



Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

Master Thesis

Vegetation Mapping in the Piedras Blancas National Park in Costa Rica Using Satellite Images

aspired academic degree
Master of Mountain Forestry

Author:	Yin-Jung Liu
Student ID-Number:	0656454
Study /Major	Mountain Forestry
Supervisor:	Univ.Prof.i.R. Dipl.-Ing. Dr.techn. Werner SCHNEIDER
Sub-Supervisor:	Dipl.-Ing. Dr.nat.techn. Tatjana KOUKAL

Vienna, Aug. 2010

Acknowledgements

I would like to thank especially my supervisor professor Schneider and sub-supervisor Tatjana for their kindly support. Without them this thesis would not be done. They gave me not only technical support and valuable experience, but also the mental comfort. I also want to thank Matteo Mattiuzzi, Franz Suppan, and Thomas Bauer for their kindly technical support. Thanks for the botanic information of local vegetation composition from Anton Weissenhofer, who provided useful information assisting this study.

Table of Contents

1	INTRODUCTION	1
2	STUDY AREA	4
2.1	Costa Rica	4
2.2	Piedras Blancas National Park.....	9
3	DATA	10
3.1	Landsat TM Images	10
3.2	DEM from Shuttle Radar Topography Mission (SRTM)	11
3.3	Piedras Blancas National Park Vegetation Map	11
3.4	Google Earth	12
4	METHODS	13
4.1	Processing Sequence	13
4.2	Pre-processing.....	14
4.2.1	Geometric Correction.....	14
4.2.2	Radiometric-Topographic Normalization	15
4.3	<i>k</i> -Nearest-Neighbour (kNN) Program.....	18
4.3.1	Methodical Background.....	18
4.3.2	Comparison of Advantages and Shortcomings	21
4.3.3	Components	22
4.3.4	Process Sequence	25
4.4	Classification Enhancements	29
4.5	Post-processing	31
4.6	Pixel Change Detection.....	31
5	RESULTS	32
5.1	Crossvalidation	32
5.1.1	Pre-classification with 12 Classes.....	32
5.1.2	Classification with 7 Classes.....	33
5.2	Classified Vegetation Map.....	34
5.2.1	Piedras Blancas National Park	34
5.2.2	Extensive Study Area.....	36
6	DISCUSSION.....	42
6.1	Sources of Errors.....	42
6.2	Possibilities of Improvement	43
6.3	Suggestions of Wildlife Corridor.....	45
7	ABSTRACT	46
8	REFERENCES	47
9	APPENDICES	50
10	TABLE OF FIGURES.....	55
12	TABLE OF TABLES	56

1 INTRODUCTION

Tropical forests are well-known carbon reservoirs and also contribute a considerable portion of biodiversity in the world. Costa Rica is considered as one of the regions with highest biological diversity. This small country covers only 0.01% of the earth landmass but approximately 500,000 species, roughly 5% of the living species in the world, found in Costa Rica (Casaretto, 2004). This country indeed plays an important role for carbon sequestration, forest preservation and biodiversity conservation like other countries with tropical forests.

Since several decades ago, these tropical countries, of course including Costa Rica, have rapidly had demographic increase and economic development. Although the development has resulted in economic prosperity, it has broken the balance between local society and the natural ecosystem which has existed for several centuries. The economic growth sacrificed the natural forest resources.

Tropical forests were destroyed during the 1980s with a high deforestation rate of estimated 8% (Holl et al., 2000). Costa Rica has also been involved in this deforestation trend. In 1940, more than 75% of Costa Rica was covered by forest (UNFCCC, 2006). Since the beginning of the 1960s, Costa Ricans cleared thousands of hectares of forest per year which led to the problems of serious deforestation and habitat fragmentation (Pfaff and Sanchez-Azofeifa, 2004). Until 1987 and 1997, only 21% and 42% forest cover, respectively, remained in Costa Rica (UNFCCC, 2006). Forests were converted to pasture or farmland in order to obtain higher economic values. In many upland regions, such as the Costa Rican Cordillera de Talamanca, past deforestation trends have given rise to landscape mosaics that were made up of patches of primary and secondary forest, shrub-land, pastureland, tree plantations, and cropland (Oosterhoorn & Kappelle, 2000). It resulted in fragmented forests and in segmentation of wildlife habitat. The deforestation obviously influenced the biodiversity and the carbon conservation not only in this area, but for the global ecosystem. There is a broad agreement that tropical deforestation is a massive source

of carbon emissions and contributes significantly to global warming and atmospheric change (Laurance, 2006). Therefore, deforestation of tropical forests is not only one single topic of those countries where the tropical forests were located, but it should also be a significant theme for the whole world.

A considerable number of studies indicate that anthropogenic land cover conversions and modifications have significant influence for the functioning of ecological systems. Especially the tropical forests, which were partly remote and inaccessible, were due to the pressure of increasing population rapidly converted into economical production, mainly agriculture, timber plantation, pasture land and for tourism (Veldkamp und Fresco,1997; Rosero-Bixby and Alberto,1996;UNFCCC,2006; Hayes and Sader, 2001). Not only croplands may accelerate deforestation trends, but also infrastructure lines and settlements have interrupted the continuity of forest distribution. The development of roads in national parks fragmented the wildlife habitats, resulting in difficulties maintaining their population.

It is important and urgent to commence related research in tropical forests. However, most of those tropical primary forests are inaccessible. This increases the temporal and economic burden of inventory works in those areas. Aerial photographs and satellite images have been broadly used due to their potential to cover large areas at low cost (Holmström et al., 2003). Using those images is a time- and cost-effective option as a convenient tool for investigating the landscape and the vegetation cover. It is recommendable not only for investigating short periods, but also for long-term change. Many research projects have used Landsat Thematic Mapper (TM) imagery for monitoring remote and inaccessible tropical forest regions with large areas, in which ground based inventory and monitoring methods are of limited use (Hayes and Sader, 2001). Satellite imagery was also described as a most powerful monitoring tool of forest-change-detection mapping for international agencies, local government agencies, and even NGOs working in the region for conservation purposes.

It is important to monitor the tropical forest in order to reduce the deforestation rate, administrate the global biodiversity resources and observe terrestrial carbon dynamics (Rahman, 2004). Therefore, using Landsat Thematic Mapper imagery is recommendable for achieving this ecological purpose in an economic way. This may achieve, better than traditional inventory, the goal of the low-cost and temporally efficient monitoring of natural resources in remote inaccessible areas.

2 STUDY AREA

The study area includes the Piedras Blancas National Park and, in addition (“extensive study area”), the adjacent northern areas of high elevation. It mainly covers the ACOSA and ACLA-P conservation areas (see Fig. 2) with roughly 1,075,500 hectare (sea area excluded). It is considered to be an important link between the OSA peninsula in the south and the wildlife corridor running through Central America in the north (personal communication by Anton Weissenhofer). More background information is given in the following.

2.1 Costa Rica

Costa Rica is located between the North American and the South American subcontinent. Due to its specific geographic position, forming the “bridge” between the two subcontinents and resulting in great ecological richness, it is considered to be one of the 20 countries with the greatest biodiversity in the world (Sanchez-Azofeifa et al., 2001; Sanchez-Azofeifa et al., 2002). More than 500,000 species, approximately 4 to 7% of the total species estimated worldwide, of which 84% are unknown to science, are found in this small country of 5,110,000 hectare (UNFCCC, 2006; Sanchez-Azofeifa et al., 2001; Weissenhofer et al., 2008). Especially the southern zone around the Golfo Dulce is an area of very rich biodiversity. It represents one of the species-richest areas in the neotropics (Weissenhofer et al., 2008).

