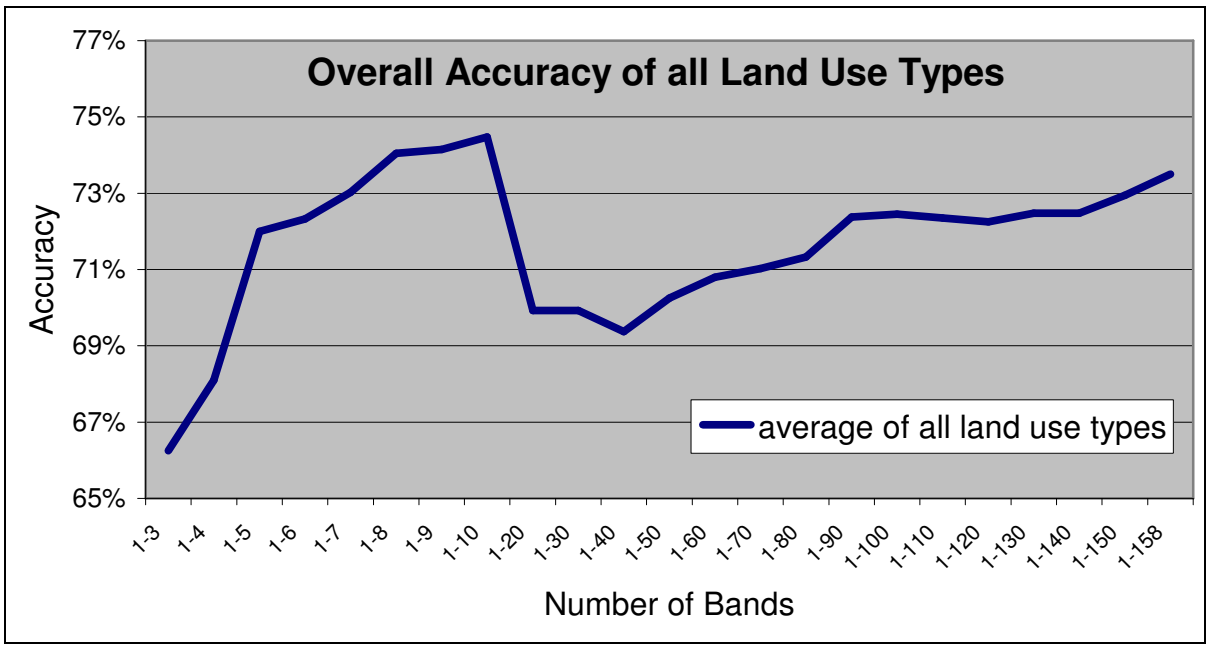


Figure 6-2 Averaged overall accuracy vs. number of bands of the ten land use types. Basis are four test runs for each different selections of input pixels per land use type (k = 5, r = 2, t = 2).

The slope of the curve, with a quick rise until the use of ten bands, which are the original bands after the geo-referencing process, and a peak at this band combination, indicates an already satisfactorily band combination for the kNN calculations. However, the collapse with the use of about thirty to forty bands and the steady rise in accuracy with increasing band numbers afterwards, suggests a negative influence on accuracy for application of specific bands or band combinations, while others are more likely to help increase accuracy.

6.2.2 Band Combination

For the accuracy assessment, varying channel combinations are considered. Particulars concerning the generation of extra bands are described in chapter 4.1.2. For the optimization of the best band combinations, bands and band combinations which do already point at a low accuracy or a negative trend, were avoided. Predominantly, these were the bands number 11 to 40. The 158 bands which are used in this study derive from the original ASTER channels 1 to 9 (including 3B). Except for these ten channels, band number 11 to 158 result from spectral

enhancements (e.g. atmospheric correction, vegetation indices) and form groups of a specific enhancement procedure in each case. Figure 6-3 shows the overall accuracy for bamboo, allotted to specific groups of band combinations. In similar manner to band group 1-10, an accuracy value of 81 % was found in band group 130-139 (principal components). An even higher overall accuracy with 83 % was found for bands 80-89, which is the group of channels that did undergo the noise reduction enhancement.

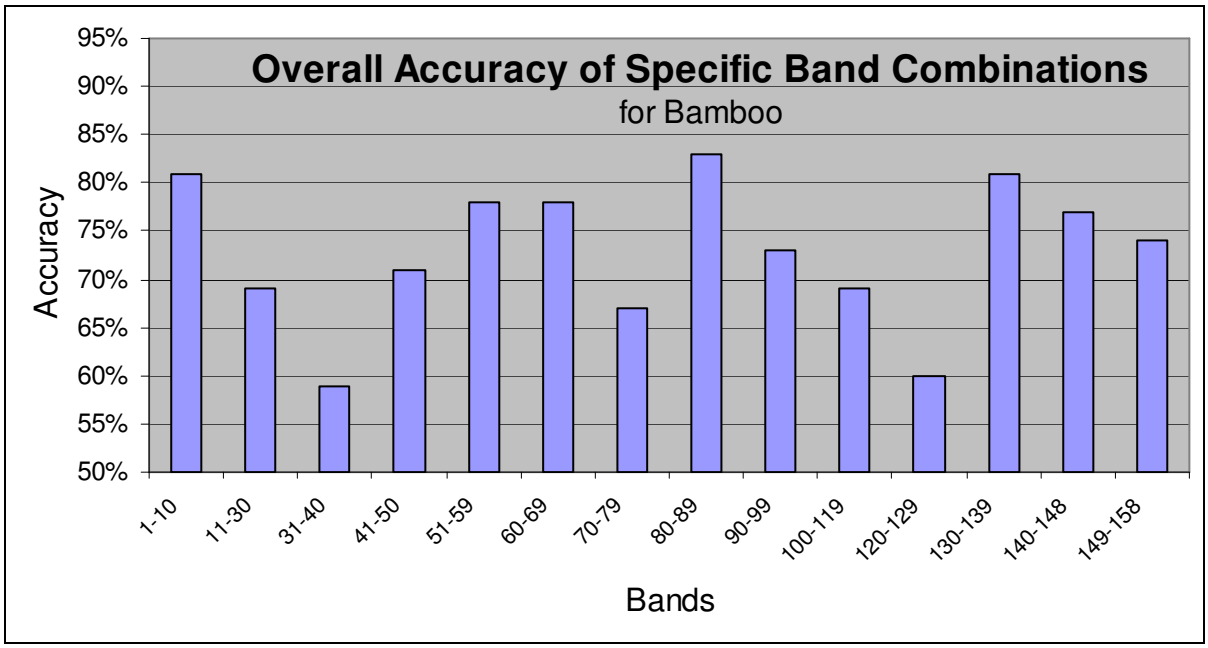


Figure 6-3 Overall accuracy of bamboo calculated for specific groups of band combinations (k = 5, r = 2, t = 2).

The overall accuracy for the average of four runs of all land use types is shown in figure 6-3. The band group combination 1-10 & 80-89 & 130-139 and the combination 80-84 & 117 show the highest accuracy with 79.3 %, whereas the combination 1-10 & 80-89 has an accuracy of 78.7 %, followed by 78.6 % each, by bands 80-84, and band group combination 1-10 & 51-59 & 80-89 & 130-139. In comparison with bands 1-10, which exhibited an accuracy of 74.5 % for all land use types, the accuracy could be increased with application of specific band combinations. These figures demonstrate that higher accuracies do not necessarily require a maximum number of bands. The selection and combination of significant bands plays a major role.

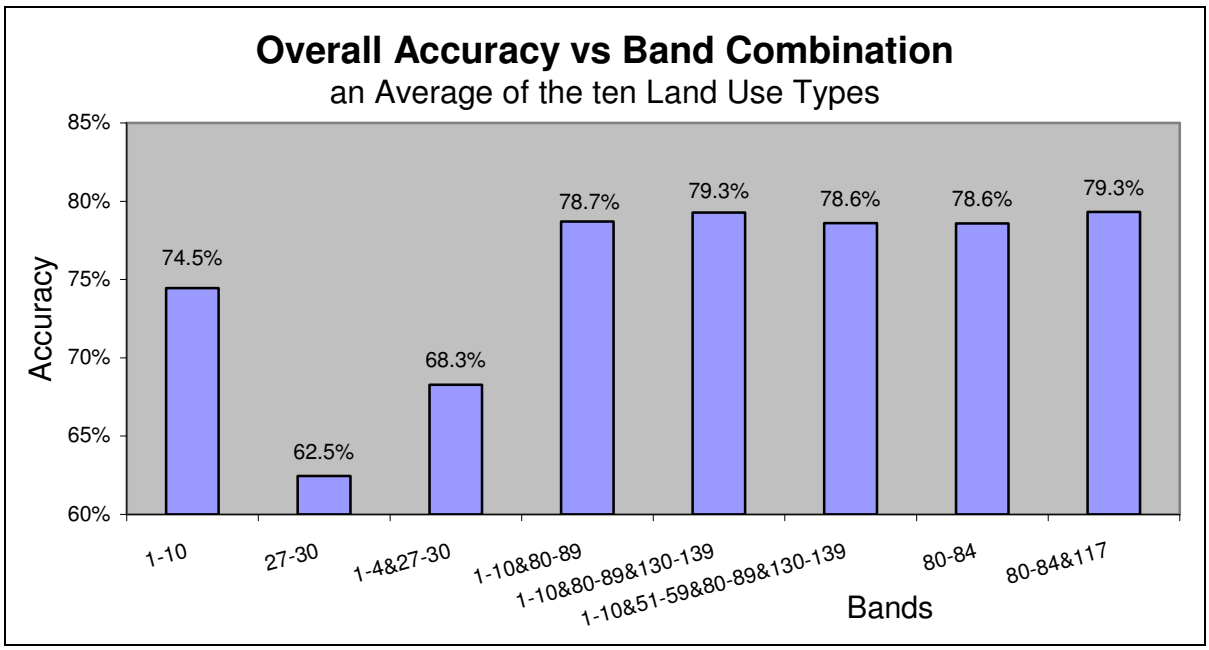


Figure 6-4 Averaged overall accuracy of the ten land use types, based on four test runs for each different selection of input pixels per land use type calculated for specific groups of band combinations (k = 5, r = 2, t = 2).

6.2.3 Land Use Type vs. Band Combination

Compared to the average of all land use types, the results for bamboo alone indicate lower accuracies. A closer look at the accuracy performance of the single land use types for selected band combinations in relation to each other is presented in figure 6-5. The accuracy differences amongst the land use systems remain largely over the different band combination. The accuracy of the oil palm plantations exhibit the highest accuracies of up to 87.5 % for band combination 80-84 and an average of 79.3 % for all band combinations, while the bush fallows (average: 67.0 %), banana plantations (average, 67.7 %), and grassy areas (average, 67.8 %) define the lower end. On the average, the accuracy differences amongst the land use systems for a specific band group are 14.7 %, varying from 9.5 % to 23.5 %. A visual appraisal of the curves indicates a parallel related curve progression for both, “poor” and “good” band combinations, hence an optimal band combination is valid for all land use classes. Simply band group 80-84, as one of the “good” band combinations, is evidently out of line. The reason for this could be due to the low number of only four bands, which makes it more vulnerable to the randomly selected input pixels.

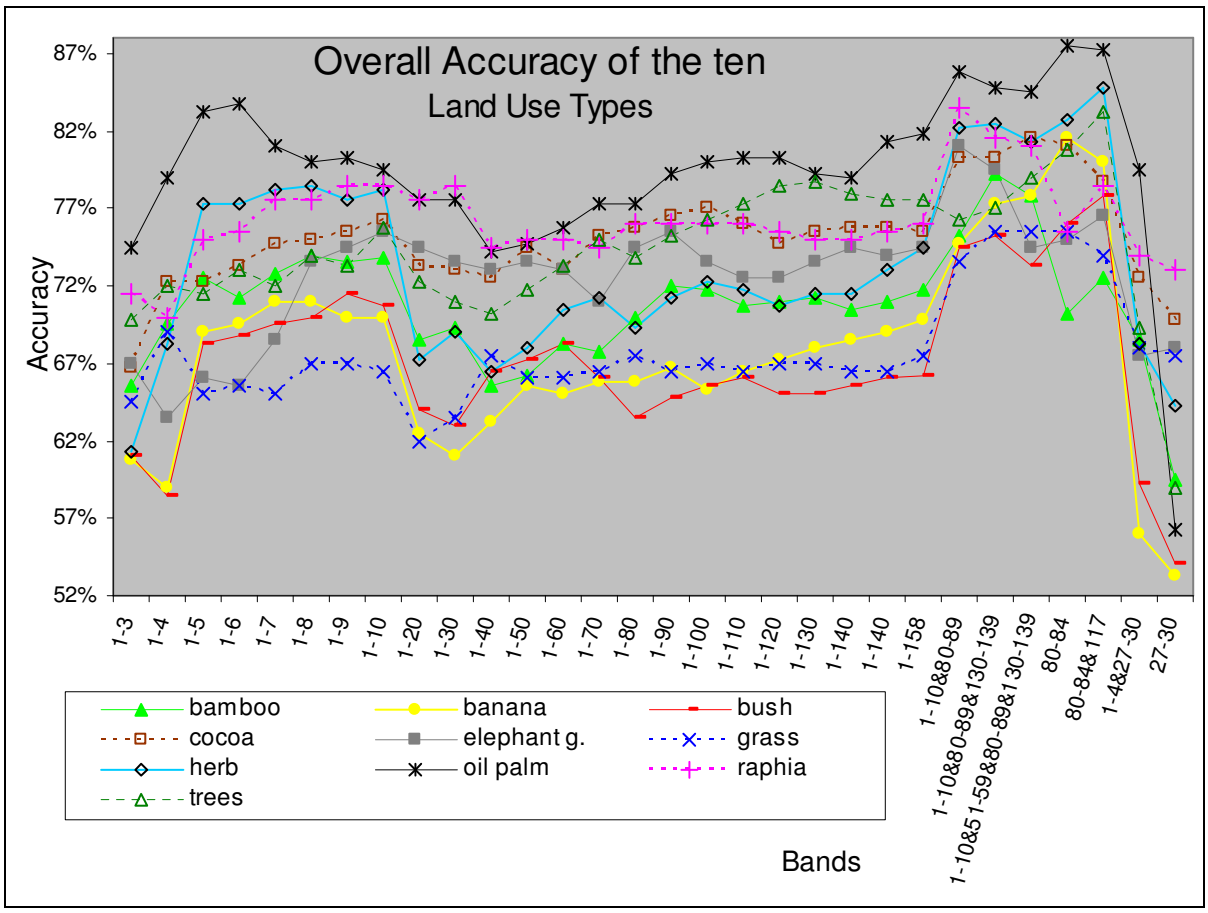


Figure 6-5 Overall accuracy of the ten land use types based on four test runs for each different selection of input pixels per land use type, calculated for specific groups of band combinations (k = 5, r = 2, t = 2).

6.2.4 Precision of the Classification Accuracies

As the results of the kNN calculations depend highly on the coincidence, which of the pixels becomes an input pixel, or, respectively, a control pixel, the progression of the standard deviation for 19 reruns (=20 runs) for the band combinations 1-10 & 80-89 & 130-139 and 80-84 & 117 is shown in figure 6-6 for bamboo. With the number of 8 reruns, the curves of the standard deviation of the two band combinations coincide, with a standard deviation of 8.8 % and show a standard deviation of 6.2 % and 6.3 % for 19 reruns, respectively. The averaging curves of the overall accuracy for both band combinations shows 3.3 % better values for the band combination 1-10 & 80-89 & 130-139, with a highest accuracy of 82.7 % and an accuracy of 79.5 % for band combination 80-84 & 117 for an average of 20 sample runs.

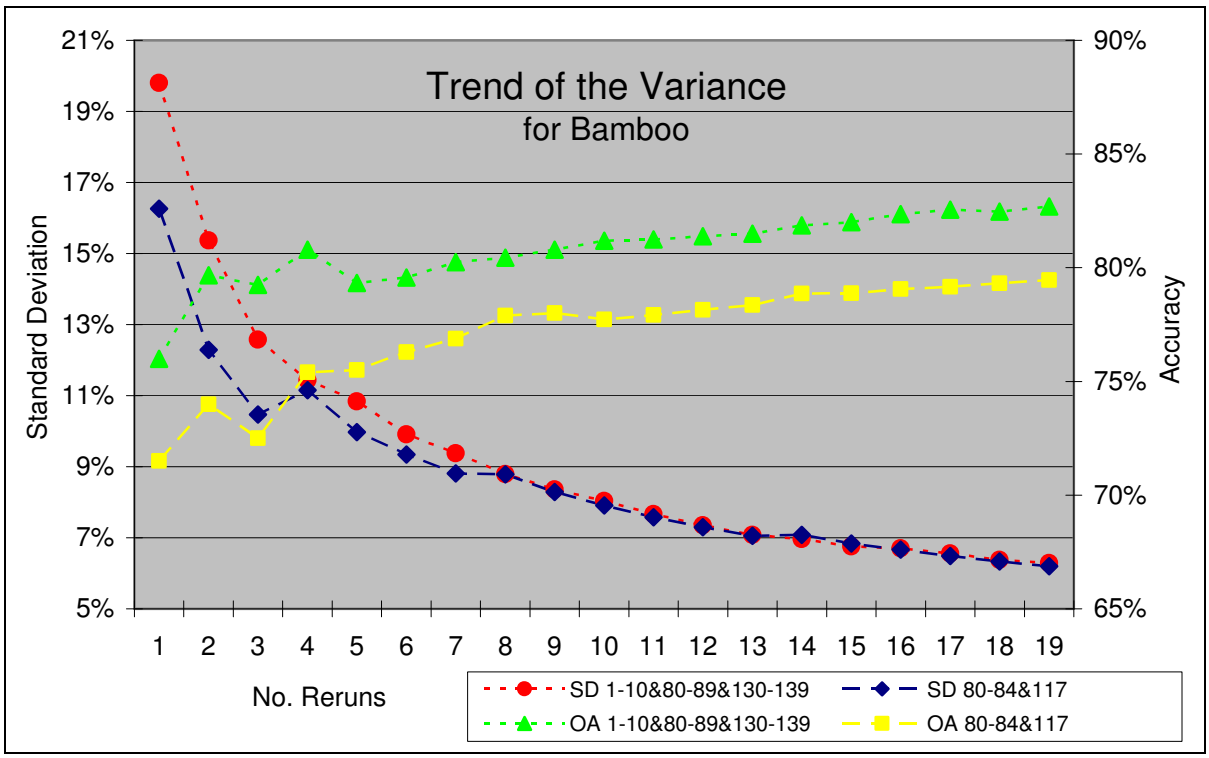


Figure 6-6 Trend of the standard deviation for bamboo. The averaged overall accuracy (OA) and the standard deviation (SD) is shown for up to 19 averaged reruns (20 runs) for the two band groups 1-10 & 80-89 & 130-139 and 80-84 & 117 (k = 5, r = 2, t = 2).

A comparison of the variance amongst the land use types, bamboo, banana, and oil palm is shown in figure 6-7. These three land use types were selected as examples and, for a small number of sample runs (<4), show that the standard deviation differs amongst the land use types. With four reruns applied, the curves of banana and oil palm – as the positive and negative extremes as regards accuracy – converge, respectively, with a standard deviation of 6.2 % for oil palm, and 5.4 % for banana, and coincide after eleven sample runs and a standard deviation of 5 %, which, additionally also defines almost the average for all three land use types at 20 applied sample runs, varying from 4.1 % (oil palm), 5.2 % (banana), and 6.3 % (bamboo).