However, Costa Rica in the 1960s was also one of the countries with the highest deforestation rates. The transformation and degradation of tropical forests represent the primary driving force in the loss of biological diversity worldwide (Sanchez-Azofeifa, 2001). Many sources in the literature mention and suggest that deforestation may be the first link in a chain of environment degradation that includes not only loss

of species diversity and loss of genetic diversity but also erosion, climatic change and so forth (Rosero-Bixby & Palloni, 1996). Deforestation in a country influences not only the country itself but also the whole earth's ecological system. Costa Rica was once one of the countries with high forest cover. In the 1940s, more than 75% of Costa Rica's territories were covered by forest. Unfortunately, this situation has altered and the deforestation has proceeded rapidly in the following decades resulting in forest reduction and fragmented forest landscapes. Costa Rica experienced in the 1960s and 1970s one of the highest deforestation rate, in the range of 3.2 to 3.7 % (Sanchez-Azofeifa, 2001). In 1987, the forest cover in Costa Rica was only 21 % (UNFCCC, 2006). Fortunately, since 1980s some reforestation strategies were gradually developed. Although these strategies are not evenly distributed over the whole of Costa Rica, they reverse the deforestation trend. The area of forest cover increased to 42 % in 1997 and to 47 % in 2000 (UNFCCC, 2006).

Demographic and non-demographic issues may be determining factors for this forest transformation. There are several research projects indicating that the high deforestation rate was significantly related to population growth (Oosterhoorn & Kappelle, 2000; Veldkamp and Fresco, 1997), although some research showed that the direct links between population growth and deforestation are weak. Rather, demographic pressure resulting in economic growth brought the massive deforestation. Expansion of agriculture and cattle ranching proceeded rapidly with more agro-export production in order to meet the economic demand brought by the increasing population. Non-demographic mechanisms resulted from credit and capital market failures, lack of suitable mediating institutions securing property right, uneven land distribution, consumption patterns in developed countries, greedy multinational companies and ignorance and bad management by colonists of frontier land (Rosero-Bixby & Palloni, 1996). The wrong policy direction brought the failure to the market and also put natural environment into high risk.

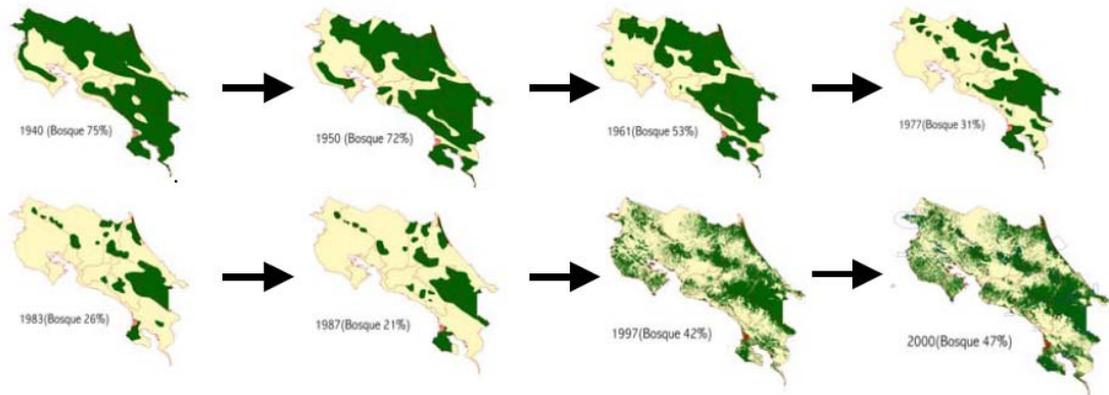


Fig.1. Deforestation in Costa Rica (UNFCCC, 2006)

As aforementioned, several tropical Latin American countries had this deforestation trend converting more land into pasture land in order to create more exportation values. Holl et al. (2000) mentioned that the primary cause of deforestation in Latin America over the past few decades has been creation of pastures for cattle grazing. Farmers cleared their land to cash in on logging or to use the land for cattle pasture or in agriculture for food production, or more importantly, for cash crops such as bananas (Rosero-Bixby & Palloni, 1996). Costa Rica was one of these countries. Ronnie de Camino(2000) also mentioned that since 1950 the main change in Costa Rican land use has been the transformation of forests into pastures and farmland, which was strongly supported and promoted by national and international loans (Sanchez-Azofeifa, 2001). Large areas of Costa Rica's land were covered by farmland, although only roughly 40% of Costa Rica's land is suitable for agriculture, and 35% of the land, even 46% in the end of the 1990s, where covered by pasture, although merely 8% are appropriate for this land use (UNFCCC, 2006; Holl et al., 2000).

However, in the 1980s the deforestation rates were significantly reduced and even reversed due to the creation of special conservation areas, reforestation promotion and forest management according to Costa Rican policies (Ronnie de Camino, 2000). The deforestation rate was dropping to less than 1.5% at the end of the twentieth century (Sanchez-Azofeifa, 2002). According to data from FAO and UNFCCC, in

2000 natural forests, plantations and special ecosystems have already covered 43-47 percent of the country's total land territory, with natural forests accounting for 38.6 percent of this. Nowadays, Costa Rica is one of the pioneer countries in reforestation, forest management, and forest protection policies. It is remarkable that Costa Rica is one of the few countries in Latin America to promote reforestation through incentives such as tax credits, direct payments, and subsidized loans that have benefited landowners, large and small(The World Bank OED, 2000).

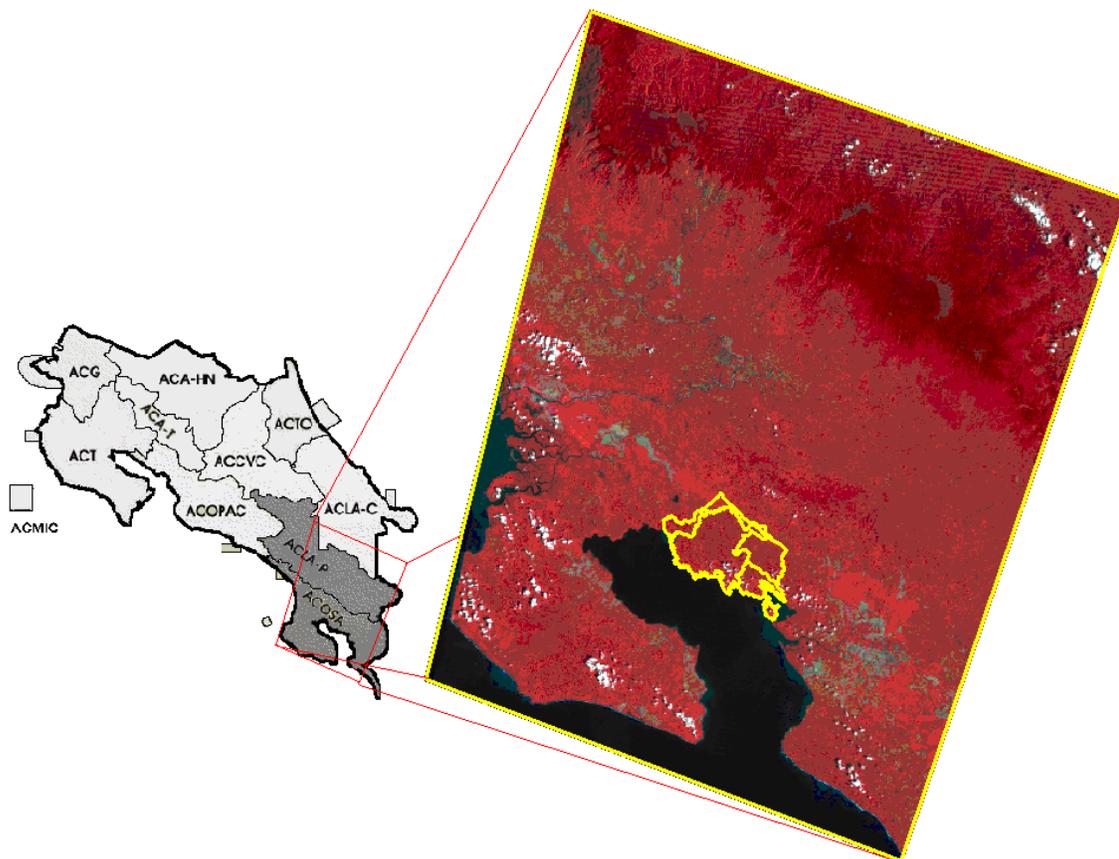


Fig.2: 11 conservation areas in Costa Rica and study area

Nowadays, more land of Costa Rica is developed and designated as protected area. The World Resource Institute indicates that in 2003 roughly 23.4% of the total land area was protected area. Some other documents also recorded approximately 25.5 % of the Costa Rica's national territory, which contains 1,304,306 hectares, dedicated to

conservation, which involves the prohibition of productive activities that could damage renewable or non-renewable resources, and to protected areas for biodiversity conservation, ecotourism, carbon sequestration and bio-prospecting for medicines (Sánchez-Azofeifa et al., 2001; Sánchez-Azofeifa et al., 2003; FAO; Costa Rica National Parks). 25 national parks, 8 biological reserves, 32 protected zones, 11 forest reserves, 58 wildlife refuges, 15 wetland/mangrove areas and 12 other categories were included in 11 conservation areas, which are spread over the whole of Costa Rica.

These 11 conservation areas were established by the Ministerio de Ambiental y Energia (MINAE) in 1998 in different regions in Costa Rica in order to undertake a decentralized management of biodiversity (Sanchez-Azofeifa, 2002; Casaretto, 2004). The conservation areas are **ACG** (Guanacaste Conservation Area), **ACA-HN** (Arenal Huetar Norte Conservation Area), **ACA-T** (Arenal Tilaran Conservation Area), **ACT** (Tempisque Conservation Area), **ACOPAC** (Central Pacific Conservation Area), **ACMIC** (Cocos Island Marine Conservation Area), **ACCVC** (Central Volcanic Conservation Area), **ACTO** (Tortuguero Conservation Area), **ACLA-C** (Caribbean La Amistad Conservation Area), **ACLA-P** (Pacific La Amistad Conservation Area), and **ACOSA** (Osa Peninsula Conservation Area). The area of this study mainly lies in ACLA-P and ACOSA.

2.2 Piedras Blancas National Park

There are two national parks, (Piedras Blancas national park and Corcovado National park) and the Golfo Dulce Forest Reserve located between the two national parks, in the southern part of the Puntarenas province, which has a total area of 429,518 ha and is situated on the Pacific slopes of Mesoamerica (Casaretto, 2004) and located in the southmost conservation area ACOSA. Corcovado National Park was developed in 1975 protecting 42,468 hectares. Piedras Blancas National park was developed in 1994 protecting 14,025 hectares of land. Golfo Dulce Forest Reserve was created in 1979 to protect 14,025 hectares of land.

Forest fragmentation and deforestation are the two main factors which have high impacts on the local biodiversity preservation. Several sources of deforestation in these three areas are mentioned by Casaretto (2004). The author indicates that in Corcovado National Park the main threats are related to human pressure, mainly the increase of socio-economic activities, from the eastern side of the Osa Peninsula. In Piedras Blancas National Park, the main source of deforestation is the river located in the Northwest section, which apparently is being used as an entrance to the park and for transportation purposes. Again, a deforestation factor in Piedras Blancas National Park is the pressure from human activities, including oil palm plantations and agriculture. In Golfo Dulce Forest Reserve, a major source of deforestation is the impact of roads. The road construction seriously interrupts the forest continuity and results in forest fragmentation.

Although the deforestation rate becomes negligible inside of national parks and biological reserve since the related forest legislation was developed in the 1980s, the areas outside of national parks and biological reserve were proven having significant forest loss (Sánchez-Azofeifa et al., 2003).

3 DATA

The data used in this study include two Landsat 5 TM satellite images, a digital elevation model from the Shuttle Radar Topography Mission (SRTM), a vegetation map, and images from Google Earth.

3.1 Landsat TM Images

According to the long period to be explored in this study, it is recommendable to use Landsat images in order to cover several decades. Landsat 5 TM satellite images were used for this analysis. The study area lies within one Landsat images in Path 15 and Row 54. The geodetic reference system WGS 84 and the projection system UTM (Zone17N) were used. Images from two dates were acquired for the study: from 1987.01.17 and from 1998.02.16. The pixel size of the images was 28.5 x 28.5m and 30 x 30m respectively.

Sanchez-Azofeifa (2001) mentioned that the highest cloud cover usually is concentrated in northern Costa Rica and the Osa Peninsula. We could observe this cloud problem when we searched through the images of several years. Most images are covered by large area of clouds. The two chosen dates had the smallest cloud cover during that period of time. The date 1987.01.17 had roughly 2.8% clouds cover, and the date 1998.02.16 had, approximately 4.2%.

Date	17.01.1987	16.02.1998
Satellite	Landsat	Landsat
Type	TM	TM
Path/Row	15/54	15/54
Spatial Resolution	28.5x28.5m	30x30m
Cloud cover	2.8%	4.2%

Table 1: Image data

The satellite image 1987.01.17 was obtained through Internet from Global Land Cover Facility center (GLCF) in March 2010. It was already georeferenced by GLCF. The 1998.02.16 TM image was obtained from U.S. Geological Survey (USGS). It was not georeferenced. This image was pre-processed and georeferenced relative to the 1987.01.17 image.

3.2 DEM from Shuttle Radar Topography Mission (SRTM)

The elevation data from the Shuttle Radar Topography Mission (SRTM) was downloaded through Internet from Global Land Cover Facility website. The SRTM data was acquired during an 11-day mission in February of 2000. The pixel size with 90m x 90m is available for the research area. The processing edition Filled Finished-A of SRTM from GLCF was used.

3.3 Piedras Blancas National Park Vegetation Map

A detailed vegetation map of Piedras Blancas National Park was employed as a reference. This vegetation map is the first detailed vegetation map in this area and was compiled in 2003 by Weissenhofer et al. based on inventory field work and aerial photographs. The map presents the state of 2003 and in some areas the state of 1998 due to clouds in the imagery of 2003 (Weissenhofer et al., 2008).

There are 23 distinct ecosystems shown in the vegetation map, considering ecotype, geographic location and land use type. They are as follows:

- Forest on plains
- Forest on hill top or ridges
- Forest on inland slopes
- Forest on coastal slopes
- Forest in ravines
- Lowland riverine forest
- Vegetation on sandy soil and gravel banks with *Gynerium sagittatum*
- Mangroves and swamp forest
- Herbaceous swamp vegetation
- Young secondary forest
- Old secondary forest
- Fern-dominated vegetation
- Regular pasture
- Swampy pasture
- Abandoned pasture (tacotal)
- Timber plantation
- Oil palm plantation
- Short-lived crops
- Preparation of land for cultivation
- Settlements and gardens
- Gravel bank
- Rivers
- Sea, lagunes

3.4 Google Earth

Google Earth was also employed in this study. Although there is no image with high spatial resolution covering the whole study area, images with high spatial resolution which covered part of the study area were used as ancillary information for the description of vegetation composition.

4 METHODS

4.1 Processing Sequence

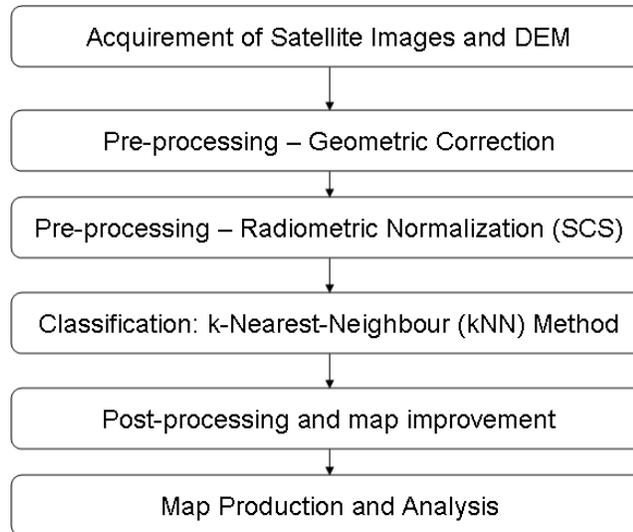


Fig. 3: Flow diagram of processing

The processing sequence is shown in figure 3. The first step was to obtain qualified satellite images for this study area. The description is given in the section 3 (data). Afterwards, the images were pre-processed. Geometric correction was performed in order to eliminate the geometric distortions of the satellite images. Another pre-processing step, radiometric normalization, was undertaken after georeferencing. Radiometric normalization compensates varying solar illumination due to the irregular shape of the terrain. Vegetation was classified employing the k-nearest-neighbour (kNN) method. This method is widely used nowadays for the estimation of forest attributes and mapping. Post-processing serves to correct the errors of classification and to improve the final results thematic map. It includes filtering for eliminating single pixels with classes differing from the surrounding area and manual editing for modifying classifications which obviously are incorrect.