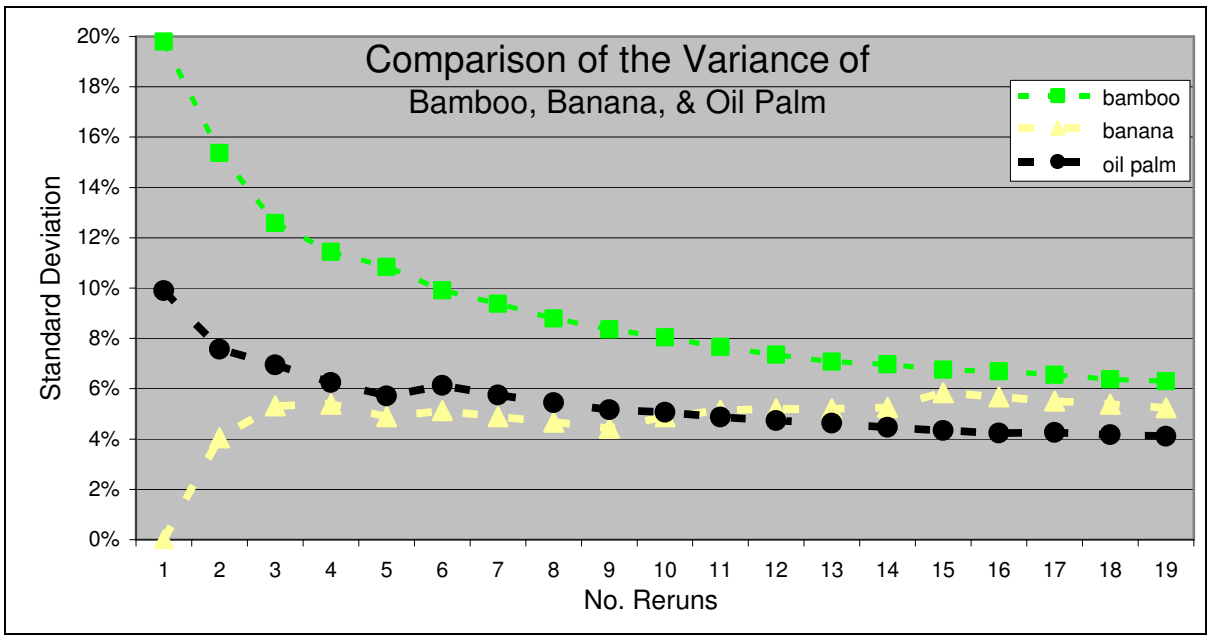


Figure 6-7 Progression of the standard deviation of bamboo, banana, and oil palm for averaged runs. The standard deviation is shown for up to 19 averaged reruns for the band group 1-10 & 80-89 & 130-139 (k = 5, r = 2, t = 2).

The look at the averaged accuracy for these land use types, versus the number of sample runs, exhibits no unique optimum for the number of sample runs (see figure 6-8). The averaged accuracy for oil palm levels off at a value of 83 % and six reruns, whereas the banana levels off at about 75 % with the same number of reruns, but bamboo requires more than 15 runs to reach a level of almost 83 %.

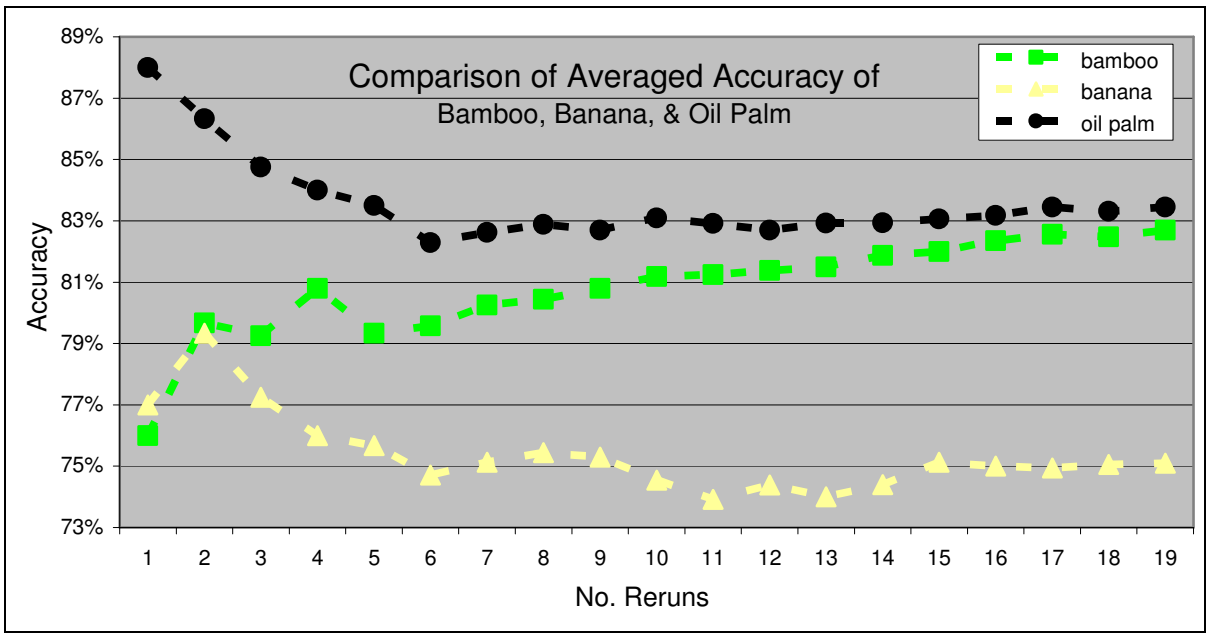


Figure 6-8 Progression of the variance of the averaged overall accuracy of bamboo, banana, and oil palm (k = 5, r = 2, t = 2; band combination 1-10 & 80-89 & 130-139).

The statistical straggling of the overall accuracy for ten sample runs of each land use type is presented in figure 6-9. The highest standard deviation of the overall accuracy was found for the grassy vegetation with a value of 15.3 %, more than twice the average of 6.7 % for all land use types. This exception is due to the extreme values of the minimum (40 %) and maximum (96 %) accuracies. The cause might be the relatively low number of 292 input pixels (see table 6-1), although raphia palms with 186, and elephant grass with 172 input pixels have an even lower number of input samples. However, the overall accuracy for grassy vegetation lies with 81.2 % above the average of 79.2 %, and varies between an accuracy of 82.7 % for oil palm plantations to 75.3 % for banana plantations.

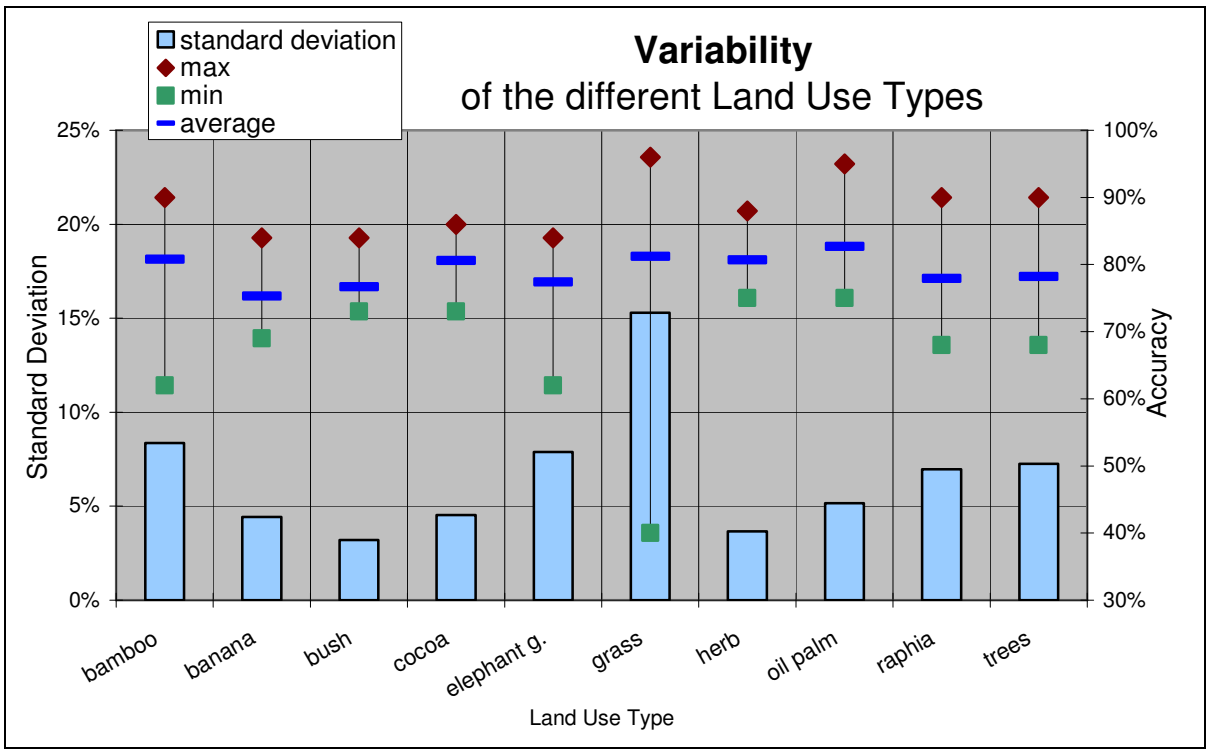


Figure 6-9 Comparison of the variability of the accuracy values within the different land use types. The standard deviation, averaged overall accuracy, minimum, and maximum values, are calculated on a basis of 10 sample runs for each land use type (k = 5, r = 2, t = 2; band combination 1-10 & 80-89 & 130-139).

The lowest standard deviations were found for bush fallow (3.2 %) and herbaceous vegetation (3.7 %), in which connection the former achieves 79.7 % overall accuracy, and the latter 80.7 %. A trend that the expected accuracy corresponds clearly with the number of input pixels was not found. Merely a tendency of lower numbers of input and control pixels, respectively, leads to higher standard deviations.

6.2.5 Distribution of Sample Plots

For the assessment of the influence of the terrestrial sample plot distribution on the accuracy of the kNN estimations, four categories of sample distributions, respectively, distribution of training pixel versus control pixel are defined: separately far-, separately close-, randomly-, and equally distributed. These four categories simulate a gradient of a very low to very high sample intensity, as the geographical distance between training and control group declines from the separately far

distribution, over a randomly to an equally distributed allocation (see figure 6-10). The chance of a training pixel located directly adjacent to a control pixel and, therefore, showing similar reflection values, is highest in the category of an equal distribution of training and control pixels, and most unlikely in the group of the separately far distribution.

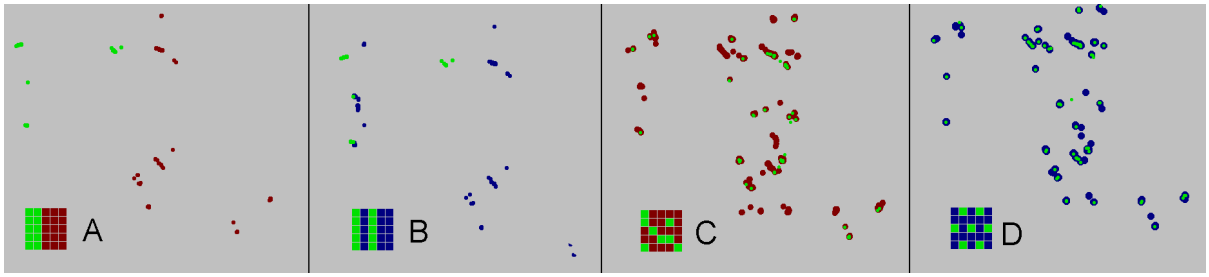


Figure 6-10 Allocation of training (dark) pixels and control (light) pixels to simulate a gradient of very low (A)to a very high (D) sample intensity: A-separate far, B-separate close, C-random, and D-equal distribution.

The averaged overall accuracy for bamboo steadily rises from 67 % to 85 % from the separate far to the equal distribution of training and control pixel (figure 6-11).
Furthermore, the least difference of the minimum and maximum value for the four runs was found in the equally distributed collective. The highest standard deviation of 12.2 % was found for the randomly distributed sample collectives, more than three times higher than the equally distributed collectives.

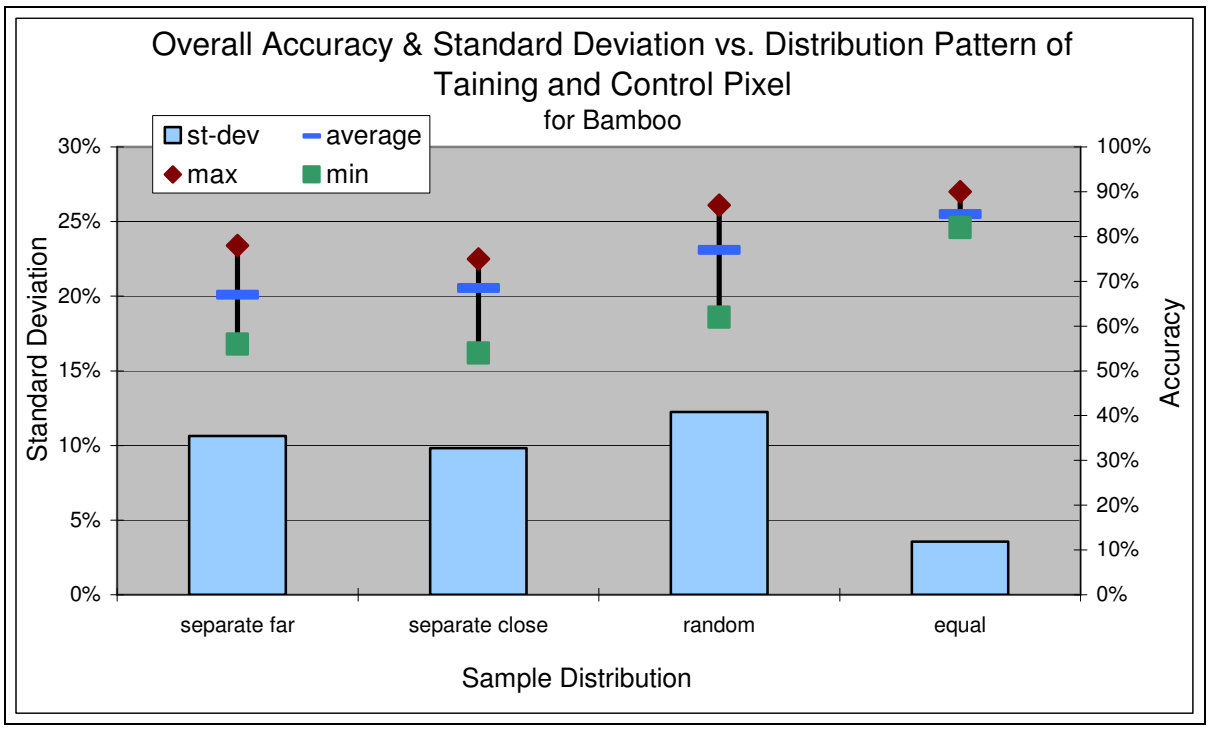


Figure 6-11 Averaged overall accuracy and standard deviation for bamboo vs. distribution pattern of the terrestrial samples. The basis for each distribution category was four runs (k = 5, r = 2, t = 2; band combination 1-10 & 80-89 & 130-139).

In contrast to bamboo alone, a comparison of overall land use types shows an averaged overall accuracy of 79.6 % for a random distribution of training and control pixel, 78.5 % for an equal distribution and the lowest overall accuracy of 76.4 % in the separately distributed collective (figure 6-12). The separately far and wide version of this category was merged. The standard deviation for a separate distribution with 15.7 % lies almost tree times higher than 5.3 % for the equally distributed samples.

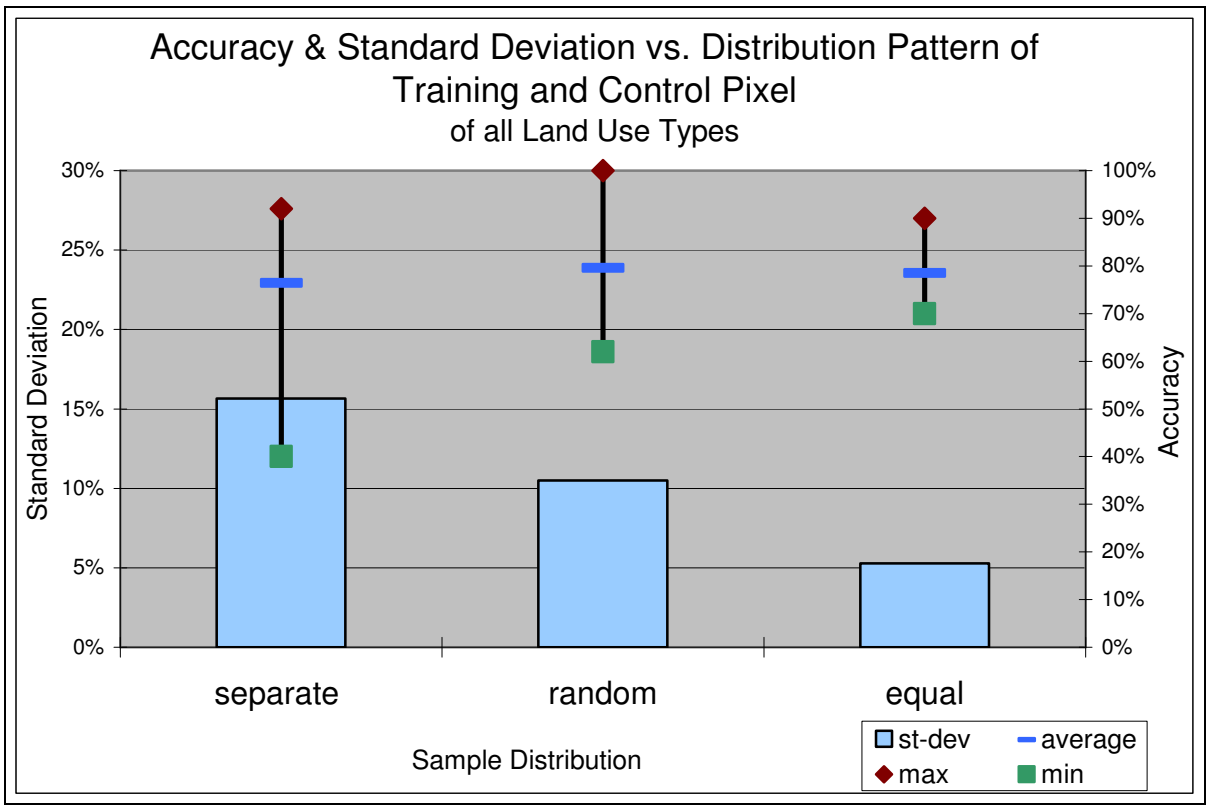


Figure 6-12 Averaged overall accuracy and standard deviation for an average of the ten land use types, depending on the distribution of the terrestrial samples. The basis for each distribution category was four runs (k = 5, r = 2, t = 2; band combination 1-10 & 80-89 & 130-139).

6.2.6 Sample Size

For the dependence assessment of sample size and accuracy of the kNN estimations with the software of Stümer (2004), the cocoa plantations were selected as an example. This land use type was chosen, as it permitted running a series of varying training pixels, up to the kNN programme's maximum extent of 999 pixels, including 499 pixels labelled 'cocoa', or, respectively, 499 pixels 'non-cocoa'. Figure 6-13 shows the averaged accuracy of five runs for the cocoa plantations for different numbers of training pixels. Although the highest maximum accuracy of 83% was found for the number of 650 training pixels, an increasing number of training pixels correlates with increasing values of the averaged overall accuracy. An averaged accuracy of 77.4 % was found for 998 sample pixels, whereas a sample unit of 50 shows an averaged overall accuracy of 69 %. Sample sizes of 800 and more exhibit

a lower standard deviation of the accuracies, approximately below 3 %, whereas lower sample sizes show a higher variation, with a standard deviation of about 4.5 %.