Some classes may be obtained with second-choice classification. Finally, vegetation classification maps are produced and these maps were analyzed with regard to land use changes.

4.2 Pre-processing

4.2.1 Geometric Correction

Raw digital images usually contain geometric distortions so they cannot be used directly as a map base without pre-processing. The sources of these distortions range from variations in altitude, attitude, and velocity of the sensor platform to factors such as panoramic distortion, earth curvature, atmospheric refraction, relief displacement, and nonlinearities in the sweep of a sensor's IFOV (Lillesand et al., 2007). Geometric correction refers to the removal and compensation of geometric distortions introduced by these factors in an image.

Geometric correction uses well-distributed ground control points, which could be obtained from GPS information of on-site measurements or from corrected maps. Using these control points the distortions occurring in an image may be eliminated.

As the image of 1987.01.17 was already geo-referenced by GLCF, this image was used as the reference image to correct the image 1998.02.16. The image geometric correction function of ERDAS was employed in this study and the DEM from SRTM was used as elevation file. The resampling method was nearest neighbor using a 30-by-30 m pixel size for image 1998.16.02 and a 28.5-by-28.5 m pixel size for image 1987.01.17. There were 15 control points used which were evenly distributed over the whole image. The total control point error (RMS Error) in 1998 is 10.9897 meters.

4.2.2 Radiometric-Topographic Normalization

The pixel values in remote sensing images may be strongly affected by the influence of terrain slope and aspect. Differences in terrain slope and aspect often create variations in signal values between pixels with even similar properties of land cover and of vegetation as a result of differences in irradiance owing to the incidence angle of illumination (Riaño et al., 2003; Soenen et al., 2005). Especially in the rugged mountainous areas, the received effective illumination may be very different due to topography. The shadowing caused by topography has strong effect on the overall canopy brightness at the pixel scale.

In addition to the rugged terrain effect, the effect of canopy self-shadowing of forested land can also be remarkable when forest related topics are studied. The canopy self-shadowing, at sub-pixel scale, results from the complexity of forest composition and the tree shade from the neighbour trees nearby. It is the unresolved shadowing due to objects within a pixel as opposed to shadowing from resolved up-sun objects in other pixels (Kane et al., 2008).

Radiometric normalization or correction is the compensation of the different solar illuminations due to the irregular shape and slope-aspect effect of the terrain. It modifies and eliminates the radiance values changes caused by terrain effect.

The radiometric correction methods may be grouped into two categories. The first group is based on band ratios. These are simpler and require no additional input data. The second group is based on modeling illumination conditions and requires a DEM of the same resolution as the image to be corrected. This group may again be divided into two categories, Lambertian, and non-Lambertian, depending on whether they assume that reflectance is independent of observation and incident angles or not (Riaño et al., 2003).

Before the Gu and Gillespie study (Gu and Gillespie, 1998), it was commonly assumed that radiance is controlled by the traditional Sun-Terrain-Sensor geometry (STS), which is based on the relationship between these three elements only, without considering the relevance of the vertical direction. Nevertheless, forest and non-forest have different radiometric characteristics. Several normalization models based on the Sun-Terrain-Sensor (STS) geometry were developed, including the *cosine correction*, *c-correction*, *statistical-empirical correction* and *Minnaert method*, in order to remove or compensate for topographic effects on forested images.

In 1998 Gu and Gillespie showed that the Sun-Canopy-Sensor (SCS) radiometric correction model,

$$L = L_o \cdot \frac{\cos \theta}{\cos \alpha \cdot \cos \theta_s} \quad (1)$$

<i>Symbol</i>	<i>Explanation</i>
L	Radiance from pixels on the sloped terrain
L_o	Radiance from pixels on the horizontal terrain
θ	Angle between terrain normal and direction to sun (degree)
α	Terrain slope angle (degree)
θ_s	Sun zenith angle (degree)

Table 2: Symbols used in SCS correction models

based on the mutual shadowing at subpixel scale, and on the sunlit area within a pixel, may have better compensation results and be more appropriate than terrain-based corrections in forested areas (Gu and Gillespie, 1998). SCS, which is a physical-structural basis to radiometric correction, preserves the geotropic nature of trees (vertical growth) regardless of terrain, view and illumination angles (Soenen et al., 2005). It was shown that SCS was more accurate for the forest canopies of 50% or higher closure comparing with cosine correction, c-correction and Minnaert

correction models, with either slightly or significantly better results. The study by Soenen et al. in 2008 also showed that the physically based approaches, SCS or the modified SCS+C method, are more appropriate for forested terrain, as they account for vegetation structure and geotropic tree structure. The other correction techniques, Minnaert, C, cosine and statistical-empirical, have effective slope-aspect correction for smooth surfaces but fail beyond this.

However, the SCS correction was found with two main problems. First the SCS correction fails to give accurate corrections for any of the real canopies due to providing a fixed value for each slope-aspect-SZA (Sun Zenith Angle) combination (Kane et al., 2008). Another problem of SCS correction is an over- or under correction on steep pro-and anti-sun slopes similar to other simple photometric functions (Soenen et al., 2005; Kane et al., 2008). One modified model (SCS+C), which was proposed in the study of Soenen et al, could mitigate the problems of SCS overcorrection for steep slopes. Another modified SCS method as described in Koukal et al. (2005) was applied for this work:

$$L = L_o \cdot \frac{c_s \cdot \cos \theta + (1 - c_s) \cdot \cos \alpha \cdot \cos \theta_s}{\cos \alpha \cdot \cos \theta_s} \quad (2)$$

C_s is the ratio of direction sun irradiances to the total (global) irradiance.

Our study area has high elevation difference from 0 to over 3500 meters above sea level and has abundant tropical rain forest resources. Therefore, it is recommended to employ SCS model. The modified SCS model by Koukal et al. was applied in this study.

4.3 *k*-Nearest Neighbour (kNN) Program

4.3.1 Methodical Background

The *k*-nearest-neighbour (kNN) algorithm is a non-parametric estimation approach and one of the widely applied classification methods. The satellite imagery data and reference data, which may come from an inventory in the field, have to be employed in the kNN method.

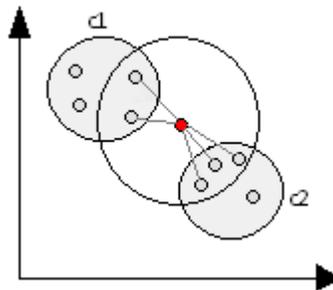


Fig. 4: Diagram of kNN Algorithm in Feature Space

The basic idea of the kNN-method is to compare an element, for which attributes have to be estimated, with elements of known attributes on the basis of other features of the elements. It is a method for classifying the objects of interest based on closest training examples and evaluating the Euclidean distance in the feature space. An object is classified in the feature space by a majority vote of its neighbours, and is assigned to the class appearing most frequently amongst its *k* nearest neighbors. These training objects have to be located on the same image as the unknown pixel to ensure that external factors such as sun elevation and atmospheric conditions at the time of image acquisition influence training and reference objects in the same way. Using the simple diagram of Fig. 4, the kNN algorithm may be explained. For instance, if the object to be classified is the red point and $k=5$, the 5 nearest

neighbours to this object have to searched for. The object is assigned to the class which appears most frequently among these 5 neighbors. In this case, the object is assigned to class 2 due to the highest appearing frequency.

Two factors may significantly influence kNN and may let it fail to work well. One is an inappropriate choice of distance metric, and another one is the presence of numerous class-irrelevant features. The most important algorithm parameters are the distance metric, which is defined in the feature space of the satellite image data, the number k of nearest neighbors, and the extent of geographic search radius (Haapanen, 2004; Tomppo et al., 2004). The Euclidian distance function which was employed in this study is as follows (Koukal, 2009):

$$ED(X_i, X_j) = \sqrt{\sum_{f=1}^F c_f (x_{i,f} - x_{j,f})^2} \quad (4)$$

X_i and X_j are vectors in F-dimensional feature space, $x_{i,f}$ and $x_{j,f}$ are the corresponding vector elements (pixel values in spectral bands f). The coefficients c_f are the weights of the spectral bands in the Euclidian space. The optimal number of nearest reference observations k is evaluated in the kNN process, which was aforementioned. Haapanen et al. (2004) described in their study that the overall accuracy is only marginally influenced by the number of neighbors considered, k, weighting of bands, band combinations etc. When k is larger than 3, a variation of k shows only approximately 2 to 3 % variations of overall accuracy.

In the kNN method, the geographic search radius can be applied, including horizontal and vertical search radius. The geographic search radius may help classify features with certain location of geographic characteristics.

The kNN method is widely applied also in natural sciences. It is proven that kNN classification can possibly be used for stratification purposes in large-scale forest inventories (Haapanen, 2004) and it was employed for the estimation of forest attributes and mapping using remote sensing data (Koukal, 2007; Haapanen et al., 2004; Gjertsen, 2007; Finley et al., 2006; Finley and McRoberts, 2008). Many countries, including Finland (Katila and Tomppo, 2001; Tomppo, 2004), Sweden (Reese et al., 2005), Norway (Gjertsen, 2007), United States (Franco-Lopez et al., 2001; Haapanen et al., 2004), and so forth, employ this method for their national forest resource inventory and related research.

In order to utilize the kNN algorithm in forestry, we have to combine the field data, which are obtained at the sample plots as reference, and, of course, remote sensing data (Koukal, 2007). Haapanen (2004) also indicated that this method may couple satellite imagery data with field-based inventory information. The inventory information may at one hand assist in determining areal extent of the real ground vegetation composition. Erroneous classification could be avoided. At the other hand, this ancillary data may assist in producing continuous digital layers of measured forest or land use attributes. In this study the vegetation map in the Piedras Blancas National Park was employed as reference data.

It was tried to apply the search radius function to correct and modify the classification. Unfortunately, it was not appropriate, neither the horizontal nor the vertical one, because most of the reference points were located in the national park and only some extra reference points were distributed in certain areas. Consequently, only the areas which are located in the search radius near the reference points could be processed. If no reference points are located in remote areas, then no results or an incorrect classification are produced. The search radius function could therefore only be applied under the situation that reference points were distributed all over the study area.

4.3.2 Comparison of Advantages and Shortcomings

Advantages

There are several important advantages of the kNN method which make this method as an attractive tool for forest attribute estimation and mapping.

- One advantage is that kNN is easy to implement. The kNN method makes no distributional assumptions, or relaxed assumptions, such as normality or homoscedasticity, in feature space about the forest variables used (Thessler et al., 2008; Koukal, 2007). It is non-parametric and thus avoids the need for explicit models (Tomppo and Halme, 2004).
- The kNN method has the potential for combining different sources of information, not only from outside of a region of interest, but from different forest inventory designs (Franco-Lopez et al, 2001).
- The kNN method produces both statistics for arbitrary area units and for wall-to-wall maps, which can be used in addition to forestry studies (Tomppo and Halme, 2004).
- Furthermore, multiple dependent forest variables (parameters) of inventory for each pixel or for area units of interest can be simultaneously computed, estimated and predicted with the kNN method (Finley et al., 2006; Franco-Lopez et al, 2001; Tomppo and Halme, 2004).
- The kNN method better preserves the covariance structure among dependent and in dependent variables (Tomppo and Halme, 2004; Finley et al., 2006).
- In addition, the kNN method is a very “transparent” technique, since it is based solely on the search for similar units (Franco-Lopez et al, 2001).

Disadvantages

However, there are also several limitations of the kNN method. Tomppo and Halme (2004) indicated several shortcomings of this method.

- The kNN method presumes that the total variation of all forest variables is well represented by the field sample plots.

- The predictions, particularly the predictions of volumes by tree species, may be biased if the area of interest is large and covers several different vegetation zones with different tree species compositions. The biases can be reduced if the set of potential nearest neighbours is somehow restricted (Tomppo and Halme, 2004).
- Varying imaging conditions within the area of a satellite image may also change the covariance structure between field data and image data
- In practice, the kNN algorithm is computationally intensive. For a small study area, this is negligible. Problems may be encountered for large study area, where the mapping efficiency may be significantly reduced. However, Finley and McRoberts (2008) described that the operation efficiency may be increased by using the kd-tree, or similar data structures, and efficient exact or approximate search algorithms.

4.3.3 Components

In this study, the kNN software developed at the Institut für Vermessung, Fernerkundung und Landinformation (IVFL) (Koukal, 2009) was used. This program consists of 3 modules: READIMAGE, CROSS-VALIDATION, and MAPPING (see Fig. 5).

The READIMAGE module is in charge of combining the input images and ancillary data and reading out the pixel values, including the TM band values, elevation values, slope values and so forth. The information is output as a “kNN-reference-file”, which is used in the next two modules.

Second module is CROSS-VALIDATION, which is a standard tool of estimating error statistics of prediction. It is easy to implement and provides nearly unbiased estimates of prediction error (Koukal, 2009).

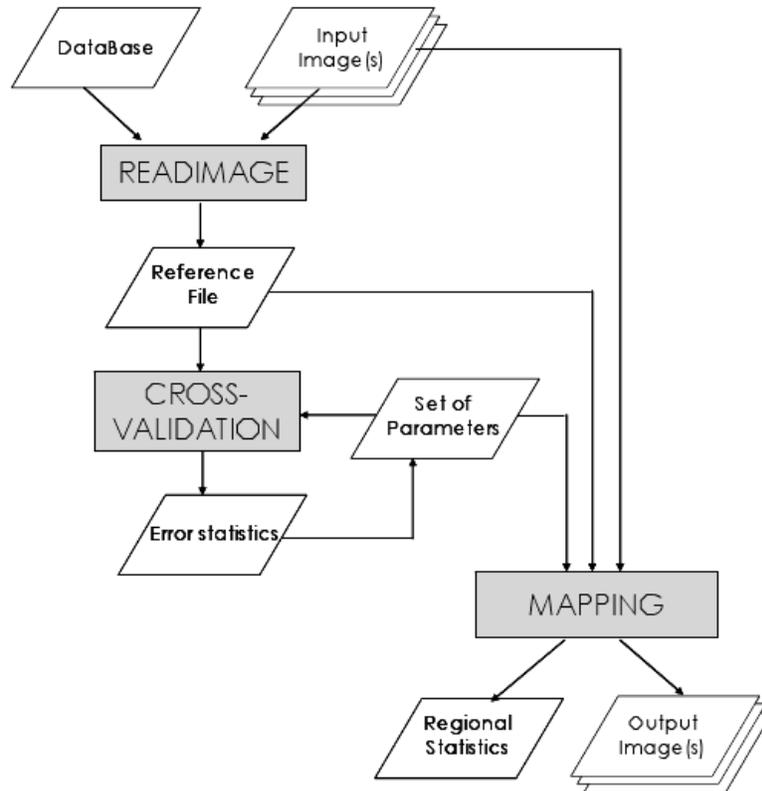


Fig. 5: kNN-program flow chart (Koukal, 2009)

Error statistics is produced, including the producer's accuracy, user's accuracy, overall accuracy and estimated Kappa value, which may be used for parameter optimization and accuracy assessment.

In addition, this tool may be used to detect doubtful sample points, namely outliers. The outliers can be eliminated for the sake of increasing overall accuracy and Kappa values. The basic idea of the outlier elimination method is the following: In the cross-validation procedure with $k=1$, it is determined, how often every reference point is used as a nearest neighbour, and how often this leads to a correct result. Reference points leading to incorrect results in more than 50% of the cases can be considered as outliers.

The CROSS-VALIDATION procedure has several forms. The multi-fold CROSS-VALIDATION, also named two-groups-method, and the leave-one-out method, which is a special kind of the two-groups-method, are two most common procedures.

Multi-fold CROSS-VALIDATION randomly splits the reference data into one training data group and one testing data group (Koukal et al., 2007; Koukal, 2009). In the leave-one out-method, the reference data is also split into two groups, training and testing. The only difference of the leave-one-out method is that the testing set consists of one single pixel and all other reference data are grouped into the training set. This form omits training sample units one by one and mimics the use of independent data (Haapanen, 2004). The leave-one-out method is especially suited for rather small reference data sets, because the maximum amount of training data is available for the classification of every pixel and the sample may be too small to get reliable estimates when splitting the sample into two exclusive data sets (Koukal et al., 2007).