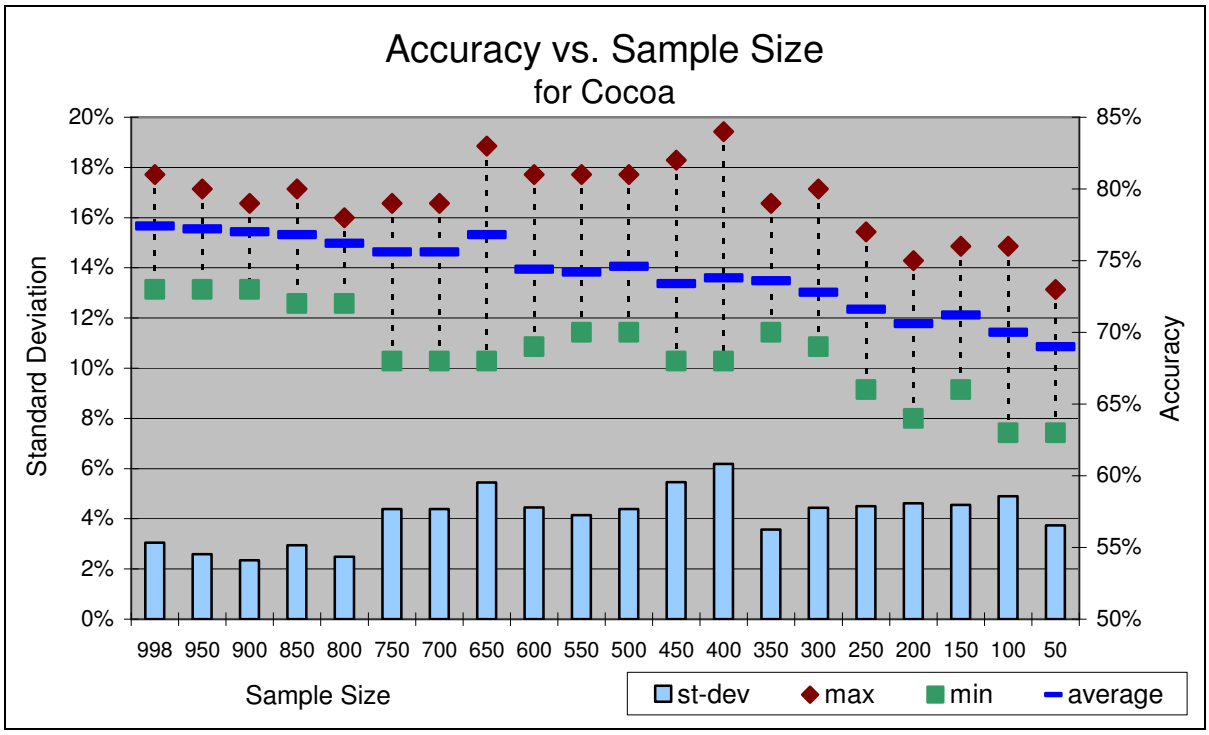


Figure 6-13 Accuracy versus sample size of training pixels. The overall accuracy is based on an average of five runs of the land use type cocoa plantation (k = 5, r = 2, t = 2; band combination 1-10 & 80-89 & 130-139).

6.2.7 Parameters *k*, *r*, *t* of the kNN Programme

The kNN programme of Stümer (2004) supports the input of variable values for the parameters *k*, representing the number of neighbours, the parameter *r* and *t*, influencing the Euclidean distance and distance weight. Varying the numbers of considered nearest neighbours, *k* is defined as the key variable of the kNN programme, whereas the variable *r* stands for the exponent of the Euclidean distance and enables the weighting of the spectral differences of the several bands. The variable *t* allows manipulating the weighting of the distance.

Below, the results for bamboo of three series of varying the number of one of the parameters each time, whilst retaining the other two parameters at the standard settings of the kNN programme, are presented.

In figure 6-14, the averaged (6 runs) overall accuracy for the different *k-nearest* neighbour value series is shown. A peak of the averaged overall accuracy is reached with 81.5 % for *k* = 4, whereas *k* = 5 estimates the presence or absence of bamboo with an accuracy that is still 81.3 % (r = 2, t = 2). With increasing numbers for *k*, the accuracy decreases continuously to 75 % for *k* = 40. An exception can be seen for eight nearest neighbours, with an abrupt rise to 80.8 %.

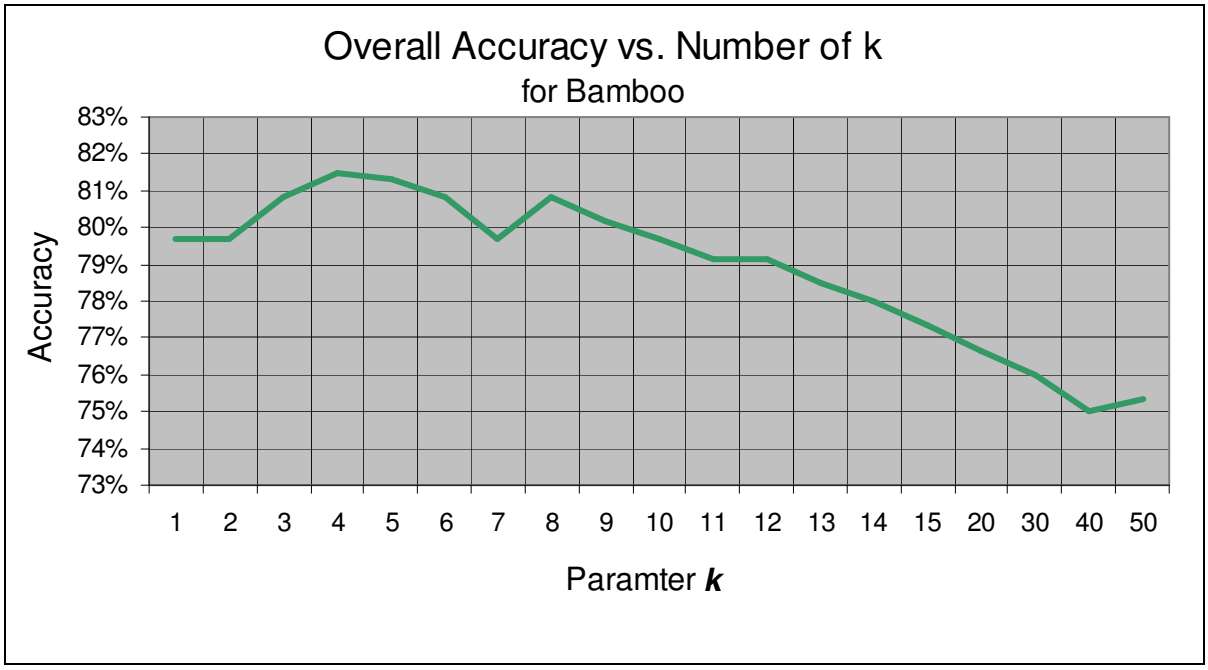


Figure 6-14 Overall accuracy of the kNN estimations for bamboo and varying values for the parameter k. The accuracy is based on six runs (r = 2, t = 2; band combination 1-10 & 80-89 & 130-139).

By replacing the constant exponent of the Euclidean distance with the variable *r*, a metric-coefficient, the individual bands can be weighted. With *r* = 1 all attribute differences are weighted equally, irrespective of their amplitude. For *r* = 2, bigger differences have a stronger weight compared to smaller amplitudes. With *r* = ∞, the maximum attribute difference is weighted with 1, all others are allocated weight 0. For the calculations, varying *r*-values are assessed for bamboo (figure 6-15). The highest averaged (4 runs) overall accuracy of 80 % was found for r = 2 and r = 3. Higher r-values decrease the accuracy to a value of 75 % for r = 15 (k = 5, t = 2).

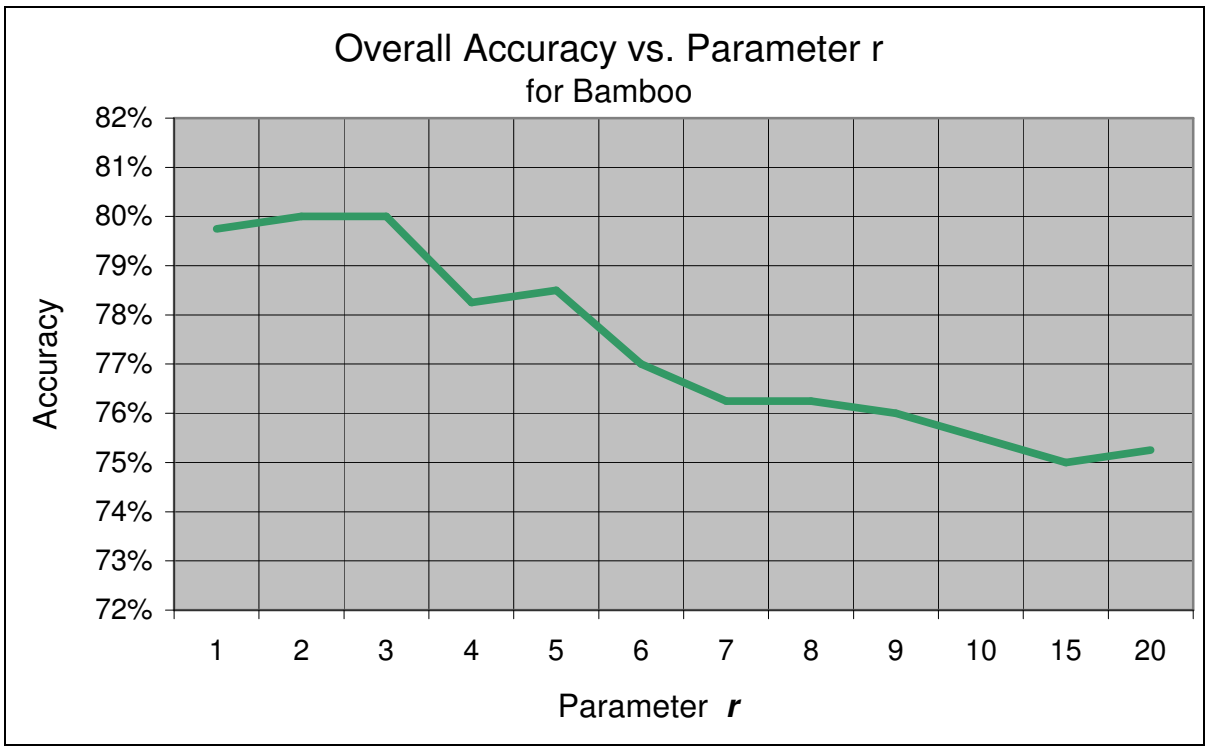


Figure 6-15 Overall accuracy of the kNN estimations for bamboo and varying values for the parameter r. The accuracy is based on four test runs (k = 5, t = 2; band combination 1-10 & 80-89 & 130-139).

The distance values actually only reflect the differences between the spectral information of two pixels. Once a pixel with its terrestrially sampled attributes is allocated to the k-nearest neighbours, these attributes are weighted according to their spectral distance. A higher similarity leads to a higher weighting of the estimated attribute. With the variable t, an exertion of influence on the weighting of the distance is possible. With the setting of higher values for t, higher weightings of pixels result in a lower spectral distance.

Accuracy calculations with t-value series are shown in figure 6-16. The overall accuracy of four runs was averaged, whereas the value for k = 5 and r = 2 remained constant. In figure 6-15, the results are displayed. Only the weighting of the selected neighbours to each other is influenced by the parameter t. The highest accuracies with 81 % were found for t = 5, whereas higher as well as lower values for t result in lessening a decreasing overall accuracy.

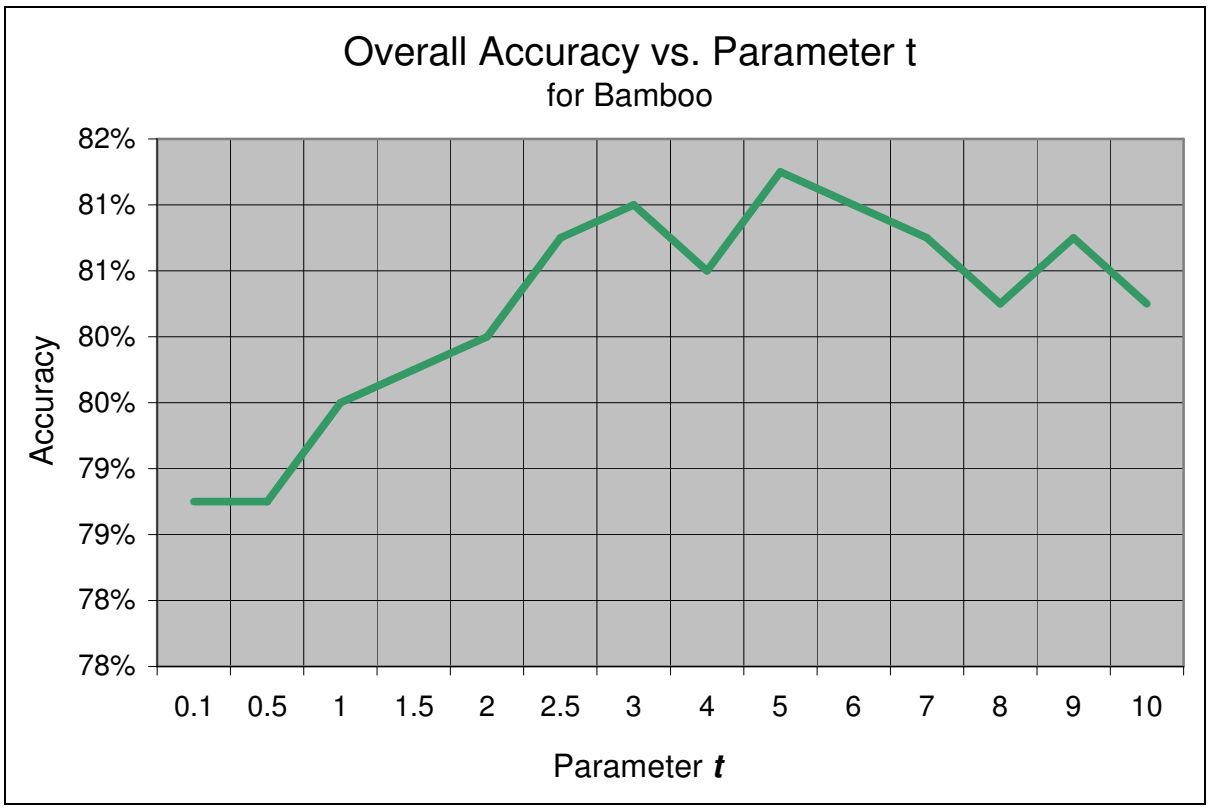


Figure 6-16 Overall accuracy of the kNN estimations for bamboo and varying values for the parameter t. The accuracy is based on four test runs (k = 5, r = 2; band combination 1-10 & 80-89 & 130-139).

The replication of the above kNN estimations with a combination of the optimised parameters k, r, and t of the kNN programme, with k = 4, r = 2, and t = 5 (respectively k = 4, r = 3, t =5), show an averaged overall accuracy of 80.5 % (79.8 %), hence similar values compared to the standard setting of the programme, and an accuracy of 80 per cent.

6.3 Types of Classification Accuracies

Comparison of overall accuracy, user accuracy, and producer accuracy for the kNN estimations for the land use type bamboo demonstrates that the values vary between 83.9 % for the producer accuracy of the attribute ‘no bamboo’ to 81.5 % for the producer accuracy of the attribute ‘bamboo’. Hence, the error of omission is slightly higher for bamboo compared to non-bamboo. The user accuracy describes an inverse situation, were the attribute bamboo shows a slightly lower error of

commission with a user accuracy of 83.5 %, compared to 81.9 % for non-bamboo (table 6-2). The calculations are based on the average of 20 runs with varying collectives of training and control pixels.

Table 6-2 Confusion matrix of the land use type bamboo, based on the average of 20 runs with different collectives of control and training pixels.

	bamboo	non-bamboo	$\sum_{line}$	user accuracy
bamboo	40.75	8.05	48.8	83.5%
non-bamboo	9.25	41.95	51.2	81.9%
$\sum_{column}$	50	50	100	
producer accuracy	81.5%	83.9%	overall accuracy	82.7%

As an illustration, the confusion matrix of the land use types banana and oil palm plantations are presented in table 6-3 and 6-4. In comparison to bamboo, the banana plantations show greater differences between the presence and absence of the attribute, amounting to a difference of 29.6 % for the producer, and 16.3 % for the user accuracy, whereas the differences for bamboo are 2.4 %, and 1.6 %, respectively. The estimated map is much more reliable for the classified non-bananas than it is for the bananas. On the other hand, the situation on the ground can be much better mapped for the attribute characteristic banana compared to non-banana.

Table 6-3 Confusion matrix of the land use type banana plantation, based on the average of 20 runs with different collectives of control and training pixels.

	banana	non-banana	$\sum_{line}$	user accuracy
banana	44.95	19.85	64.8	69.4%
non-banana	5.05	30.15	35.2	85.7%
$\sum_{column}$	50	50	100	
producer accuracy	89.9%	60.3%	overall accuracy	75.1%

Within the oil palm plantations, the producer accuracy of 88.5 % for the presence, and 78.4 %, respectively, for the absence of the attribute exhibits quite a difference in the errors of omission. The user accuracy of 80.4 % for the presence and 87.2 % for the absence of oil palm plantations shows a higher reliability of the estimated non-oil palm pixels, compared to pixels which are classified as oil palm plantations.

Table 6-4 Confusion matrix of the land use type oil palm plantation, based on the average of 20 runs with different collectives of control and training pixels.

	oil palm	non-oil palm	$\sum_{line}$	user accuracy
oil palm	44.25	10.8	55.05	80.4%
non-oil palm	5.75	39.2	44.95	87.2%
$\sum_{column}$	50	50	100	
producer accuracy	88.5%	78.4%	overall accuracy	83.5%

6.4 Kappa Coefficient

The overall agreement of a classification is given by the kappa coefficient. The averaged kappa values of ten runs for the several land use types are presented in table 6-5. The kappa values vary between 0.669 for oil palm plantations and 0.654 for bamboo, to a minimum of 0.502 for banana/plantain plantations, thus showing a

substantial (e.g. oil palm, bamboo) to moderate (banana) agreement with the classification according to Landis & Koch (1977). An average over all land use types results in a kappa value of 0.583, which is defined as moderate agreement.

Table 6-5 Kappa coefficient of the land use types.

LUT	bamboo	banana	bush	cocoa	elephant	grass	herb	oil palm	raphia	trees forest	average
Kappa Coefficient	0.654	0.502	0.534	0.604	0.548	0.624	0.614	0.669	0.517	0.564	0.583

6.5 Classification Procedure including all Land Use Types at once

The optimization analysis of the kNN method was conducted with the selection of a single land use type, attributed with the presence or absence of this land resource. For map generation of single resource distribution maps, this procedure is valid. For making multi-attribute maps, such as land cover maps and land use classifications, these single land resource maps can be combined, the layer of each land use type of interest can be overlaid, respectively. Another possibility is to include all land use types of interest at once, in the form of a multi-attribute classification. Below, the kNN procedure is applied to all the ten land use types, defining ten different classes, attributes for the kNN programme, respectively. The results of the overall accuracy for an occurrence probability $p > 0.2$ and $p > 0.5$, as they are calculated and output by the kNN programme of Stümer (2004), are shown in figure 6-17. The calculation of overall accuracy and standard deviation are averaged, based on ten runs with different collectives of training and test pixels.

The overall accuracy is increased from 55 % to 70 %, if only output pixels with an occurrence probability $p > 0.5$ are included in the accuracy calculations, compared to the standard value of $p > 0.2$ for a ten-attribute classification. Compared with the single attribute classifications above, lower accuracies are found for multi-resource estimations. However, the standard deviation of the several test runs show a lower value of 4.2 % for $p > 0.2$, compared to 5.9 % for an occurrence probability of $p > 0.5$.