Gjertsen (2007) mentioned that the accuracy assessment in CROSS-VALIDATION may be enhanced in two ways. In addition to the original bands of the images, it is helpful to use derived measures such as band ratios to improve the level of association. These band ratios may improve the separability of the forest variables. This may result in enhanced overall accuracy. Another possibility is to use auxiliary information such as site index maps (e.g. soil maps) to reduce the set of reference plots for each pixel to be classified, and to use only plots with the same site index class as the pixel to be classified. In this study, the slope information which was extracted from SRTM was inserted in the kNN-reference-file as an extra band 8 together with the image data in order to improve the overall accuracy.

MAPPING module is applied as the last step in this program. Pixel by pixel is estimated and the result is written to raster files (maps). Besides the maps, regional statistics are also generated in this module.

4.3.4 Process Sequence

The detailed vegetation map of the Piedras Blancas National Park, which was generated by Weissenhofer et al. in 2003, was employed as ancillary information for reference point selection. Although this vegetation map represents the state of 2003 and in some areas the state of 1998 due to clouds in the imagery of 2003, it was the only land use type information which could be acquired as reference data base. This vegetation map and the images of two acquisition dates (chapter 3.1) were combined for generating the kNN-reference-file.

ArcGIS was employed to produce pointsets as reference points in the area within the vegetation map. Extracting enough ground sample points was essential for successful application of the kNN method. The accuracy of the generated land-cover information is highly dependent on the measurable quality of the reference information. Haapanen (2004) mentions that it is important for the reference data to capture the range of spectral variability, which reflects the terrestrial characteristics, within a class in order to obtain good class estimates with the kNN method. It was assumed that the spectral characteristics of most of the land use types are represented best in the homogeneous areas of these land use types. However, pixel values within one land use type may also vary, due to various species composition, regional soil humidity or stand microclimate condition. It is essential to attain continuous variable estimates covering the range of the spectral variability of every land use type. The reference points should be well distributed. The images which were employed in the kNN-method should be covered by the terrestrial inventory area to an extent as large as possible (Koukal et al., 2007).

First of all, two kinds of reference points were generated, one with interval 500 meters and the other one with interval 250 meters. In the regions with mono-vegetation types and large areas, the reference points with interval 500 meters were used. For instance, within the primary forest which occupied large areas, reference points were used with interval 500 meters. In the areas where the vegetation types are

more discontinuous and fragmented, the reference points with 250 meters were used in order to obtain more reference information. These pointsets were also visually modified by moving points from doubtful areas, which had uncertain land use type, and from land use type edges, where the points may lie on mixed pixels of different land use type. The points in cloud and shadow areas also had to be moved to non-cloud or non-shadow areas. Those points were moved to areas of uniform land use. Some additional reference points outside the vegetation map area, which could be visually classified with certainty, e.g. oil palm plantations, mangroves, fallow land, or primary forest in the remote high elevation area, were also generated.

As aforementioned in chapter 3.3, there were 23 distinct ecosystems shown in the Piedras Blancas National Park vegetation map. First, these 23 classes were preliminarily classified into 12 classes (see table 3). The total numbers of reference points were 2496 and 1994 for image 1987.01.17 and 1998.02.17, respectively. There were at least 30 reference points per vegetation class, except for class 7, timber plantation. This class had only 20 and 15 reference points for image 1987.01.17 and image 1998.02.16, respectively.

In the crossvalidation, the results of pre-classification showed that the accuracy of class 7 timber plantation wasn't trustable. It seems that timber plantation has similar spectral characteristics like secondary forest, producing indistinct results of crossvalidation.

In order to have more reliable results, the classes were further combined into 7 new classes (see table 3). Primary forest and secondary forest were of high interest in this study, hence they were kept invariant in the second classification system. Mangrove is one class with specific spectral characteristics and easily detected and classified. It was proven having qualified classification results, and therefore class mangrove was kept in the new classification system.

<i>Pre-classification</i>		<i>classification</i>	
1	Primary Forest	1	Primary Forest
2	Secondary Forest	2	Secondary Forest
3	Mangrove	3	Mangrove
4	Regular Pasture	4	Grassland
5	Marsh Grassland		
6	Fallow Land		
7	Timber Plantations	5	Arable Land
8	Oil Palm Plantations		
9	Arable Land		
10	Settlement	6	Non-vegetation
12	Gravel Bank		
11	Waters	7	Waters
0	Clouds, Shadow, Out of Research Area	0	Clouds, Shadow, Out of Research Area

Table 3: Categories of the classification

Regular pasture and marsh grassland were combined to grassland. Grasses and shrubs may colonize the pasture soon after abandonment (Holl et al., 2000). Abandoned pasture or marsh grassland may be reasonably classified to grassland. Fallow land could be considered as abandoned land, which was recovered by grasses and shrubs very soon. Therefore, fallow land was also classified to the new class grassland. The preliminary class 7 timber plantation, 8 oil palm plantation and 9 arable land were considered as anthropogenic arable land. Accordingly, they were reclassified as new class arable land. The preliminary class 10 settlement and class 12 gravel banks were the land use types without vegetation. They were combined as the new class non-vegetation.

The coefficients c_f in equation (4) are significant factors influencing the overall accuracy and the Kappa values in the kNN-method. The coefficients enhance or reduce the influence of bands. As aforementioned, there are 7 bands in Landsat TM images. All the bands, except band 6, were employed in the kNN-method. Band 3,

which is the visible red band with wave length 0.63-0.69 μm , helps to discriminate vegetation and soil. Band 4, which is near infrared with wave length 0.76-0.90 μm , also discriminates vegetation and soil and has the ability to distinguish vegetation varieties and conditions. Arable land can be separated from the land use types without vegetation. Besides, arable land and grassland may show higher reflectance than the forest vegetation types. Band 5, which is mid-infrared with wave length 1.55-1.75 μm , detects the water content. It is distinctly separating water bodies from other land use types. This band is sensitive of the amount of water in vegetation and also can assist separating arable land and forest. Several research studies have proven that the wave band combination of the visible red, near infrared and mid-infrared spectral regions (band 3, 4, 5), has great potential for vegetation discrimination (Hayes and Sader, 2001). Accordingly, the coefficients of band 3, 4 and 5 were given enhanced weight and set to 1, and bands 1, 2, and 7 were given lower weight and set to 0.4. The weight of bands 1, 2, and 7 was determined by trial-and-error optimizations of the overall accuracy.

Band 6 was not available for the image 1987.01.17. Consequently, it was not considered to use this band for classification.

Some land use types are restricted to flat areas, others to sloped areas. For instance, mangrove forest appears only in the flat area near the river mouth or coast below 20 meters above sea level, and also oil palms and arable land only occur on flat terrain. To use this information, the slope information was also extracted and inserted to the kNN-reference-file as band 8 in order to assist the classification. The coefficient of band 8 was set to 1.

Mask files were separately used for each image in this study in order to eliminate clouds, cloud shadows and the regions outside of the study area. Two benefits of the use of mask files can be mentioned. First, the pixel values from clouds or shadows which would affect the kNN classification are avoided. Second, the computing time could be reduced.

4.4 Classification Enhancements

The classification results obtained with the kNN program were checked first by comparison with the detailed vegetation map in the Piedras Blancas National Park. This was done in order to ensure that the classification results are reliable. Furthermore, the whole study area had to be verified. The doubtful areas with uncertain vegetation type or incorrect vegetation classification results were checked with ancillary images from Google Earth or through the interview with a botanic expert who is familiar with the local vegetation. Additional reference points were set in those doubtful areas, e.g. in swampy areas which were classified as oil palms, or in high elevation areas which were classified as mangrove forests. The correct classes were assigned to these reference points. Reference points were also set in remote areas, for instance mountainous areas above 2500m sea level. These mountain areas were covering largely distinct ecological life zones. These additional reference points were necessary due to the gradual change of vegetation properties along the latitude and elevation.

A maximum distance can be set in both horizontal and vertical directions (search radius function) in order to avoid selecting the nearest neighbours (spectrally similar plots) from a region in which the response of image variables to field variables is assumed to be not similar to that of the pixel under consideration (Tomppo et al, 2004). The search radius function may modify and improve classifications. For instance, mangroves appear only at roughly 0 meter sea elevation and near river mouths. Another example is that arable land and oil palm plantation may only be developed on flat terrain at lower elevation. These land use types occur rarely at higher elevations and in remote mountainous areas, therefore pixels which are classified as these classes should be modified. However, it is not recommendable to use the search radii in this study. The elevation ranges from 0 to 3802 m and, as aforementioned, the reference points are mostly located in the Piedras Blancas national park, where the highest elevation is roughly 200 m above sea level. Therefore, no reference points exist for higher elevations. The pixels which were

located at higher elevation are not classified when a vertical search radius is applied. The search radius function could only be applied if the reference points were evenly distributed in the whole study area, at all elevations.

The second-choice selection was also considered for classification enhancement. It means that not the class with the highest number of neighbours is chosen, but the class with the second-highest number of neighbours. The second-choice class is obtained from the “fuzzy file” produced by the kNN mapping process. The fuzzy file lists, for every pixel to be classified, all classes represented by the nearest neighbours, together with the “membership degree” of the pixel to each class. A weight function for the neighbours can be chosen in fuzzy kNN classification. It was chosen proportional to the inverse of the Euclidean distance so as to avoid the situation that two or three of the classes had the same membership degree. Under the situation that several classes had the same membership degree, the selection of the class to be assigned was difficult. Although the problems of the doubted areas could be partly solved by the second-choice selection, the pixels which have only one choice (100%) still created problems and made this process meaningless. Therefore, increasing the number of reference points was more recommendable than other methods for classification enhancement in this study.

4.5 Post-processing

After the vegetation classifications were performed, a majority filter was employed. This filter is implemented in ERDAS under Interpreter, GIS analysis. Filter size 7 x 7 was applied in this study. It means that the target pixel is located in the center of a 7 x 7 window, and it is assigned to the class which appears most frequently among these 7 x 7 pixels. This function can filter out isolated pixels with a class different from the surrounding area. For instance, some pixels belonging to mangrove appeared in the large primary forest area. Those pixels were certainly assigned to the incorrect class. The neighborhood filtering function can assist to eliminate these incorrectly classified pixels.

4.6 Pixel Change Detection

Before applying the pixel change detection function, one preparation had to be done. Image 1987.01.17 had spatial resolution 28.5m x 28.5m and image 1998.02.16 has 30m x 30m. In order to compare the class change pixel by pixel, the image 1987.01.17 was transformed to the 30m x 30m pixel size and pixel alignment of image 1998.02.16, using the project raster function in Arc GIS.

The combine function in Arc Toolbox was applied to detect the pixel value change of land use types. This function can combine multiple raster files, assigning a unique output value to each unique combination of input values.

According to the main aim of this study, the 3 “change paths” “primary forest preserved”, “primary forest to secondary forest or grassland” and “primary forest to other types” were extracted from this change matrix of 64 possibilities of change.

5 RESULTS

5.1 Crossvalidation

5.1.1 Pre-classification with 12 Classes

The results of crossvalidation are shown in the appendices. The error matrices for the classification from the images of the two dates (1987.01.17, 1998.02.16) for $k = 6$ may be seen in tables 4 and 5. This value of k was found to give the highest overall accuracy and the highest Kappa value. The overall accuracies of image 1987.01.17 and 1998.02.16 were roughly 77% and 82%. The Kappa values are 72% and 77%, respectively.

In both pre-classifications, class 7 timber plantation showed the most problematic results. In image 1998.02.16, the producer's accuracy of this class was only 0.1. The image 1987.01.17 showed even worse results. There was not one single pixel of this class classified correctly. Except the influence of time difference between vegetation map and images, there are two possibilities to explain this. One possibility is that there are too few reliable reference points for classification. According to the vegetation map, the areas of timber plantation are rare, and they are scattered in small patches. The reference points could be erroneously assigned to other classes of neighbouring areas. The second possibility could be that timber plantation is spectrally very similar to secondary forest. Besides, depending on the species which are cultivated, it may be misclassified to other class too.

When the timber is quite young with tree height not much above the height of grass, it may also be mistakes for grassland or farmland.

5.1.2 Classification with 7 Classes

As aforementioned, a new classification with seven classes was performed (see table 3). Several pre-classification classes were combined together to get higher overall accuracy and more reliable results. The new classes are primary forest, secondary forest, mangrove, grassland, arable land, non-vegetation and waters.

Tables 6 and table 7 in the appendix show the crossvalidation results of the classification with 7 classes of both years. Apparently, the overall accuracies for both images were increased, from 77% to 81% for image 1987.01.17 and from 82% to 86% for image 1998.02.16. The Kappa values were also raised, from 72% to 81% for the former image and from 77% to 86% for the later image. These new classes show more general instead of more detailed information, and accordingly the classification results are more reliable. Except class secondary forest, which shows 49% and 54% producer's accuracy for the two years, the other classes have moderate to high producer's accuracies. The low accuracy for secondary forest is not surprising, because secondary forest is showing vegetation characteristic between primary forest and grassland. Secondary forest was relatively poorly distinguished from primary forest and grassland. It acts as a residual group between these two groups. Grassland may be rapidly occupied by bushes or the invading pioneer tree species. The secondary forest may grow very fast and become primary forest. Tropical ecosystem succession is relatively dynamic and quick. Besides, the definition of secondary forest also has an impact on the classification result.

5.2 Classified Vegetation Map

5.2.1 Piedras Blancas National Park

The results of vegetation classification in the Piedras Blancas National Park are shown in Fig.6. 7 classes are distinguished (see table 3). This classification is solely based on the Piedras Blancas National Park vegetation map of 2003 as a reference. The classification results therefore are doubtful, especially for the image 1987.01.17, despite the use of the outlier classification method. Comparing the classification results of imagery 1987.01.17 and 1998.02.16, the former image obviously has more classification errors. The result of image 1987.01.17 has more pixels classified to arable land which does not exist in the reference vegetation map. This error may come from the time difference between the reference vegetation map which represents the situation in 2003 and the image acquisition date. Land use types may have significantly changed during this period of time. Especially in tropical ecosystem, the vegetation may change rapidly. The abandoned pasture or arable land may be invaded by shrubs within a few years.

In the classification of the image of 1987.01.17, there are more primary forests and fewer secondary forests. It is quite plausible that more undisturbed forests existed in 1987. The locations of the patches of secondary forest and grassland appear relatively correct for both dates. Mangroves, near the river mouth, were precisely classified. Except the north-south belt in the western part of the national park and some fragmented patches which were incorrectly classified, the arable land in the north and east of the national park is relatively correctly classified.