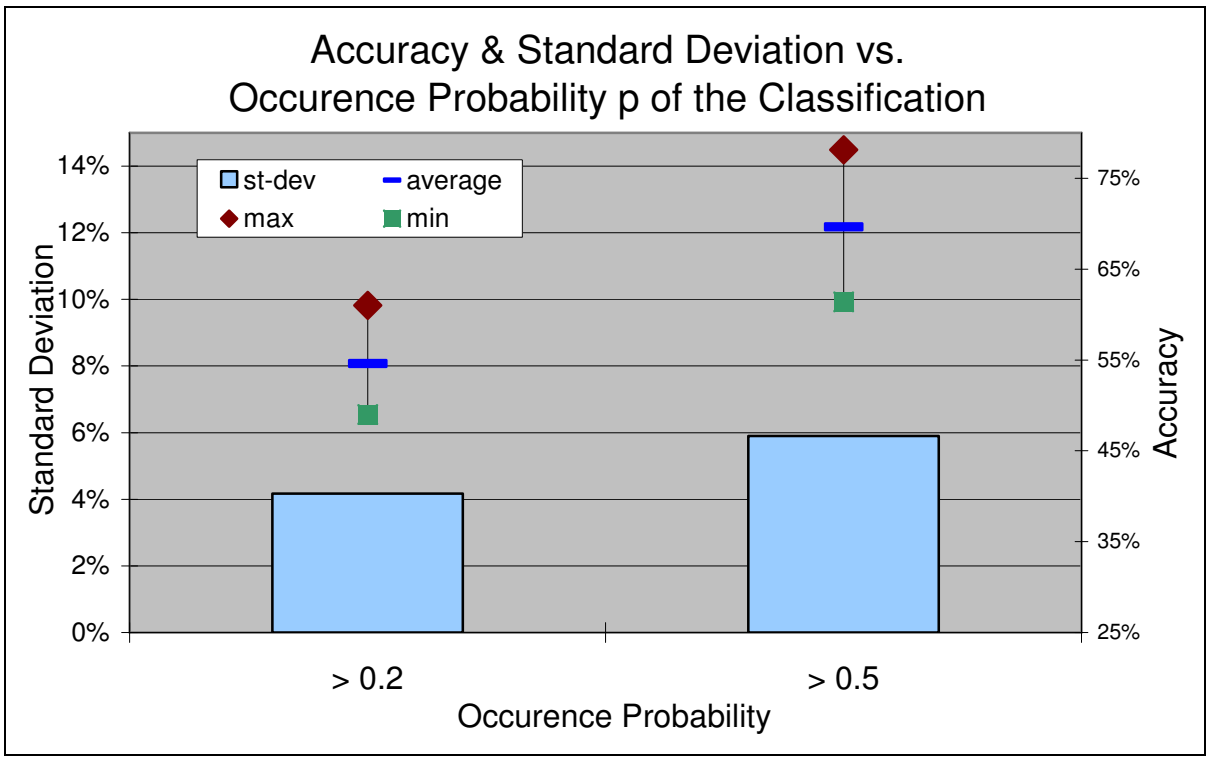


Figure 6-17 Overall accuracy & standard deviation of the classification of the ten land use types with different occurrence probabilities p.

The confusion matrix (table 6-6 & 6-7) of the multi-resource classification exhibits in which way individual pixels are misclassified. It permits the detection of potential similarities of reflection for specific land use types. In this study, bush fallow and banana plantation stand out in that high proportions of up to 23 % are found to be misclassified and confused with each other. Similar values are found for the raphia palms, where 19 % are wrongly classified as cocoa plantations, whereas 12 % of cocoa were wrongly classified as banana plantations.

By comparing the producer accuracy of the different land use types, it can be seen that the land use for elephant grass and grassy vegetation can be mapped best of all, an occurrence probability $p > 0.2$, and $p > 0.5$. The producer accuracy of these land use types is 68 % for $p > 0.2$, whereas elephant grass reaches a producer accuracy of 89.4 % for $p > 0.5$. The lowest producer accuracy was found for bush fallows, ranging from 37 to 37.5 %. The highest reliability of the estimations was also found for the grassy vegetation with a user accuracy of 71.6 % ($p > 0.2$). For $p > 0.5$, the highest user accuracy was calculated for bamboo (84.6 %). The lowest map

reliabilities were found for banana (40.9 %, p > 0.2) and bush fallow (44.1 %, p > 0.5).

By restricting classified pixels to an occurrence probability p of at least 0.5 instead of 0.2, 42.1 % of the pixels are rejected and will not distribute to the resulting classification. Output pixels below the minimum occurrence probability have to be labelled as “unclassified” on the map. However, the remaining pixels show an increase in overall accuracy from 54.6 % (p > 0.2) to 69.8 % (p > 0.5) and an increase of the kappa value from 0.496 to 0.663, as is shown in table 6-6 and 6-7.

Table 6-6 Confusion matrix of the land use classification with the standard occurrence probability p > 0.2. The results of ten runs with varying collectives of training and test pixels are summed up.

p > 0.2	bamboo	banana	bush	cocoa	eleph. Grass	grass	herb	oil palm	raphia	trees	$\sum_L$	user a. [%]
bamboo	49	2	2	8	4	2	1	1	6	3	78	62.8
banana	4	54	23	12	1	6	10	10	4	8	132	40.9
bush	1	12	37	8	3	6	4	3	4	8	86	43.0
cocoa	12	4	5	47	1	3	3	2	19	10	106	44.3
eleph.	4	9	7	2	68	5	4	4	0	7	110	61.8
grass	2	3	3	2	5	68	2	9	1	0	95	71.6
herb	9	5	7	6	7	3	60	8	0	2	107	56.1
oil palm	5	6	3	0	4	7	6	56	2	2	91	61.5
raphia	7	1	2	6	2	0	1	0	57	10	86	66.3
trees	7	4	11	9	5	0	9	7	7	50	109	45.9
$\sum_C$	100	100	100	100	100	100	100	100	100	100	1000	
producer acc. [%]	49	54	37	47	68	68	60	56	57	50	OA:	54.6
											K:	0.496

Table 6-7 Confusion matrix of the land use classification with an occurrence probability p > 0.5. The results of ten runs with varying collectives of training and test pixels are summed up.

p > 0.5	bamboo	banana	bush	cocoa	eleph.	grass	herb	oil palm	raphia	trees	$\sum_L$	user a. [%]
bamboo	33	1	1	1	1	0	0	1	0	1	39	84.6
banana	0	37	11	3	0	4	3	3	1	1	63	58.7
bush	0	6	15	1	0	4	3	0	1	4	34	44.1
cocoa	6	2	3	33	0	0	1	1	10	6	62	53.2
eleph.	1	8	3	0	59	4	1	1	0	5	82	72.0
grass	1	2	2	1	2	59	1	6	0	0	74	79.7
herb	1	1	3	1	1	0	50	3	0	1	61	82.0
oil palm	2	1	0	0	2	2	2	43	2	1	55	78.2
raphia	4	0	0	4	0	0	1	0	50	6	65	76.9
trees	1	1	2	3	1	0	5	2	4	25	44	56.8
$\sum_C$	49	59	40	47	66	73	67	60	68	50	579	
producer acc. [%]	67.3	62.7	37.5	70.2	89.4	80.8	74.6	71.7	73.5	50.0	ov. acc.:	69.8
											K:	0.663

6.6 The kNN Classification Maps

For the compilation of land resource distribution or land use maps, the kNN method was applied for ordinal numbers in this study. Besides the geographic coordinates, the output file of the kNN programme comprises the probability of the occurrence of the relevant land resource and thus, the derived class, each pixel is allocated to. For a one-attribute classification, an output pixel is classified as ‘existing’ or ‘non-existing’; for the land use type bamboo, for instance, a pixel is, therefore, either estimated to represent ‘bamboo’ or ‘non-bamboo’. The probability threshold of such a one-attribute classification with two parameter values constitutes 0.5. A pixel is attributed to the characteristic feature ‘bamboo’ with a probability threshold above 50 per cent. At a probability below 50 %, the pixel is estimated to represent ‘non-bamboo’ for that specific location.

In figure 6-18, a detail of the estimated bamboo distribution map is shown for three different allocations of training pixels. The number of bamboo input pixels remained at 254, only the number of non-bamboo pixels varied from 254, 508, and 745,

respectively. This was to display the difference of the classified maps in respect to the influence of varying ratios of bamboo and non-bamboo training samples. In general, a similar distribution pattern of bamboo can be recognised for the three images, though higher proportions of bamboo output pixels, as well as higher occurrence probabilities are found with increasing proportions of bamboo input pixels.

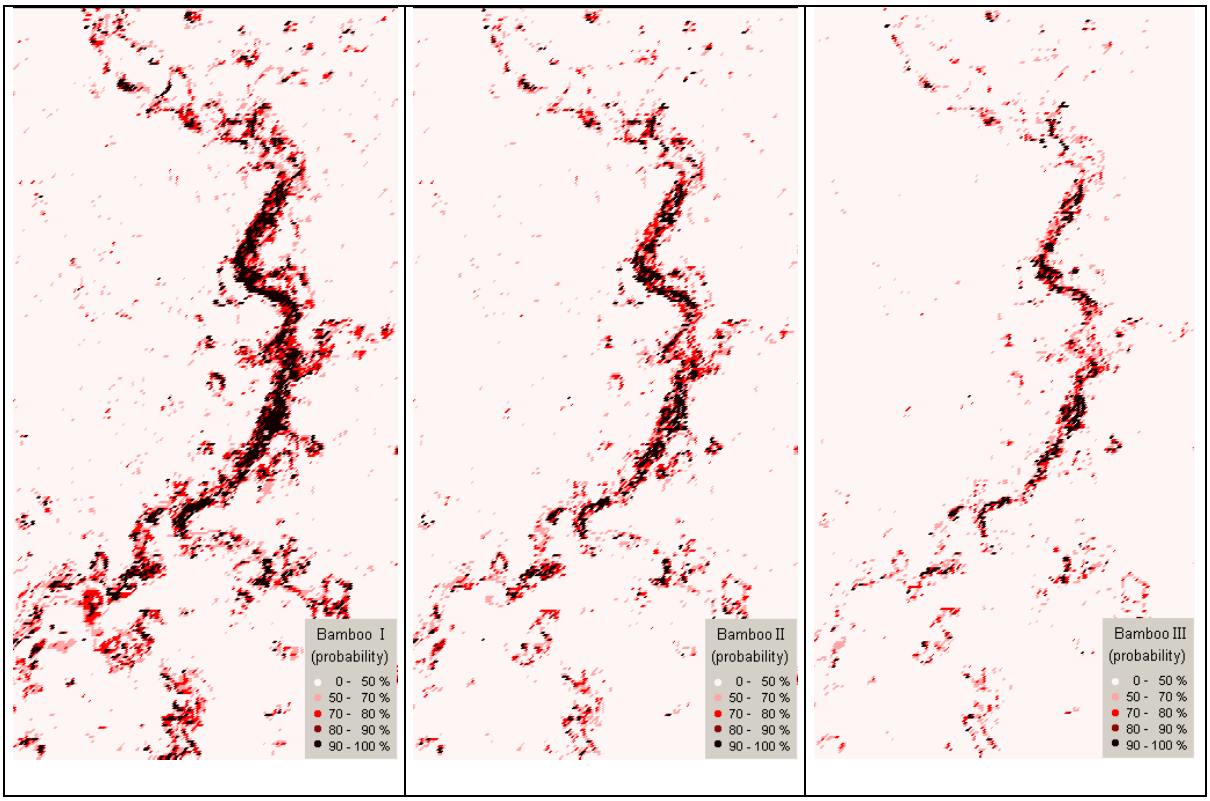


Figure 6-18 Extraction of the bamboo distribution maps and the occurrence probability of each resulting pixel. The three maps represent different ratios of bamboo and non-bamboo input pixels: 254:254 (I), 254:508 (II), and 254:745 (III).

To quantify the influence of the sample distribution of bamboo to non-bamboo training pixels, the number and proportion of bamboo and non-bamboo output pixels of the classified image is listed in table 6-8. Additionally the values of a multi-attribute classification, with a ratio of 97 to 899 pixels per land use type, are listed. With an equal proportion of bamboo and non-bamboo training pixels, bamboo is estimated to cover 22 % of the surface of the study area. With increasing proportions of non-bamboo training samples, the estimated bamboo cover drops to 6 % of the area, at a ratio of 1:10 bamboo to non-bamboo training pixels.

Table 6-8 Distribution of bamboo and non-bamboo pixels within the collective of training pixels and the classified resource map of the study area.

Origin:	Bamboo I		Bamboo II		Bamboo III		Multi-Attr. Class.	
Attribute Feature	Bamboo	Non-bamboo	Bamboo	Non-bamboo	Bamboo	Non-bamboo	Bamboo	Non-bamboo
No. Training Pixels	254	254	254	508	254	745	97	902
Proportions	50 %	50 %	33 %	67 %	25 %	75 %	10 %	90 %
No. Output Pixels	1677313	5970761	985394	6662680	691430	6956644	439293	7208782
Proportions	22 %	78 %	13 %	87 %	9 %	91 %	6 %	94 %

The corresponding results for banana and oil palm plantations are similar (see tables 6-9 & 6-10). The estimated banana surface cover of 41 % at a ratio of 1 to 1 for banana and non-banana input pixels, drops to 12 % surface cover for a proportion of only 10 % of the input pixels, registered as banana. With a proportion of 50 % of the input pixels being oil palms, the surface of the study area is estimated to cover 28 %. This value drops to 6 % surface cover, for a proportion of 10 % of the input pixels describing oil palm plantations. Hence, the partitioning of training samples has a clear effect on the estimation results of all three land use types; this effect is non-linear however.

Table 6-9 Distribution of banana and non-banana pixels within the collective of training pixels and the classified resource map of the study area.

Origin:	Banana I		Banana II		Multi-Attribute Classification	
Attribute Feature	Banana	Non-banana	Banana	Non-banana	Banana	Non-banana
No. Input Pixels	325	325	325	674	97	902
Proportions	50 %	50 %	33 %	67 %	10 %	90 %
No. Output Pixels	3099156	4548918	2021388	5582626	959982	6688092
Proportions	41 %	59 %	27 %	73 %	12 %	87 %

Table 6-10 Distribution of oil palm and non-oil palm pixels within the collective of training pixels and the classified resource map of the study area.

Origin	Oil Palm I		Oil Palm II		Multi-Attribute Classif.	
Attribute Feature	Oil Palm	Non-oil palm	Oil Palm	Non-oil palm	Oil Palm	Non-oil palm
No. Training Pixels	317	317	317	682	97	902
Proportions	50 %	50 %	32 %	68 %	10 %	90 %
No. Output Pixels	2126330	5521744	1340343	6307731	459203	7188871
Proportions	28 %	72 %	18 %	82 %	6 %	94 %

The resource distribution maps for banana (figure 6-19) and oil palm plantations (figure 6-20) show the distribution estimations of banana, respectively, oil palm plantations and the corresponding occurrence probabilities. For a greater clarity, detailed extracts of the study area were chosen. Two map versions, based on different allocations of input pixels, are displayed. Analogous to bamboo, for each land use type, individual spatial distribution patterns can be recognised for banana and oil palm plantations.

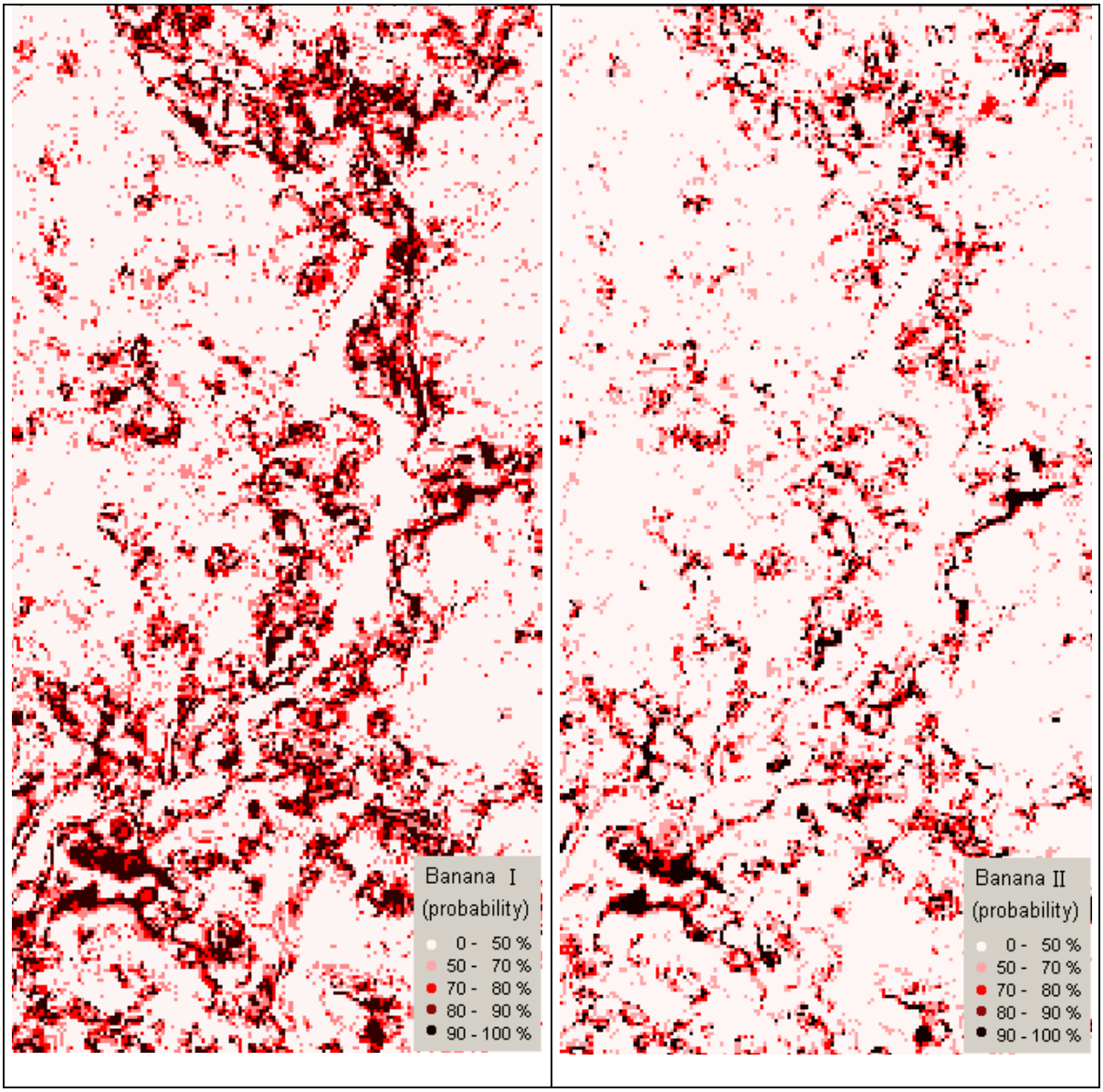


Figure 6-19 Extraction of the occurrence probability map of banana with different ratios of banana and non-banana input pixels: 325:325 (I) & 325:674 (II).

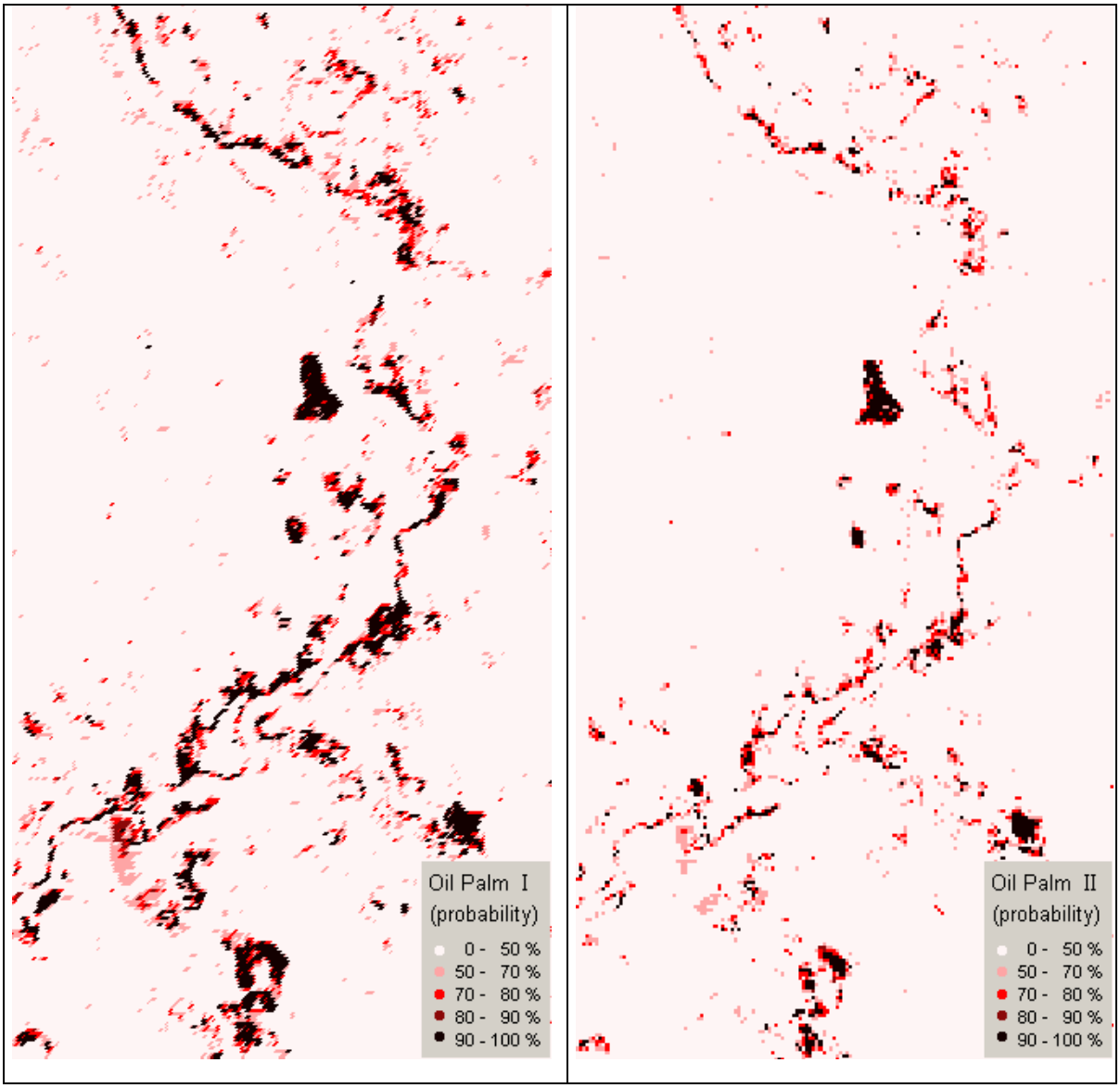


Figure 6-20 Extraction of the occurrence probability map of oil palm with different ratios of oil palm and non-oil palm input pixels: 317:317 (I) & 317:682 (II).

To compare the result variability of multi-attribute classifications, two kNN estimation runs with randomly selected training samples (999 out of 3360) are shown for the ten land use types in figure 6-21. The maps show a detail of the classified study area and correspond to the bamboo, banana, and oil palm distribution maps above. The double s-curve of bamboo occurrence (light green) along an invisible stream in the centre of the image, similar to the bamboo maps above, can be identified, but at a significantly lower occurrence. At first glance, the distribution pattern of the ten land use types looks similar for the two images. Nevertheless, differences can be visibly recognised.

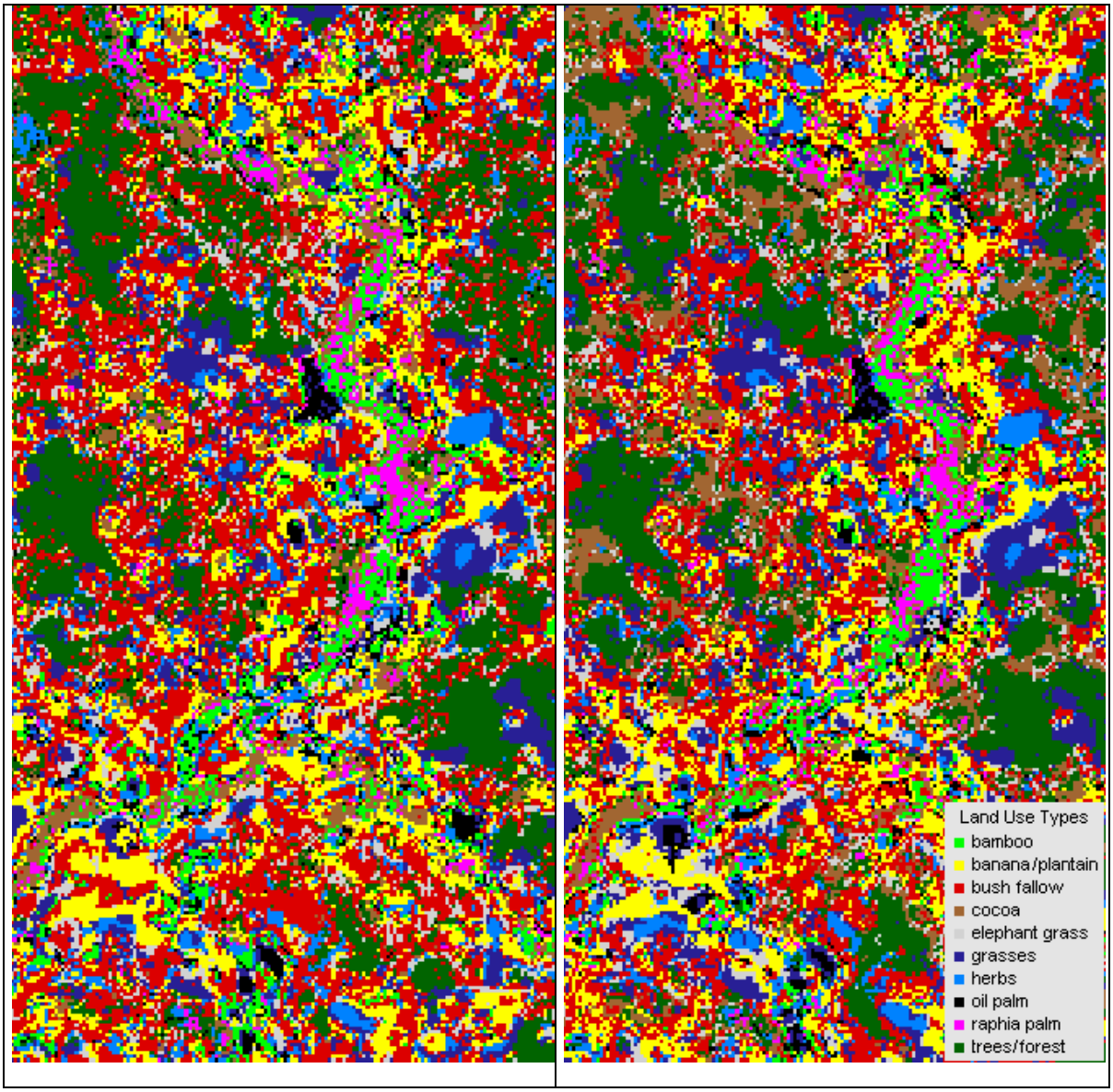


Figure 6-21 Extraction of the classified map of the study area with an occurrence probability p > 0.2.

The detailed land cover proportions for each land use type of the whole study area are listed in table 6-11. Within the study area, the highest proportions (averaged) of land are covered by grassy vegetation (15 %), forest (15 %), and bush fallows (14 %). A medium cover is observed for banana plantations (12 %), elephant grass (11 %), and cocoa plantations (9 %). Lower covers are shown by herbaceous vegetation (8 %), oil palm plantations (7 %), bamboo (6 %), and raphia palms (5 %). Slight differences of land cover, smaller than two percentage points, are observed for the two classification runs, although a general trend can be concluded. The classification image of the complete study area is shown in figure 6-22.

Table 6-11 Surface cover of the land use types estimated with the kNN method. The results of two runs with different selections of training plots are listed.

Land use type	classification 1		classification 2		average
	no. pixels	proportion	no. pixel	proportion	
bamboo	411045	5.4%	439572	5.7%	6 %
banana/plantain	914138	12.0%	857517	11.2%	12 %
bush fallow	1091525	14.3%	987070	12.9%	14 %
cocoa plantation	643241	8.4%	702813	9.2%	9 %
elephant grass	843221	11.0%	892218	11.7%	11 %
grassy vegetation	1076893	14.1%	1229268	16.1%	15 %
herbaceous veg.	629813	8.2%	591246	7.7%	8 %
oil palm	518546	6.8%	506479	6.6%	7 %
raphia palm	375300	4.9%	335255	4.4%	5 %
trees/forest	1144352	15.0%	1106636	14.5%	15 %
total	7648074	100%	7648074	100%	100 %

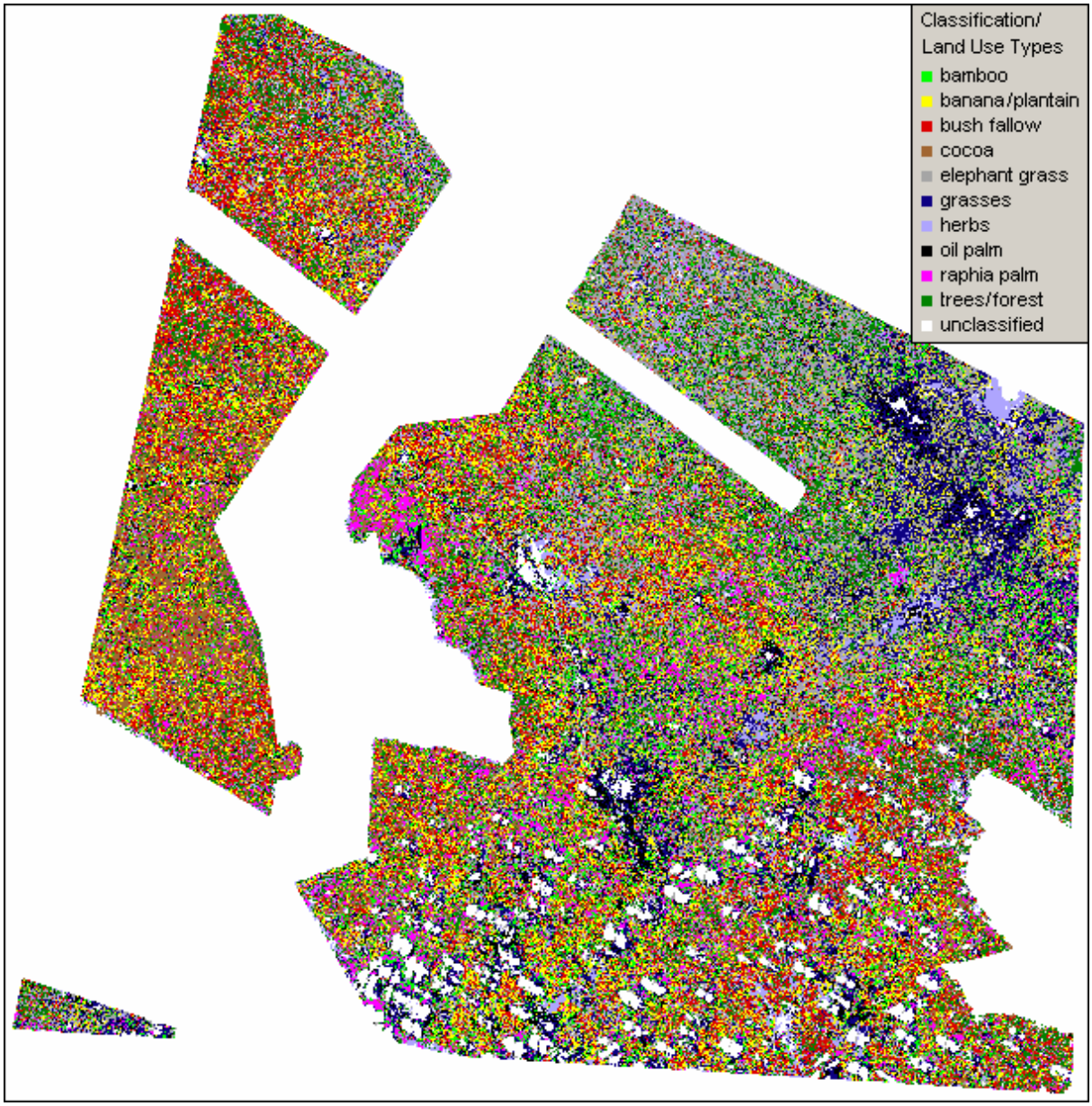


Figure 6-22 Land use type classification of the study area with the kNN method. The white areas are unclassified pixels and represent forest reserves, built up areas, clouds, and shade of clouds. The occurrence probability for a pixel was p > 0.2.

7 Discussion and Conclusions

7.1 Discussion of Sample Size and Design

The reasons for the chosen inventory design have already been mentioned in chapter 5.1, nevertheless further aspects can be described as follows. The study area stretches across an area of more than 3,500 km^2. A 15 by 15 m pixel based classification of such a large area, is necessarily at the cost of the sample size intensity. General stratification failed due to the lack of reliable and sufficient data and information concerning the area, though at least for bamboo and raphia palms, stratification in the form of a defined buffer along streams and rivers would be conceivable, as these land use types mostly occur along streams. The problem here was the small number of sample units, inadequate stream maps, and partially dry riverbeds during fieldwork and satellite image taking. For the sampling of abandoned rice fields, successionally vegetated by elephant grass, a stratification on the basis of their close connection with lowland conditions is conceivable.

Beyond this, the coarsely meshed 7 by 7 km grid entailed the disadvantage of missing out formative landscape patterns or even complete and rare land use types, such as rice fields or teak plantations. Nevertheless, half a dozen teak plantations were recorded and being aware of the fact that in some parts of the study area, large proportions of rice fields are present, they have been completely omitted from the accuracy analyses and are not represented in the training collectives. Hence, despite their existence, they do not occur in the classified maps, but are represented instead by the remaining land use types. Teak plantations, depending on their age, might be covered by the land use type forest, rice fields by their subsequent successional stage after they were abandoned, the elephant grass or grassy vegetation.

Another issue is the procedure of, at least partially, selectively seeking examples of each land use type and not completely sampling them randomly, a procedure which entails bias. This structural weakness is mildened by a maximised sample size for each land use type. It is assumed that the most typical characteristics of each land use type were sampled, consequently the variety of reflection values for each land use type were likely to be sufficiently covered with a sample size of 111 to 600 pixels

for each type. Merely mosaics of altering land uses, smaller than pixel size, could not be recognised.

The overall agreement of an image classification strongly depends on the degree of conformance between the measured position of a GPS receiver and its true geographic position, respectively, the applied satellite image. If training pixels, collected during fieldwork, have already been recorded with wrong geographic coordinates, or errors are unknown, the results of the classification lack validity. Analyses with regard to accuracy of the geometric image correction, as well as the expected position errors during terrestrial data collection, were made. The geo-referenced satellite image already features errors of up to five metres in 79 %, 5-10 metres in 15 %, and 10-15 metres in 6 % of the evaluation points. The geographic spread of the GPS receiver during field sampling brings in another uncertainty of up to nine metres (1 %) to the geographic data. Hence, deviations of 20 metres between satellite image and corresponding sample pixel are possible. For the terrestrial sampling, only sample plots with a minimum radius of at least 15 metres were collected to reduce the problem of position errors. However, particularly underneath tall bamboo clusters, in densely vegetated stands of forest and bush fallow, GPS position errors tend to increase. For some land use types, such as banana or grassy vegetation, the small scaled mosaic of alternating vegetation, intercropping and mixtures of land uses precludes increasing the minimum sample area. Consequently, an unknown number of falsely classified sample pixels went into the database and might have a considerable influence on classification accuracies.

After exclusion of the forest reserves, the remaining off-reserve forests still covers an area of 1720 km^2, featuring a sampling intensity of about 0.02 plots per ha. Similar studies show sampling intensities of 0.015 to 0.1 plots per hectare (Wallerman, 2003; Stümer, 2004; Gagliano et al., 2007). For the classification of forest/non-forest/water, the study area of Haapanen et al. (2004) encompasses approximately 29,748 km^2 and features a sampling intensity of 0.03 samples per square kilometre. Here it has to be considered that the selected land covers feature defined reflection values.

However, for the achievement of references and guide values for the development of a monitoring system and to extract recommendations for the classification of these areas, the present sampling intensity and design is presumed to be sufficient.

7.2 Band combination

In comparison with similar studies (Holmström, 2001; Stümer, 1994; Cabaravdic, 2007), where combinations of several optical satellite data, radar data, or other additional information like Digital Elevation Model (DEM) increased the results of kNN estimations, for this study only one scene of the ASTER sensor was available. In the present study, a general increase of the bands used does not necessarily improve the classification accuracy. Indeed, gradually adding up the original ASTER bands 1 to 9 (including 3b) lead to an accuracy increase. In order to obtain a further improvement, particular band combinations or indices have to be added. Whereas some particular band combinations lead to an accuracy increase, the same bands may have a negative effect on another land use type. This effect was also noticed within one land use type when training and test collective were interchanged. It is likely that, in these cases, a stronger interconnection between pixel selection, compared to band selection exists. Altogether the band combinations 1-10 & 80-89 & 130-139 and 80-84 & 117 showed the best results as regards all land use types. With reference to band combination 80-84 & 117, all bands belong to the visible-NIR wavelength region, complemented by the vegetation index (band 117). Surveying vegetative cover basically relies on VNIR spectral bands, but also includes vegetation indices (Lacaze, 1996; Becker et al. 2005; Khunrattanasiri, 2006). Band combination 1-10 & 80-89 & 130-139 also comprises bands in the SWIR wavelength region. The ´Tasselled Cap´ transformation, which was represented by bands 130-139, also showed a positive effect, whereas the NDVI did not contribute so clearly. For estimations of the leaf area index (Lee et al. 2006), NIR reflectance showed positive correlations, although the correlation strength is weaker than in SWIR and visible regions and no significant correlation was found for NDVI.

Despite a higher number of applied bands, thirty bands compared to only six bands, hence longer computation time, band combination 1-10 & 80-89 & 130-139 was chosen for most of the adjacent analyses. The reason here was a slightly lower variability in accuracy values amongst the different land use types. The higher number of bands applied might also help to adjust extreme results, as the risk of chance is reduced, compared to band combination 80-84 & 117. For large image files though, priority could be given to a lower number of bands, if computer restrictions and efficient computing time plays the decisive role. In contrast to the accuracy analyses in chapter 6, band combination 80-84 & 117 was used for the generation of land resource distribution maps. This was done with respect to the large image file, covering more than 7.5 million pixels. The computation of a single image with the kNN programme and an Intel Pentium M processor, 1.73 GHz, 6 bands (80-84 & 117), k = 5, and a sample size of 999 training pixels took about 36 hours. With band combination 1-10 & 80-89 & 130-139 it would have been a multiple period of time.

This study is not meant to identify the only and optimal band combination for the estimations. On this account, detailed analyses on the ideal band combination were set aside or done only cursory and when apparently justified. However, it became obvious that the selection of spectral bands is of high significance. Particularly in the case of large investigation areas, the number of selected bands should be minimized and optimised. If major canopy trees (shading timber species) are ignored, the different land use types exhibit typical maximum vegetation heights. Radar data and aerial images might be of vital help to mark-off the boundaries of the land use types, thus significantly increasing the classification results. Haapanen et al. (2002) found that using multiple image dates for an area typically considerably improves results.

7.3 Accuracy, Precision and Overall Agreement

For some land use types (e.g. grassy vegetation), higher variances in overall accuracy for repeated test runs with various, random selections of training and test pixels were found, compared to others (e.g. bush fallow, herbaceous vegetation). This offers the assumption that some land use types show a more characteristic

pattern of reflection values than others. The high precision of bush fallow and herbaceous vegetation is most likely caused by the fact that these lands have been abandoned and define a specific successional stage with dense vegetation. These land use types are not exposed to any human activities or treatment until they have been transformed into another land use type. Comparison of different locations shows a similar variety of growth patterns and species composition. Therefore, sample location, hence pixel selection for training and test collective, might not play such an important role. On the other hand, the grassy vegetation showed a comparatively low precision, which could result from the intensive use as pasture and frequent or seasonal changes in treatment, for example, burning or weeding. At one location, appearance is quite similar and homogeneity is quite high. Compared to another location, the situation might have changed completely, as grazing intensity is lower or soil humidity differs if one location is rather situated at the top of a hill and the other one in a flat, plain area. Additionally the low number of sample units, hence insufficient cover of variety for the grassy vegetation, might contribute to the low precision.

Other examples of a lower precision of the repeated estimations are bamboo and elephant grass. As bamboo is a land use type that hosts almost no shade trees or additional vegetation layer and the bamboo clusters permanently remain in one location, the reason must lie somewhere else. Probably the different density of the clumps and the extent of the exploitation (up to 90 % was recorded) of the shoots and canes mean the bare soil on the ground of the area has some influence on the reflection values at one location of bamboo clusters. Elephant grass appears to grow fairly homogenously and, like bamboo and oil palm plantations, it is quite unusual that shading trees, which influence reflection, are found on these sites. On the other hand, it was found that ground conditions as regards soil dampness may vary dramatically within and amongst the different sites. The low number of samples might contribute to this result as well.

The different number of sample pixels used for each land use type is very likely to be another important factor for the different estimation precision, as well as the accuracies of the results. As the only reason it must be rejected, because highest

and lowest accuracy was found for oil palm and banana plantation, whereas both had a middle position as regard to sample size. Nevertheless, the two land use types with the lowest number of sample size, elephant grass and raphia palms, respectively, exhibited accuracies below average. This tendency was also proven with the varying sample sizes for the cocoa plantations versus accuracy. Higher sample sizes showed a trend to higher accuracies. Similar results were observed for the precision; a higher sample size showed the tendency to lower standard deviation of repeated runs with varying selections of training and test pixels. When the several land use types are taken into consideration, this rule could also be applied in the reverse direction.

By comparison of the averaged accuracy with several runs of randomly selected training and test pixels, it was found that the curves for the three selected land use types bamboo, oil palm plantation, and banana plantation flattened, or at least the volatility dramatically declined, after about six reruns. It is certainly true that the selected examples show some differences among one another, but this has to be considered for a monitoring design and for practical applications, respectively. Particularly when considering the remaining seven land use types, it must be presumed that every land use type itself has different minimum requirements as regards variance of the expected results.

Observing the accuracy for the particular land use types, it was found that some land uses tend to have higher estimation accuracies compared to others for both aspects, that is in comparison with different band combinations as well as for replications with varying sample collectives. Classification accuracies above average were found for oil palm plantations, followed by grassy vegetation, herbaceous vegetation, and cocoa plantations. The oil palms are a long-lasting cash crop, with well defined boundaries, intensively maintained and not inter-planted with any other crops. The ground is generally covered with low growing grasses and herbs. Merely plantation age, hence spacing, differs to a great extent, particularly in the case of younger plantations. These facts might be the reason for the relatively high estimation accuracy. The grassy vegetation, mainly used as grazing land for cattle, is subject to high maintenance as well. Connected with that is the fact that this land use type basically involves grasses and has a homogenous appearance within one pixel. With

higher proportions of herbaceous species, the land use type is defined as herbaceous vegetation, then supporting less homogeneity per pixel, with respect to species richness and vegetation height. Together with the bamboo clusters, these land use types lack shade trees and occur relatively homogenously within one pixel.

The cocoa plantations, the major cash crop in the region, experienced the highest maintenance and treatment over time. Intercropping is rather unlikely, except for fruit trees at times and homogeneity of this land use type is very high. However, a shading canopy of timber trees, or at least scattered tree individuals dispersed within a coca farm, influence the reflection values to a high proportion.

In contrast, banana plantations are very heterogeneous. They occur densely packed with tall plants, intercropped with all types of seasonal vegetables, show high proportions of bare soil, particularly at a young stage. As they function as a subsistence crop as well as a cash crop, sold at local markets, they are planted in a vast variety of shapes, sizes and intermixtures with other crops. They might cover areas of several hectares with only few giant trees above, as well as on a patch, smaller than pixel size, surrounded by dense cocoa plantations or adjacent scattered tall trees. The sample size of the bush fallows was the highest of all land use types and might have contributed positively as regards, precision of the estimations. The last but one position with respect to overall accuracy was surprising though. This land use type was found at almost every location, the reason for the high number of sample units, and, consequently, meaning that it was not very specific as regards topography, moisture conditions, etc. Just this fact might lead to a vast variety of all kinds of reflection values, as this land use type appears to be densely packed with overgrowing herbs and only scattered small shrubs, or, contrary to that, with a dense stand of tall shrubs and only small proportions of ground vegetation. In-between, all types of successional stages might appear, sometimes interrupted by slash and burn activities and intercropping with seasonal vegetables, or abandoned for many years, almost representing a forest patch.

Not all land use types are explicitly analysed here. The recorded sample size and observed occurrences of raphia palms, for example, were considered to be

substantially insufficient. The tree and forest patches involved the problem that most land use types do host tree individuals, or more or less scattered timber trees already form a canopy layer above a specific land use type. Another issue is that isolated giant trees are sampled as individuals, with any kind of other land use type underneath.

The problem of all interpretations in this connection is that assumptions are based on the accuracy and precision of the classification results. Quantitative analyses of actual reflection values and the differences in reflection for individual sample pixels are lacking. With the kNN software used in this study, this is not applicable immediately. Though the arrangement of such analyses is possible with the present data, they are very time-consuming if no adequate software is available.

The classification accuracies obtained in the study vary dramatically, depending on the observed land use type and particularly with regard to the number of classes applied. Averaged overall accuracies of up to 83 % are acquired for oil palm plantations, whereas banana plantations were located at 75 %. Far beneath these figures are the results when classifying all ten land use types at once. The classification accuracy then lay at 55 %. With a mark up of the occurrence probability from > 0.2 to > 0.5, this value could be improved to 70 %, with the limitation of about 40 % of the pixels rated as being unknown or unclassified, respectively. For the evaluation of the results, it has to be considered that each sample plot might comprise several land uses. The dominant land use type with a surface cover of at least 50 % defined a pixel's land use type. For single class estimations (e.g. bamboo), output pixels included another uncertainty of up to 50 %, given by the occurrence probability for each pixel.

For a classification of 14 forest cover types with the kNN method and a 10-point cluster plot design, Haapanen et al. (2002) defines the best estimation accuracies at pixel level, at 45 % to 55 % correctly classified. The application to forest/non-forest classification showed an accuracy of 86 % and corresponds in the main with the findings of this study. The kNN classification of forest/non-forest/water achieved overall accuracies ranging from 87% to 91% (Haapanen et al., 2004).

Holmgren & Thuresson (1998) reviewed several studies focusing on land classification. For a dozen or more forest classes ranging from regeneration areas to mature stands, classification accuracy was consistently in the range of 65-85%, regardless of the sensor used or local climatic conditions. Accuracy increased to above 90 %, if water bodies and land uses other than forestry were included. Comparing the results with other studies using the kNN method, Stümer (2004) calculated accuracies of 60 % to 73 % for the attribute deadwood.

The land cover classification of Dias (2003), featuring similar land conditions as the present study, with the seven land cover types, villages, forest, bush, fallow, cocoa, oil palm, and annual crops (including plantain, cassava, maize), showed an overall classification accuracy of about 72 %. This result is slightly better than the multi-classification of the present study, presumably due to the lower number and a more distinct selection of classes. Voado (2004) presented an assessment for a nine class land use map of the Goaso forest district and described the overall accuracy as being about 82 %. Using the software eCognition, Hailemariam (2004) classified similar areas of the Goaso forest district, defining seven land cover classes. They were annuals, cocoa with trees, fallow, grass, build-up, bare, and perennial without trees. The overall accuracy was estimated to be about 93 %. Compared to this study, the overall accuracy of the classifications of the Goaso forest district mentioned above seem to be significantly higher. This could be caused by the different selection of classes. Built-up areas, with a much defined reflection pattern, were not included in this work besides the fact that marshland, for instance, was subdivided into raphia patches, elephant grass, and bamboo. The different classification methodology might also cause the differences in the results.

The overall agreement of the classifications of this study, defined by the kappa coefficient, may be considered to show a moderate to substantial agreement. Kappa values of single land use type estimations ranged from 0.502 (banana) to 0.669 (oil palm). The overall agreement of the multi-attribute classification of the ten land use types showed a kappa coefficient of 0.496 and was improved to 0.663 when limiting the estimations to an occurrence probability of $p > 0.5$, instead of $p > 0.2$ per pixel. In comparison with other studies, using the kNN method for map making of forest cover

types, Gagliano et al. (2007) showed similar values with kappa values between 0.523 and 0.531.

7.4 Plot Distribution

With the defined spatial arrangement between training and test pixels in the categories separate (far, close), random, and equal distribution of the two collectives, different sampling designs, respectively, sample intensities are simulated. Theoretically, sample pixels of immediate vicinity tend to feature similar conditions in terms of soil conditions, topographic characteristics, treatment, etc. Whereas a further spatial distance from training pixel to test pixel (respectively, pixel that has to be estimated), tend to show greater site differences, that is reflections. With a gradient from far to close spatial distance of training and test pixels, the influence on the classification results is quantified.

For bamboo the classification accuracy correlates significantly with spatial distance of training and test collective. This leads to the assumption that the closeness of training and test pixel improves the accuracies in general. Similar results are shown for the precision of several test runs. A more equal distribution exhibits higher classification precision, compared to a higher spatial distance of training and test collective. However, lowest precision was found for a random distribution, which, on the contrary, is supposed to occupy a medium position. The majority of bamboo pixels came from sites that covered a multitude of bamboo clusters at one location, hence featuring very similar site conditions. The number of different bamboo locations was very limited, thus reflection values might show a low variability, which leads to a relatively low standard deviation of the repetitions and a high averaged classification accuracy of 85 % for the equal distribution, compared to 67 % for the separate far distribution.

In comparison, this substantial increase in classification accuracy for a more equal distribution was not observed for the average of all land use types. An increase of about three per cent points from separate to random distribution, but a decrease of

one per cent point to an equal distribution was observed. The number of only four runs for each land use type might not suffice. An earlier test run with a random distribution and ten sample runs showed a slightly lower classification accuracy, but still higher than the average of four runs and an equal distribution. The classification again exhibits a definite tendency: classification precision correlates with reduction of spatial distance of training and test pixels.

For an optimal sampling design, the crucial factor is not merely the equal or random distribution of sample plots over the area. Additionally, it is important that typical examples of the participating land use types are recorded to cover the variety of reflection values within the classes. Some land us types are distributed more or less equally over the whole area (e.g. cocoa plantation) and are easily encompassed by a random or systematic sampling design. Other land use types feature very large areas, but are limited to specific site conditions (e.g. rice field in lowlands). Such land use types can be located selectively, for example by using aerial photographs, where they can easily be identified, because of the large dimensions. Samples within such an area can then be distributed randomly. Other examples are the bamboos and raphia palms. In very rare cases, the raphia palms grow plentifully within marshes, whereas the bamboos prefer small meandering streams. Both types are also found widely distributed over the area, limited to only small numbers or even individuals at one location. The sampling design then has to be adapted to the particular land use type of interest.

Haapanen et al. (2002) described the effect on overall accuracy of forest cover type estimation by limiting the inclusion of subplots, comparing the tree categories using all subplots (59 - 71 %), excluding subplots closer than 40 metres (approx. 50 %), and excluding subplots from same cluster (33 - 44 %). These categories define a similar gradient, as the allocation in far, random, and equal distribution of this study. The findings of Haapanen et al. (2002) correspond to the results of this study: Since the subplots are located very close to each other, the reflection values for nearby pixels tend to be very similar within a cluster. In cover type classification, the entire subplot cluster was often of the same cover type and thus the nearest neighbours were usually found within the cluster. This impact has to be taken into consideration

for inventory-wide calculations and interpretations of overall accuracies in classifications.

7.5 Sample Size of Training Pixels

The software used for the kNN estimations of this study was limited to 999 training samples in one procedure. The number of input pixels per class – equal distribution assumed – correlates with the number of classes estimated in one procedure. By using only one attribute (e.g. bamboo) with two characteristic features (bamboo & non-bamboo), one characteristic feature can be supported by 499 input/training pixels. For a classification of ten land use types at one and the same time, only 100 training pixels per land use type, minus one all in all, can be handled. Therefore, the number of sample plots per land use type strictly depends on the number of attributes (classes) to be classified. Generally, for this study, training pixels are distributed equally over the involved attributes/classes.

A clear correlation of sample size and classification accuracy was found representatively for the attribute cocoa. Higher numbers of training pixels caused higher accuracies with an increasing flattening of the curve for higher sample sizes. Surprisingly, even the very low number of 50 sample plots (25 pixels cocoa, 25 non-cocoa) showed an averaged estimation accuracy of 69 % compared to 77.4 % with 998 sample pixels. The best precision were found for sample sizes above 750 pixels, whereas the lowest precision for the five runs for each category were found for the medium values of 400, 450, and 650 sample plots.

The pure maximisation of sample size is not sensible. Of greater importance is the representative distribution within the feature space. It has to be considered, depending on the selected k-nearest neighbours, that there is a sufficient number of samples available for the particular classes. Land uses with a very homogeneous feature, such as the cocoa plantations, require a lower sample size compared to banana plantations, which happen to appear in all kinds of mixtures with other land

uses. An optimisation of the sample design and sample size per class also reduces expenses for terrestrial data collection.

To obtain reliable estimates with the kNN estimation method, Nilsson (1997) claims that at least 500 sample plots are needed for Nordic forest conditions, which is applicable for Minnesota (Haapanen et al., 2002).

7.6 Parameters k, r, t

K is a parameter representing how many samples are considered to classify one object, whereas r and t enable the weighting of the spectral differences of the several bands. The kNN programme applied for this study already suggests values in the presetting (k = 5, r = 2, t = 2). Every land use type is most likely to respond differently to adjustments of these variables. Yu et al. (2006) found different classes achieved the highest classification accuracy at different k values. Precise specifications for each land use type are not given. This study appraised the constellation for bamboo, as a representative example, observing the parameter adjustments in a test series, with the two remaining parameters fixed at presetting.

The variable k, as the key parameter, is of particular interest. It is also particularly dependent on the sample size. Larger k values tend to favour larger sample sizes (Yu et al., 2006), whereas r and t tend to be more influenced by the reflection patterns of the specific land use type. A maximum classification accuracy was found for k = 4 (81.5 %), whereas lower, and particularly higher values, result in a decrease. The sample size was 408, which is the possible maximum for this land use type. Therefore, the optimum of this study is very close to the presetting of the software. With analyses as regards kNN methods for forest mapping, Haapanen et al. (2002) affirm that a value of k between 1 and 3 seems appropriate for mapping; larger number of neighbours reduces the overall estimation error, but it also leads to a reduction in the producer's accuracy. Stümer (2004) found that for the attribute dead wood, the best accordance between kNN estimations and reference samples were found for k-values between 1 and 7, which is consistent with the findings of this

study. One should note that when k = 2, the mode estimate is identical to that for k = 1. For a sample size of about 400 pixels, the best classification accuracy was approximated for k-values of about 7 to 9 (Yu et al., 2006). Tomppo (1996; 1997) has typically used k = 5.

The r-values enable a different weighting of individual bands. For bamboo, r-values of 1 to 3 showed the highest classification accuracies. With an increase of r, high spectral differences of individual channels get a higher weighting in the overall distance, which leads to a decrease in quality of the overall distance. In other words, pixels with relatively homogeneous spectral differences of the individual bands are favoured.

With the introduction of the variable t, control of the weighting of the selected k-nearest neighbour pixels and their allocated characteristic feature is enabled. With bigger values for t, higher weight is given to pixels with low spectral distance. The optimum for the t-values of bamboo is five.

With the weighting of spectral bands (k = 1), Haapanen et al. (2002) improved the cover type classification accuracy, whereas the volume RMSEs dropped with similar parameters. Determination of weights was not straightforward. Varying the distance decay parameter t had little influence (87.3 – 89.2 %) on overall accuracy (Haapanen et al., 2004) and showed a similar variability to that observed in this study.

The parameter findings for k, r, and t observed in this study can only serve as a first preliminary for applications of the kNN method of bamboo with a very limited number of spectral data. For other vegetation types, investigation areas and available remote sensing data, individual and more detailed sample calculations are absolutely recommended.

7.7 Assessment of the Classification Results

Image Classification

By comparison of multi-attribute classifications of the whole study area with randomly selected input pixels (999 out of 3360), the two resulting images displayed very similar distribution patterns for the ten land use types. The proportions of land cover showed a maximum deviation of 2 percentage points for the grassy vegetation, between the two images. The average deviation of all land covers was 0.76 %.

The classifications of single land resources with variable ratios of 'presence' and 'absence' input pixels had an impact on the estimated land cover proportion of these resources. Therefore, precise estimations of land use covers are not valid at this point. However, distribution patterns of specific land resources can be visualized at the present stage. For a monitoring system of specific land resources, prior assessment with respect to optimum input sample ratios between 'presence' and 'absence' of a land use type has to be undertaken. The same procedure is necessary for a multi-attribute classification. In all cases, it is recommended that the same attribute input ratios be applied for a monitoring with repeated classifications over time.

The land cover proportions estimated by the multi- and single-attribute classifications cannot be used as an indicator for the real proportions in the area at this moment. Indeed, the land cover proportions within a classification do not vary intensely in the case of a repeated investigation with another, randomly selected collective of training pixels. On the other hand, the chance of being selected as input pixel was determined by the proportion of samples for each land use type within the pool of the 3360 sample plots, which had been previously determined with a specific ratio. A detailed assessment with respect to adequate partitioning of samples for each land use type, or proportion for each class, has been omitted. Therefore, it is most likely that the estimated cover proportions of land use types, like bamboo and raphia palms, have been rather overestimated, particularly as these land uses could hardly be found with a random sampling design.

Confusion Matrix

The confusion matrix of the classification assessment means that an assessment of accuracy differences and relations amongst the attributes within a classification is possible. High proportions (10 to 28 %) of registered (ground truth) banana and bush fallow samples, for instance, were confused with each other and thus misclassified, hence similarities in reflection patterns are assumed. Similar proportions in misclassification (7 to 19 %) occurred between cocoa plantations and raphia palms.

7.8 Inventory of NTFPs and Tree/Forest Resources

The estimation and inventory of non-timber forest products with help of remotely sensed data strictly depend on the growth characteristics and distribution patterns of the particular resource. For this study, the kNN method has been applied to bamboo and raphia palms, whereas rattan had to be neglected due to inadequate size of the sample units. The assessment results exhibit that the distribution of bamboo and raphia palms can be estimated at a similar level of correctness as other land resources. However, bamboo clusters and particularly raphia palms also appear to occur as individual plants, not only in the form of clusters and groups. In these cases, estimation is rather inaccurate, and for individual raphia palms, almost impossible. Because mature raphia palms exhibit a relatively unique size and growth pattern, the number of palms correlates most likely with the covered area. Thus, inventory of raphia occurrences of a specific minimum size, indicates positive results for the application of the kNN method and quantities may be estimated based on the land cover. However, previous assessment of the correlation of surface cover and quantity have to be carried out for raphia, as well as bamboo. Similar results are indicated for forest patches and even individual giant trees. For a quantitative and more precise inventory of tree and forest resources outside the reserves, further studies are suggested. A separation of scattered trees within other land use types, and, on the other hand, forest patches, are advisable.

7.9 Recommendations for the Development of a Monitoring Design

The dramatic land use changes in tropical countries are becoming more and more of vital interest with increasing prices of crude oil and the adherent transformation from mixed farm and forest lands into treeless monocultures of oil palm plantations. Another big problem is the rapid deforestation caused by the continuing illegal and unsustainable practices of chainsaw operators (Kufuor, 2004). It is reported that pure terrestrial inventories and forest registers of Ghanaian off-reserve forest areas are already failing due to lack of funding. The development of an efficient remotely sensed monitoring system is recommended to detect and record land use changes over time.

The results of this study proved the applicability of the kNN method for a monitoring system of the land resources in the off-reserve forests in Ghana. The analyses give an overview of the potential estimation accuracies and overall classification agreement of several land use types. Observation made during terrestrial sampling and the analyses of the results provide preliminary indications and thus general guidelines and recommendations for the development of a monitoring system for the different land resources of the region. Optimisation of parameters and design of a monitoring system has to be developed individually with respect to objectives and land resources of interest.

In a first step, it has to be defined which land resources are actually of interest and what quality level is expected from the estimation. A basic rule is that an increasing number of classes leads to lower classification accuracies. Some land uses are more likely to have better accuracy results than others. Particularly land resources that show manifold mixtures of intercropping and a tendency to smaller units are less likely to be monitored and require higher sampling intensities. Individual optimum parameters have to be determined for the resources of interest and the objectives of the monitoring, respectively.

With respect to the small-scale mosaic of the different land uses, a pixel based classification with variable sample plots is advisable. Object based classification is

confronted with the problem of indistinct land use boundaries and intercropping. It is suggested that the minimum radius of the sample plots be 15 metres to increase the chance of geographically correct sampling. Indeed, most land use types, except fallow lands, show an average unit size of 0.3 ha (Hailemariam, 2004). The problem here is that scattered canopy trees are dispersed within most land use types and might have stronger influence on reflection, than the vegetation underneath. So it is advisable that one distinguishes between locations with and without shading timber trees. When intercropping cannot be registered by the image resolution, it should be defined and proportions recorded. Depending on the type of land use, proportions of other site conditions (swamp, proportion of bare soil, proportion of other vegetation, etc.) should be recorded.

The number of sample plots strongly depends on the homogeneity of the land use type, as well as upon cover proportion of the specific land resource. Particularly for multiple land resource monitoring, preliminary studies are essential. However, a minimum of 800 sample plots for an area of the Goaso forest district is advisable to increase precision of the estimation. If local information on resource distribution is available, stratification might help to reduce costs and to optimize sample size for specific land resources. Otherwise, a systematic, preferably equal spread of plots over the area, with fragmentation into subplots, is recommended. To enable a representativeness of the typical reflection conditions within one land use type, parts of the subplot locations could be selectively chosen. It is suggested that optimum ratios of sample units distributed over the different attributes/classes be determined. If only general distribution patterns of several land uses are of interest and precise proportions of land cover are of minor interest, an equal distribution of the samples distributed amongst the different classes is acceptable. However, for repeated estimations, it is strongly recommended that these ratios be kept. Depending on the objective of monitoring, a stabilisation of the estimation results could be achieved by resampling the distribution map into a 2 by 2 or 3 by 3 pixel matrix. Bamboo clusters, for example, happen to appear clumped. Single bamboo output pixels, correct or misclassified, could be ignored with this technique. Only agglomerations of bamboo output pixels would then be recognised.

GPS position accuracy plays a fundamental role. Correction signals to determine the GPS signal errors are not available in many regions. In tropical conditions of developing countries, using a small, cheap and user-friendly GPS receiver with medium position accuracy should be preferred to heavy equipment, assuming higher position accuracy. The advantages of an average function provided by the receiver should be fully utilised if possible. Within dense stands, facing poor GPS signals, it is worth recording and interpolating peripheral positions. Bigger circle plots reduce the uncertainty of position errors and reduce wrong registrations. The remotely sensed images applied should already be professionally geo-referenced when purchasing to reduce the accumulation of position errors. It is advisable to use cloud free satellite images of different dates, preferably complemented by radar data and a digital elevation model. Application of indices might help to considerably improve the classification results. The selection of specific band combination as well as the optimum values for the parameters k, r, and t should be determined by preliminary analyses.

8 Summary

Continuously increasing population and concomitant demand for food have been devastating forests by exploitation and shifting cultivation at an alarming rate. With the awareness the exhaustion of fossil fuels and recent price peaks for crude oil, the tendency towards forest destruction has been accelerated with the transformation of forest lands into oil palm plantations. In the same way, non-timber forest products like rattan, bamboo or raphia palms are disappearing. Local populations are more and more dependent on the illegal gathering of rattan from protected forest reserves and chainsaw lumbering practices in Ghana's off-reserve forests – prohibited since 1997 – serve as a source of livelihood for a good number of Ghanaians. The monitoring of forest lands is vital to enable the sustainable management and development of the area. The mission of the internationally acting organisation Tropenbos, where this study is imbedded, is to generate scientific input for sustainable management of natural resources of tropical countries.

The application of the k nearest neighbour (kNN) method in the combination of terrestrial data with remotely sensed data for forest attribute estimation and mapping has become an integral part of forest inventory methods. The object of this study is to assess the potential of the kNN method for the development of a cost efficient monitoring system of specific non-timber forest products, forest resources, and different land use types, in the off-reserve forests of the Goaso forest district in Ghana. For this purpose ten land use types and land resources, were identified, these being the following; bamboo, banana plantations, bush fallow, cocoa plantations, elephant grass, grassy vegetation, herbaceous vegetation, oil palm plantations, raphia palms, and forests. Based on selectively chosen sample circles with variable diameters, 3360 pixels were distributed within the off-reserve forests, covering an area of about 1710 km^2. For the classification assessment, the registrations were each divided respectively into a collective of training and test pixels. The method was applied to ASTER data and out of it mathematically generated indices were deduced to identify how various spectral band combinations, sample sizes, and sample distributions contribute to the overall accuracy of the various land use types.

The kNN classification achieved overall accuracies ranging from 75 % for plantain, 78 % for tree resources and raphia palms, 81 % for cocoa plantations and bamboo, up to 83 % for oil palm plantations. With increasing sample sizes or simulation of a particular sample plot distributions, the results could be improved. Likewise, selection of optimal parameters for the kNN programme settings were found. K-values between 3 and 6 showed the highest values for the land resource bamboo. For a multi-attribute classification of the ten land use types, overall accuracy and overall agreement in the classification could be increased by limiting the occurrence probability from the results of the kNN programme to at least 0.5. This, however, caused a waste of 42 % of the estimated pixels, which had to be labelled as unclassified. The partly broad differences of the kNN estimations, as regards accuracy and precision within the several land uses show, that optimisation of parameters ought to be identified individually, depending on the specific selection of resources which are to be monitored.

Finally, the results of the study demonstrate a high potential for the application of the kNN method for a monitoring system of the land resources of Ghana's off-reserve forests. Adaptation of individual parameters, for instance the value for k, optimisation of spectral data, sample size and design applied to specific land resources, will further increase the expected estimation results and classification accuracies of this study. At this point, the estimation and inventory of non-timber forest products, in particular raphia palms, is limited to habitats bearing a specific minimum ground cover and exhibiting an agglomeration of individuals to be captured by the satellite sensor. For detailed estimation and inventory of bamboo and raphia, the proportions of land cover can be used to estimate quantities of NTFP resources. However, further studies on the correlation of stock quantities versus habitat cover are highly suggested.

9 Zusammenfassung

Der ungebrochene Bevölkerungszuwachs und der damit verbundene Bedarf an Nahrung trägt, etwa durch Wanderfeldbau, unvermindert zur Waldzerstörung bei. Mit der zunehmenden Bewusstwerdung der Endlichkeit fossiler Energieträger und damit einhergehender Höchstpreise für Erdöl, nimmt der Trend der zunehmenden Entwaldung durch Abholzung und anschließender Umwandlung in Ölpalmplantagen noch zu. Zugleich verschwinden natürliche Vorkommen von Nicht-Holz Waldressourcen wie Rattan, Bambus oder Raphiapalmen. Die einheimische Bevölkerung sieht sich damit mehr und mehr dazu gezwungen, Rattan illegal aus Ghanas geschützten Waldreservaten zu entnehmen oder weiterhin eigenmächtig – und seit 1997 verboten – Holzeinschlag mit der Motorsäge in den Waldflächen außerhalb der Reservate vorzunehmen, was einer Großzahl von Ghanaern ermöglicht, ihren täglichen Lebensunterhalt zu bestreiten. Ein Monitoring dieser bewaldeten Flächen ist somit unabdingbar, um eine nachhaltige Nutzung und Landentwicklung dieser Gebiete zu ermöglichen. Die vornehmliche Aufgabe der international agierenden Organisation Tropenbos, in die diese Studie eingebunden ist, ist es, wissenschaftliche Forschung zu unterstützen, um damit Handlungsempfehlungen und Managementstrategien für die nachhaltige Nutzung natürlicher Ressourcen in tropischen Ländern bereitzustellen.

Die Anwendung der k-nächsten Nachbarn (kNN) Methode, in Kombination von terrestrisch erhobenen Daten mit Fernerkundungsdaten, zur Schätzung von forstlichen Kennwerten und zur Kartenerstellung ist mittlerweile ein integraler Bestandteil von Forstinventurmethoden. Das Ziel dieser Studie ist es, das Potential der kNN Methode für die Entwicklung eines kostengünstigen Monitoring Systems für bestimmte Nicht-Holz Waldprodukte, Waldressourcen und verschiedenen Landnutzungstypen, in nicht ausgewiesenen Reservaten des Forstdistrikts Goaso in Ghana, zu beurteilen. Dazu wurden die zehn Landnutzungstypen bzw. Landressourcen Bambus, Kochbanane, Buschbrache, Kakaoplantage, Elefantengras, Grasvegetation, Krautige Vegetation, Ölpalmplantage, Raphiapalme und Waldflächen identifiziert. Basierend auf selektiv ausgewählten Probekreisen mit variablem Durchmesser wurden 3360 Pixel außerhalb der Waldreservate im Goaso Forstdistrikt verteilt. Damit wurde eine Fläche von etwa 1710 km^2 abgedeckt. Für die

Genauigkeitsanalyse wurden die Aufnahmen in jeweils zwei verschiedene Kollektive aufgeteilt, in Trainings- und Testpixel. Die Methode wurde auf ASTER Satellitendaten und den daraus erzeugten Indizes angewandt, um dadurch den Einfluss einzelner spektraler Bänderkombinationen, des Stichprobenumfangs und der Probenverteilung, in Hinsicht auf die Gesamtgenauigkeit der Klassifizierung einzelner Landnutzungstypen, zu bestimmen.

Die kNN Klassifizierung erreichte Gesamtgenauigkeiten von 75 % für Plantagen von Kochbanane, 78 % für Baumressourcen und Raphiapalmen, 81 % für Kakaoplantagen und Bambusvorkommen und bis zu 83 % für Ölpalmplantagen. Mit zunehmendem Stichprobenumfang oder der Simulation von speziellen Probenverteilungen konnte die Genauigkeit der Schätzungen noch erhöht werden. Ebenso konnten die einstellbaren Parameter des kNN Programms optimiert werden. Für Bambus wurden die besten Ergebnisse für k-Werte zwischen 3 und 6 ermittelt. Für die Multi-Attribut-Klassifizierung der zehn Landnutzungstypen, konnte die Gesamtgenauigkeit und die Gesamtgültigkeit der Klassifizierung erhöht werden, wenn die Eintrittswahrscheinlichkeit von den Ergebnissen des kNN Programms auf mindestens 0,5 erhöht wurden. Dies führte jedoch gleichzeitig zu einer Verwerfung von 42 % der ausgegebenen Pixel, die dann als unklassifiziert angegeben werden mussten. Die teilweise beträchtlichen Unterschiede der Schätzungen in Bezug auf Genauigkeit und Präzision der kNN Schätzungen zeigen, dass eine Optimierung der einzelnen Parameter individuell ermittelt werden sollte, je nachdem für welche Ressourcen ein Monitoring erfolgen soll.

Abschließend läßt sich ein hohes Potential der kNN Methode für die Anwendung in einem Monitoring System der Landressourcen Ghanas außerhalb der Waldreservate prognostizieren. Unter anderem kann durch individuelle Anpassung der Parameter etwa von k und Optimierung spektraler Daten, des Stichprobenumfangs und Aufnahmedesigns für individuelle Landressourcen, die Schätz- und Klassifizierungsgenauigkeiten aus dieser Studie noch weiter erhöht werden. Zum gegenwärtigen Zeitpunkt ist eine Inventur und Schätzung von Nicht-Holz Waldprodukten, insbesondere von Raphiapalmen, auf eine Mindestgröße des Habitats und einer Anhäufung von Einzelindividuen begrenzt, um vom

Satellitensensor erfasst werden zu können. Für eine detaillierte Schätzung und Inventur von Bambus und Raphiapalmen sind weitere Studien bezüglich des Zusammenhangs der eingenommenen Fläche und der Vorratsmenge empfehlenswert, um damit vom Flächenanteil auf die Quantität einzelner Ressourcen schließen zu können.

10 References

Abagale, F. K., Addo, J., Adisenu-Doe, R., Anthony, K. M., Apana, S., Boateng, E. A., Owusu, A. N., & Parahoe, M., 2003: The potential and constraints of agroforestry in forest fringe communities of the Asunafo District - Ghana. Tropenbos International, Wageningen, The Netherlands, 50 p.

Abbiw, D., Agbovie, T., Akuetteh, B., Amponsah, K., Dennis, F., Ekpe, P., Gillett, H., Ofosuhene-Djan, W., & Owusu-Afriyie, G., 2002: Conservation and sustainable use of medicinal plants in Ghana. http://www.unep-wcmc.org/species/plants/Ghana Accessed 08/15/2007

A.D.A, 2002: District medium term development plan: 2002-2003. Goaso: Asunafo District Assembly, Brong Ahafo region, Ghana.

Affum-Baffoe, K., 2001: Modified procedures for tree resource assessment in off-reserve areas of Ghana high forest zone. Ghana Forestry Commission. Paper presented at IUFRO 4.11 Conference on Forest Biometry, Modeling and Information Science. June 26-29, 2001, Greenwich, England.

Albertz, J., 1991: Grundlagen der Interpretation von Luft- und Satellitenbildern: Eine Einführung in die Fernerkundung. Wiss. Buchgesellschaft, Darmstadt, 93-153.

Alo, A. C. & Pontius, R. G., 2006: Identifying systematic land cover transitions using remote sensing and GIS: The fate of forests inside and outside protected areas of Southwestern Ghana. Department of International Development, Community and Environment. Graduate School of Geography, Clark University, Worcester MA.

Altmann, D. G., 1991: Practical statistics for medical research. London, 611 p.

Altmann, N. S., 1992: An introduction to kernel and nearest neighbour nonparametric regression. Am. Stat. 46: 175-184.

Amanor, K. S., 2003: Natural and Cultural Assets and Participatory Forest Management in West Africa. In: Conference Papers Series No. 8. International Conference on Natural Assets, Political Economy Research Institute and Centre for Scientific Environment, Philippines.

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E., 1976: A land use and land cover classification system for use with remote sensor data. Geological Survey Professional Paper 964. A revision of the land use classification system as presented in U.S. Geological Survey Circular 671. United States Government Printing Office, Washington, 41 p.

Anttila, P., 2002: Nonparametric estimation of stand volume using spectral features of aerial photographs. Can. J. For. Res. 32: 1849-1857.

Asamoah-Boateng, B., 2003: Distribution and diversity of tree-resources outside forest (TROF) in Southern Ghana. Thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

ASTER GDS, 2007: ASTER GDS Web Site, Earth Remote Sensing Data Analysis Center. http://www.gds.aster.ersdac.or.jp/gds_www2002/index_e.html Accessed 08/22/2007

Atkeson, C. G., Moore, A. W., & Schaal, S., 1997: Locally weighted learning. Artificial Intelligence Review, 11(1-5): 11-73.

Bailey, T. & Jain, A., 1978: A note on distance-weighted k-nearest neighbor rules. IEEE Trans. Systems, Man, Cybernetics 8: 311-313.

Ball, D. W., 1995: Defining terms. Spectroscopy 10: 16-18.

Banko, G., 1998: A review of assessing the accuracy of classifications of remotely sensed data and of methods including remote sensing data in forest inventory. International Institute for Applied Systems Analysis (IIASA). Interim report IR-98-081, Laxenburg, Austria.

Barnsley, M. J., Moller-Jensen, L., & Barr, S. L., 2000: Inferring urban land use by spatial and structural pattern recognition. In: Donnay. J. P., Barnsley, M. J., & Longley, P. A., (eds.). Remote sensing and urban analysis. Taylor and Francis, London, 115-144.

Becker, B. L., Lusch, D. P., & Qi, J., 2005: Identifying optimal spectral bands from in situ measurements of Great Lakes coastal wetlands using second-derivative analysis. Remote Sensing of Environment 97(3): 238-248.

Benneh G. & Agyapong, G. T., 1990: Land degradation in Ghana. London: Commonwealth Secretariat and Legon: Department of Geography and Resource.

Bih, F., 2006: Assessment methods for non-timber forest products in off-reserve forests. Case study of Goaso district, Ghana. Dissertation, Fakultät für Forst- und Umweltwissenschaften, Albert-Ludwigs-Universität Freiburg, 140 p.

Bishop, Y., Fienberg, S., & Holland, P., 1975: Discrete multivariate analysis - theory and practice. MIT Press, Cambridge, MA. Development, University of Ghana. In: BIRD, KNUST (2001). Research Sites Community Survey, Final Report-TGP, 33 p.

Bortz, J., 1993: Statistik für Sozialwissenschaftler. 4. Auflage. Springer, Berlin.

Cabaravdic, A., 2007: Efficient Estimation of Forest Attributes with k NN. Dissertation, Fakultät für Forst- und Umweltwissenschaften, Albert-Ludwigs-Universität Freiburg, 129 p.

Campbell, J. B., 2002: Introduction to remote sensing. 3rd ed. The Guilford Press, New York.

Clark, R. N., 1999: Spectroscopy of rocks and minerals and principles of spectroscopy. In: Rencz, A. N. (ed.). Remote sensing for earth sciences. Manual of remote sensing, 3rd ed. New York, 3: 3-58.

Cochran, W. G., 1977: Sampling Techniques. John Wiley & Sons, New York, 3rd ed., 428 p.

Cohen, J., 1960: A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1): 37-40.

Congalton, R., 1991: A review of assessing the accuracy of remotely sensed data. Remote Sensing of Environment 37: 35-46.

Congalton, R. & Green, K., 1999: Assessing the accuracy of remotely sensed data. Principles and practices. Lewis Publishers, Boca Raton, Florida, 137 p.

Congalton, R. & Mead, R., 1983: A quantitative method to test for consistency and correctness of photointerpretation. Photogrammetric Engineering and Remote Sensing, 49(1): 69-74.

Congalton, R., Oderwald, R., & Mead, R., 1983: Assessing Landsat classification accuracy using discrete multivariate statistical techniques. Photogrammetric Engineering and Remote Sensing, 49(12): 1671-1678.

Cover, T. M. & Hart, P. E., 1967: Nearest neighbour pattern classification. IEEE Transactions on Information Theory, IT-13(1): 21-27.

de Bie, C. A. J. M., 2000: Comparative performance of analysis of agro-ecosystems. Thesis, Wageningen University, The Netherlands. Universal Press, Veenendaal, 232 p.

de Vries, A. P., Mamoulis, N., Nes, N., & Kersten, M., 2002: Efficient k-NN search on vertically decomposed data. ACM SIGMOD 2002, Madison, Wisconsin.

Dias, H. U., 2003: Analysis of the spatial distribution of tree resources outside the forests in Ashanti region, Ghana. Thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands, 73 p.

Duch, W. & Grudzinski, K., 1999: The weighted k-NN with selection of features and its neural realization. 4th Conference on Neutral Networks and their Applications, May 1999, Zakopane.
http://www.fizyka.umk.pl/publications/kmk/99scal-knn.html
Accessed 08/08/2006

ERDAS IMAGINE, 1997 : V 8.3 Tour Guides. ERDAS, Inc., Atlanta, Georgia, 454 p.

ERDAS IMAGINE, 1997: V 8.3 Field Guide, Fourth Edition, ERDAS, Inc. Atlanta, Georgia, 656 p.

Dudani, S. A., 1976: The distance-weighted k-nearest-neighbor rule. IEEE Trans. Syst., Man, Cybern. 6: 325–327.

Downs, G. M. & Barnard, J. M., 2002: Clustering of very large datasets. Sheffield, UK BCI Barnard Chemical Information Ltd.

FAO, 1976: Management and utilization of the tropical moist forest - from the FAO Committee on forest development in the tropics - extracts. Pasca, T. M. (ed.). Unasylva - No. 112-113, Vol. 28.

FAO, 2003. State of the World's Forests. Rome.

Fehrmann, L., 2006: Alternative Methoden zur Biomasseschätzung auf Einzelbaumebene unter spezieller Berücksichtigung der k-nearest neighbour (k-NN) Methode. Dissertation, Fakultät für Forstwissenschaften und Waldökologie, Georg-August-Universität Göttingen.

Fenstermaker, L., 1991: A proposed approach for national to global scale error assessments. Proceedings GIS/LIS '91, ASPRS, ACSM, AAG, AM/FM International and URISA, 1: 293-300.

Fisher, P. F., Comber, A., & Wadsworth, R., 2005: Land Use and Land Cover: Contradiction or Complement. In: Fisher, P. F. & Unwin, D. J. (eds.). Representing GIS. John Wiley, London, 85-98.

Fix, E. & Hodges, J. L., 1951: Discriminatory analysis, nonparametric discrimination, consistency properties. Technical Report 4, United States Air Force, School of Aviation Medicine. Randolph Field, Texas,

Franco-Lopez, H., Ek, A. R., & Bauer, M. E., 2001: Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment 77: 251-274.

Gagliano, C., De Natale, F., Incerti, F., & Maselli, F., 2007: Alternative application of the k-NN method for mapping forest cover type. In: Schaepman, M. E., Liang, D., Groot, N. E., & Kneubühler, M. (eds.). 10th Intl. Symposium on Physical Measurements and Spectral Signatures in Remote Sensing. Intl. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI. Davos, Switzerland.

Garmin, 1997: Garmin GPS II plus, Owner´s Manual & Reference. Garmin Corporation, Olathe, Kansas, 100 p.

Goetz, A. F. H., 1992: Imaging spectrometry for earth remote sensing. In: Toselli, F. & Bodechtel J. (eds.). Imaging Spectroscopy. Fundamentals and Prospective Applications, 1-20.

Goetz, A. F. H., Solomon, G. V., & Rock, B. N., 1985: Imaging spectrometry for earth remote sensing. Science 228: 1147-1153.

Gustafsson, K., 2002: Demonstration of methods to monitor sustainable forestry. EU/LIFE project 1998-2001, Rapport 8. Skogsstyrelsen, Jönköping.

Haapanen, R., Ek, A. R., Bauer, M. E., & Finley, A. O., 2004: Delineation of forest/nonforest land use classes using nearest neighbor methods. Remote Sensing of Environment 89(3): 265–271.

Haapanen, R., Lehtinen, K., Miettinen, J., Bauer, M. E., & Ek, A. R., 2002: Progress in adapting k-NN methods for forest mapping and estimation using the new annual forest inventory and analysis data. In: McRoberts, R. E., Reams, G. A., Van Deusen, P. C., & Moser, J. W. (eds.). Proceedings of the 3rd Annual Forest Inventory and Analysis Symposium. St. Paul, MN. U.S. Department of Agriculture, Forest Service, North Central Research Station, 87-95.

Haara, A., Malamo, M., & Tokola, T., 1997: The k-nearest neighbour method for estimating basal-area distribution. Scan. J. For. Res. 12: 200-208.

Haendel, L., 2003: Clusterverfahren zur datenbasierten Generierung interpretierbarer Regeln unter Verwendung lokaler Entscheidungskriterien. Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Universität Dortmund, 120 p.

Hailemariam, S. N., 2004: Assessment of trees outside forest using object-oriented classification and medium resolution satellite imagery. A case study in Brong Ahafo region, Ghana.

Hall, J. B. & Swaine, M. D., 1976: Classification and ecology of closed-canopy forest in Ghana. Department of Botany, University of Ghana, Legon. The Journal of Ecology 64(3): 913-951.

Hall, J. B. & Swaine, M. D., 1981: Distribution and ecology of vascular plants in a tropical rainforest. Forest Vegetation in Ghana. 25 p.

Häussler, T., Köhl, M., Lautner, M., Scheuber, M., Stümer, W., & Ziese, H., 2000: Teilvorhaben HEZEW. In: ProSmart Endbericht, Daimler Chrysler Aerospace, Dornier Satellitensysteme, Friedrichshafen: 9/1 - 9/119.

Hawthorne, W. D. & Abu-Juam, M., 1993: Forest protection in Ghana. Report of the ODA Forest Inventory and Management Project, Kumasi, Ghana.

Hessenmöller, D., 2001: Modelle zur Wachstums- und Durchforstungssimulation im Göttinger Kalkbuchenwald. Dissertation, Fakultät für Forstwissenschaften und Waldökologie, Georg-August Universität Göttingen. Logos Verlag Berlin, 163 p.

Hessenmöller, D. & Elsenhans, A. S., 2002: Zur Schätzung des Zuwachses bei Rotbuche Fagus sylvatica L. Ein Vergleich parametrischer Verfahren mit der k-nearest neighbour Methode. Allg. Forst- u. J.-Ztg. 173(11/12): 216-223.

Hildebrandt, G., 1996: Fernerkundung und Luftbildmessung. Wichmann Verlag, Heidelberg, 676 p.

Holm, S., Hägglund, B., & Mårtenson, A., 1997: A method for generalization of sample tree data from the Swedish National Forest Survey. Swedish University of Agricultural Sciences. Department of Forest Survey. Report No. 25, 94 p.

Holmgren, J., Joyce, S., Nilsson, M., & Olsson, H., 2000: Estimating stem volume and basal area in forest compartments by combining satellite image data with field data. Scan. J. For. Res. 15: 103-111.

Holmgren, P. & Thuresson, T., 1998: Satellite remote sensing for forestry planning: A review. Scan. J. For. Res. 13(1): 90- 110.

Holmström, H., 2001: Data acquisition for forestry planning by remote sensing based sample plot imputation. Doctoral Thesis, Swedish University of Agricultural Sciences, Umeå, 41 p.

Holmström, H., Nilsson, & M., Ståhl, G., 2001: Simultaneous estimations of forest parameters using aerial photograph interpreted data and the k-nearest neighbour method. Scan. J. For. Res. 16(1): 67-78.

Jensen, J. R., 2000: Remote sensing of the environment: an earth resource perspective. Prentice-Hall.

Jozwik, A., 1983: A learning scheme for a fuzzy k-nn rule. Pattern Recognition Letters 1: 287–289.

Kellenberger, T. W., 1996: Erfassung der Waldfläche in der Schweiz mit multispektralen Satellitenbilddaten – Grundlagen, Methodenentwicklung und Anwendung. Remote Sensing Series, Universität Zürich.

Keller, J. M., Gray, M. R., & Givens, J. A., 1985: A fuzzy k-NN neighbor algorithm. IEEE Trans. Syst., Man, Cybern. 15(4): 580–585.

Kerwin, T., 2006: Classification of natural language based on character frequency. http://www.cse.ohio-state.edu/~kerwin/language_classification.pdf Accessed 07/20/2007

Khunrattanasiri, W., 2006: Development of forest inventory techniques with remote sensing for forest resources assessment. Dissertation, Fakultät für Forst- und Umweltwissenschaften, Albert-Ludwigs-Universität Freiburg, 129 p.

Kilkki, P. & Päivinen, R., 1987: Reference sample plots to compare field measurements and satellite data in forest inventory. Proceedings of Seminar on Remote Sensing-aided Forest Inventory, Finnland 10-12 December 1986. University of Helsinki, Department of Forest Mensuration and Management, 209-215.

Kohavi, R. & Provost, F., 1998: Glossary of terms. Editorial for the special issue on applications of machine learning and the knowledge discovery process. Machine Learning, 30: 271–274.

Köhl, M. & Lautner, M., 2001: Erfassung von Waldökosystemen durch Hyperspektraldaten. Photogrammetrie Fernerkundung Geoinformation 2: 107-117.

Köhl, M., Stümer, W., Scheuber, M., Ziese, H., Häussler, T., & Lautner, M., 2000: Combining hyperspectral data and ground surveys for the assessment of non-wood goods and services of forests. In: A Decade of Trans-European Remote Sensing Cooperation (Buchroithner, M. F.), Proceedings of the 20th EARSel Symposium, Balkema Publishers, Tokyo.

Komagata, N., 2002: Chance Agreement and Significance of the kappa Statistic. http://www.tcnj.edu/~komagata/pub/Kappa.pdf Accessed 05/20/2004

Korhonen, K. T. & Kangas, A., 1997: Application of nearest-neighbour regression for generalizing sample tree information. Scan. J. For. Res. 12: 97-101.

Kotey, E. N. A., Francois, J., Owusu, J. G. K., Yeboah, R., Amanor, K. S., & Antwi, L., 1998: Falling into place. Policy that works for forests and people series No. 4. International Institute for Environment and Development, London.

Kraus, K. & Schneider, W. 1988: Fernerkundung. Band 1. Physikalische Grundlagen und Aufnahmetechniken, Bonn.

Kufuor, K. O., 2004: New institutional economics and the failure of sustainable forestry in Ghana. Nat. Res. J. 44: 743-760.

Kulbach, D., 1997: Systematischer Vergleich von Verfahren zur überwachten Klassifikation von multispektralen Luftbildern mit Bodenauflösungen im Meterbereich. Diplomarbeit, Fachbereich Physik, Universität Hamburg, 139 p.

Lacaze, B., 1996: Spectral characterisation of vegetation communities and practical approaches to vegetation cover changes monitoring. In: Hill, J. & Peter, D. (eds.). The use of remote sensing for land degradation and desertification monitoring in the mediterranean basin - Proceedings of an Expert Workshop, 13 - 15 Juni 1994, Valencia, 149 - 165.

Landis, J. R. & Koch, G. G., 1977: The measurement of observer agreement for categorical data. Biometrics 33(1): 159-174.

Lautner, M., 2001: Hyperspektrale Fernerkundungsdaten zur Charakterisierung von Waldbeständen. Dissertation, Lehrstuhl für Biometrie und Forstliche Informatik, Technische Universität Dresden, 116 p.

Lee, K. S., Kim, S. H., Park, J. H., Kim, T. G., Park, Y. I., & Woo, C. S., 2006: Remote sensing estimation of forest LAI in close canopy situation. Kor. J. Rem. Sen. 22(5): 305-311.

LeMay, V. & Temesgen, H., 2004: Comparison of nearest neighbour methods for estimating basal area and stems per hectare using aerial auxiliary variables. For. Sci. 51(2): 109-119.

Lemm, R., Vogel, M., Felber, A., & Thees, O., 2005: Eignung der k-nearest neighbour (kNN-) Methode zur Schätzung von Produktivitäten in der Holzernte – Grundsätzliche Überlegungen und erste Erfahrungen. Allg. Forst- u. J.-Ztg. 176: 189-200.

Löffler, E, 1994: Geographie und Fernerkundung. Eine Einführung in die geographische Interpretation von Luftbildern und modernen Fernerkundungsdaten. 2. Auflage, Stuttgart, 251 p.

Logicon Geodynamics, 1997: Multispectral Imagery Reference Guide, Fairfax: Logicon Geodynamics, 101 p.

Malinen, J., 2003a: Prediction of characteristics of marked stand and metrics for similarity of log distribution for wood procurement management. Dissertation, Faculty of Forestry, University of Johensuu, Finland.

Malinen, J., 2003b: Locally adaptable non-parametric methods for estimating stand characteristics for wood procurement planning. Silva Fennica, 37(1): 109-118.

Malinen, J., Maltamo, M., & Verkasalo, E., 2003: Predicting the internal quality and value of Norway spruce trees by using two non-parametric nearest neighbour methods. Forest Products Journal, 53(4): 85-94.

Maltamo, M. & Kangas, A., 1998: Methods based on k-nearest neighbour regression in the prediction of basal area diameter distribution. Can. J. For. Res. 28: 1107-1115.

Mayers, J., Howard, C., Kotey, E. N. A., Prah, E., & Richards, M., 1996: Incentives for sustainable forest management: A study in Ghana. IIED Forestry and land use series No. 6. International Institute for Environment and Development, London.

McRoberts, R., Nelson, M. D., & Wendt, D. G., 2002: Stratified estimation of forest area using satellite imagery, inventory data, and the k-nearest neighbour technique. Remote Sensing of Environment 82: 457-468.

Moer, M., 1987: Nearest neighbour inference for correlated multivariate attributes. Proceedings of IUFRO Conference on Forest Growth Modelling and Prediction, Minneapolis, 23-27 August 1987. USDA Forest Service, General Technical Report NC-120 (Saint Paul, MN, US Department of Agriculture, Forest Service, North Central Forest Experiment Station), 716-723.

Moer, M. & Hershey, R. R., 1999: Preserving spatial and attribute correlation in the interpolation of forest inventory data. In: Spatial accuracy assessment. Lowell, K., Jaton, A. (eds). Land information uncertainty in natural resources. Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Quebec City, Canada, 20-22 Mai 1998. Ann Arbor Press, Chelsa, Michigan, 419-430.

Moer, M. & Stage, A. R., 1995: Most similar neighbour: an improved sampling inference procedure for natural resource planning, For. Sci. 41: 337-359.

Muinonen, E., & Tokola, T., 1990: An application of remote sensing for communal forest inventory. The usability of remote sensing for forest inventory and planning. Proceedings from SNS/IUFRO workshop. Umeå, Sweden: Remote Sensing Laboratory, Swedish University of Agricultural Sciences (Report 4), 35-42.

Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L., 1995: The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience and Remote Sensing 33: 481-486.

Nakov, P. & Dobrikov, P., 2004: Non-parametric SPAM filtering based on kNN and LSA. In Proceedings of the 33th National Spring Conference of the Bulgarian Mathematicians Union, 1 - 4 April 2004, Borovets, Bulgaria, 231-235.

NASA, 2004: Jet Propulsion Laboratory. California Institute of Technology. ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer. http://asterweb.jpl.nasa.gov/index.asp Accessed 09/07/2005

NASA, 2007: National Aeronautics and Space Administration. Terra - The EOS Flagship. http://terra.nasa.gov Accessed 06/15/2007

Ngamabou, R. S., 2006: Evaluating the efficacy of remote sensing techniques in monitoring forest cover and forest cover change in the Mount Cameroon region. Dissertation, Fakultät für Forst- und Umweltwissenschaft, Albert-Ludwigs-Universität Freiburg.

Niemann, H., 1983: Klassifikation von Mustern. Springer, Berlin.

Nieschulze, J., Böckmann, T. H., Nagel, J., & Saborowski, J., 2005: Herleitung von einzelbestandesweisen Informationen aus Betriebsinventuren für die Zwecke der Forsteinrichtung. Allg. Forst- u. J.-Ztg. 176: 169-176.

Niggemeyer, P., 1999: Schätzung von Durchmesserverteilungen mit der k-nearest neighbour Methode. Diplomarbeit, Fakultät für Forstwissenschaften und Waldökologie, Georg-August-Universität Göttingen, 71 p.

Niggemeyer, P. & Schmidt, M., 1999: Estimation of the diameter distributions using the k-nearest neighbour method. In: Pukkala, T. & Eerikäinen, K. (eds). Growth and yield modelling of tree plantations in South and East Africa. University of Johensuu, Faculty of Forestry. Research Notes 97: 195-209.

Nilsson, M., 1997: Estimation of forest variables using satellite image data and airborne lidar. Dissertation, Acta Universitatis Agriculturae Suecicae, Silvestria 17.

Oehmichen, K., 1997: Satellitengestützte Waldflächenkartierung für die Bundeswaldinventur. Dissertation, Fakultät für Mathematik, Informatik und Naturwissenschaften, Universität Hamburg.

Ortiz, M. J., Formaggio, A. R., & Epiphanio, J. C. N., 1997: Classification of croplands through integration of remote sensing, GIS and historical database. International Journal of Remote Sensing 18(1): 95-105.

Owubah, C. E., Le Master, D. C., Bowker, J. M., & Lee, J. G., 2001: Forest tenure systems and sustainable forest management: The case of Ghana. Forest Ecology and Management 149: 253-264.

Qi, H., 2002: Feature selection and kNN fusion in molecular classification of multiple tumor types. Electrical and Computer Engineering Department, University of Tennessee.

QWOD, 2001: Agricultural Systems in Ghana. QWOD Group, London, United Kingdom. http://whitesys.net/qwod_farms/agricsysgha.htm
Accessed 08/21/2006

Rajaniemi, S., Tomppo, E. , Ruokolainen, K., & Tuomisto, H., 2003: Estimating number of pteridophyte and Melastomataceae species from satellite images in Western Amazonian rain forest. In: Advances in forest inventory for sustainable forest management and biodiversity monitoring. Forestry Sciences. Kluwer Academic Publishers,76: 57-64.

Ratanopad, S. & Kainz, W., 2006: Land cover classification and monitoring in Northeast Thailand using Landsat 5 TM data. ISPRS Technical Commission II Symposium, 12 - 14 July 2006, Vienna.

Richards, J. A. & Xiuping, J., 1999: Remote sensing digital image analysis. An Introduction. 3rd ed. Springer, Berlin, 363 p.

Richards, J. A. & Xiuping, J., 2005: Remote sensing digital image analysis. An Introduction. 4th ed. Springer, Berlin, 439 p.

Rosenfield, G. & Fitzpatrick-Lins, K., 1986: A coefficient of agreement as a measure of thematic classification accuracy. Photogrammetric Engineering and Remote Sensing, 52(2): 223-227.

Rouw, de, A., 1991: The invasion of chromolaena odorata L. King & Robinson (ex Eupatorium odoratum), and competition with the native flora, in a rain forest zone, South-West Cote d'Ivoire. Journal of Biogeography 18(1): 13-23.

SBRC, 1994: Space Sensors. Goleta, CA: Santa Barbara Research Center, 33 p.

Sellers, P. J., 1985: Canopy reflectance, photosynthesis, and transpiration. International Journal of Remote Sensing 6: 1335-1372.

Sironen, S., Kangas, A., Maltamo, M., & Kangas, J., 2001: Estimating individual tree growth with the k-nearest neighbour and k-most similar neighbour methods. Silva Fennica 35(4): 453-467.

Sohlberg, S. & Sokolov, V. E. (eds.), 1986: Practical application of remote sensing. Proceedings of a Seminar on the Practical Application of Remote Sensing in Forestry (Jönköping), 1985. Martinus Nijhoff Publishers, Boston, 228 p.

Stümer, W., 2004: Kombination von terrestrischen Aufnahmen und Fernerkundungsdaten mit Hilfe der kNN methode zur Klassifizierung und Kartierung von Wäldern. Dissertation, Fakultät für Forst-, Geo- und Hydrowissenschaften, Technischen Universität Dresden, 153 p.

Temesgen, H., 2003: Estimating stand tables from aerial attributes: A comparison of parametric and most similar neighbour methods. Scan. J. For. Res. 18(3): 1-10.

Tokola, T., Pitkänen, J., Partinen, S., & Muinonen, E.,1996: Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. International Journal of Remote Sensing 17(12): 2333- 2351.

Tommola, M., Tynkkyen, M., Lemmetty, J. Harstela, P., & Sikanen, L., 1999: Estimating the characteristics of a marked stand using k-nearest-neighbour regression. Journal of Forest Engineering 10(2): 75-81.

Tomppo, E., 1991: Satellite imagery-based national inventory of Finland. International Archives of Photogrammetry and Remote Sensing 28(7-1): 419-424.

Tomppo, E., 1996: Multi-source national forest inventory of Finland. In: Päivinen, R., Vanclay, J. & Miina, S. (eds.). New Thrusts in Forest Inventory. Proceedings of the Subject Group S4.02-00 'Forest Resource Inventory and Monitoring' and Subject Group S4.12-00 'Remote Sensing Technology'. Volume I. IUFRO XX World Congress, 6-12 August 1995, Tampere, Finland. EFI, EFI Proceedings 7: 27-41.

Tomppo, E., 1997: Application of remote sensing in Finnish national forest inventory. In: Kennedy, P. J. Application of remote sensing in European forest monitoring. International Workshop, Vienna, Austria, 14th - 16th October 1996. Proceedings. European Commission, 375-388.

Tomppo, E., 2005: The Finnish multi-source national forest inventory. Small area estimation and map production. Kangas & Maltamo (eds.). Forest inventory. Methodology and applications. Springer, Netherlands.

Tomppo, E. & Pekkarinen, A., 1997: Methodenerprobung der finnischen nationalen Multiquellen-Waldinventur in Nordrhein-Westfalen. In: Landeswaldinventur - Konzeption, Verfahrenstest, Ergebnisse, Schriftenreihe der Landesforstverwaltung Nordrhein-Westfalen 5: 52-67

Tropenbos, 2007: Tropenbos International Ghana Programme. http://www.tropenbos.nl/sites/site_ghana.php Accessed 07/24/2007

Tuncel, E. & Rose, K., 2002: Towards optimal data clustering for approximate similarity searching. In: IEEE International Conference on Multimedia and Expo. August 2002, Lausanne, Switzerland.

Van Leeuwen, L., 2006: Off-forest reserve tree management. Geo-information applications for off-reserve tree management in Goaso district, Ghana (GORTMAN). International Institute for Geo-Information Science and Earth Observation. http://www.itc.nl/research/policy/spearhead4/vanleeuwen.asp Accessed 09-22-2006

Voado, G. C., 2004: Regeneration of trees in a tropical forest-agro ecosystem. A case study in Goaso forest district of Ghana. Thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands, 73 p.

Wallermann, J., 2003: Remote sensing aided spatial prediction of forest stem volume. Doctoral Thesis, Swedish University of Agricultural Sciences, Umeå, 100 p.

Warfield, S., 1996: Fast k-NN classification for multichannel image data. Pattern Recognition Letters 17: 713-721.

Wills, J. B., 1962: Agriculture and land use in Ghana. Oxford Universtiy Press, London.

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., & Schirokauer, D., 2006: Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. International Archives of Photogrammetry and Remote Sensing 72(7): 799-812.