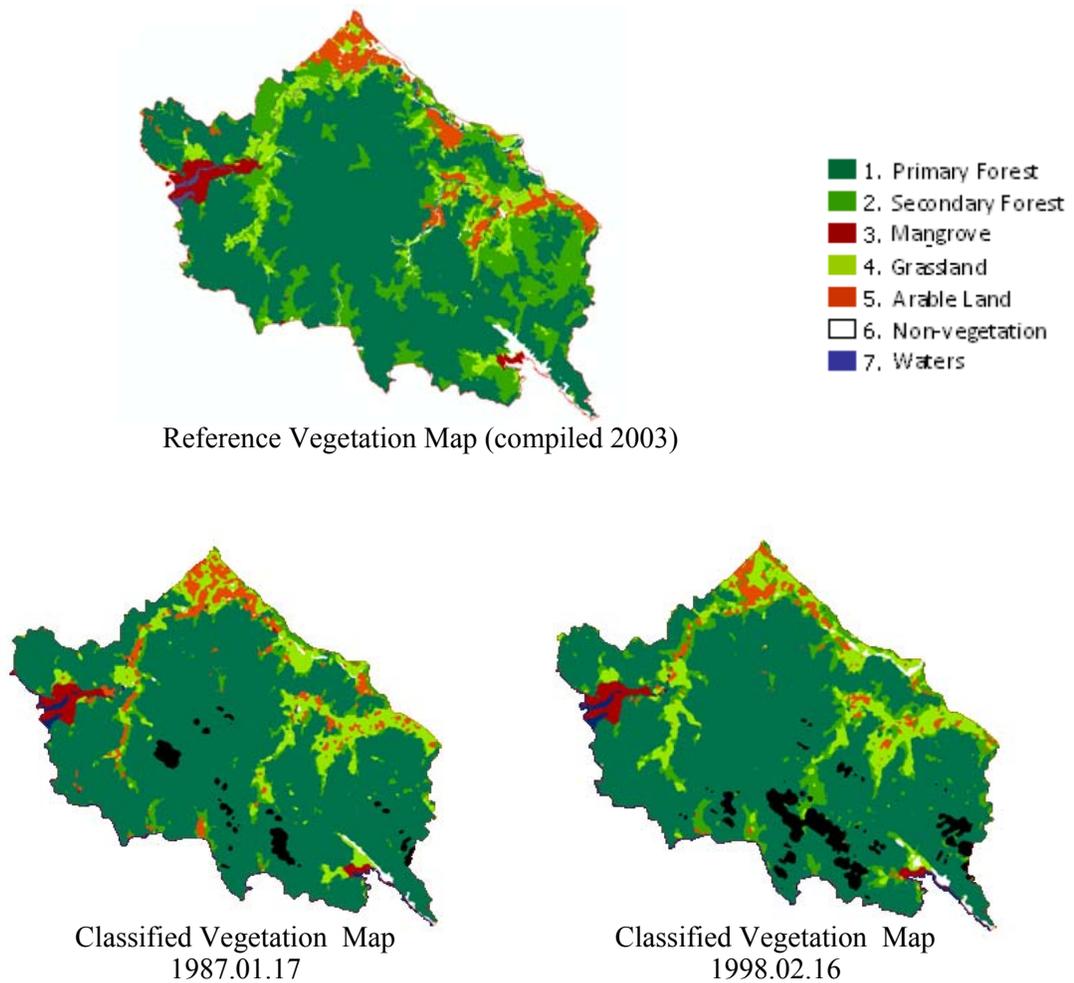


Fig. 6: Results of vegetation classification in the Piedras Blancas National Park

It can be assumed that the classification of image 1998.02.16 is more reliable compared with image 1987.01.17. There are fewer erroneous pixels assigned to arable land in the west part, and the pattern of land use types fits better to the vegetation map. In general, this image shows more primary forests than the vegetation map, similar to image 1987.01.17, and it shows fewer patches of secondary forest and less arable land. More pixels which belong to secondary forest in the vegetation map were classified as primary forest in this image. The classification of mangrove was also reliable in this image. As aforementioned, the classifications of these two classes seem quite correct, although more grassland and less arable land are shown compared with the vegetation map. This may again result

from the time difference between the image acquisition and the compilation of the vegetation map. A conclusion can be made so far that the results of classification in the national park area are moderately reliable for image 1987.01.17 and reliable for image 1998.02.16.

5.2.2 Extensive Study Area

The results of the classification with 12 classes are shown in the appendix, figures 12 and 13. Only the classifications with 7 classes which are reproduced in figures 7 and 8 are discussed here. Nevertheless, one interesting point found in the pre-classification result with 12 classes may be mentioned here. The pre-classification map of image 1987.01.17 (Fig. 12) shows much more fallow land than the classification map of image 1998.02.16. Both images were taken in January and February, within the dry season, which lasts from December through April in Costa Rica. Therefore, the weather factor might not be responsible for this substantial difference. There is one possibility to cause this phenomenon. In the 1980s, Structural Adjustment Programs (SAPs) introduced by the World Bank reduced the profitability of agriculture and cattle ranching in marginal forest lands. It was also mentioned by Weissenhofer et al. (2008) that many banana plantations were closed or abandoned when the United Fruit Company withdrew its farmland and left Costa Rica in 1986. The image 1987.1.17 was taken right after the withdrawing year. Grass and shrubs may have not invaded these farmlands a year after abandonment. Therefore large areas of fallow land may appear in this image.

Although the classification with 7 classes shows reliable results in the Piedras Blancas National Park, it shows more problems in the whole study area. For the image 1987.01.17, the problems of class 5, arable land, are extending from the national park to the whole area. There are many misclassifications of arable land. The image 1998.02.16 shows more accurate classification results. Although this later image yield better results, there are also some areas with doubtful classification.

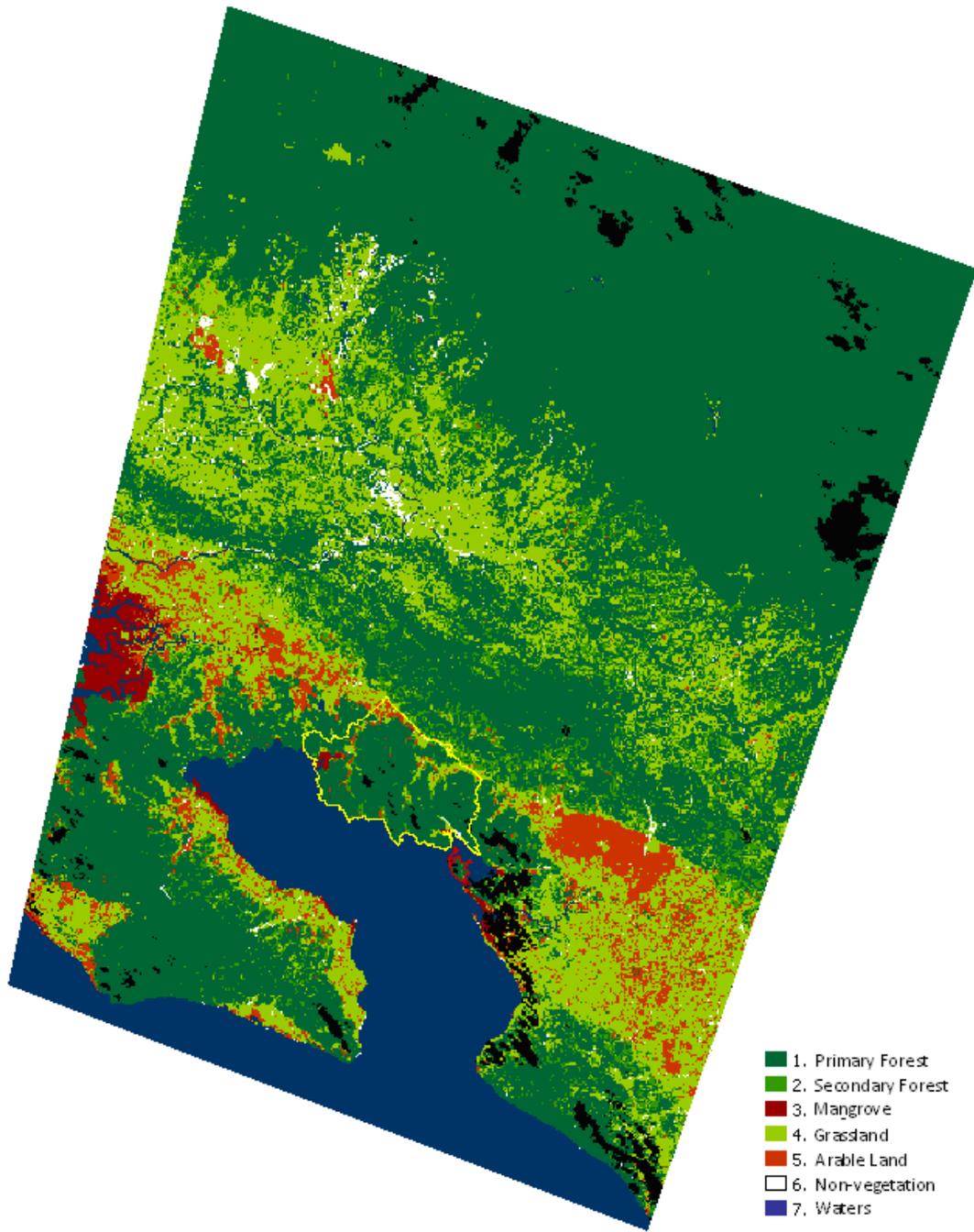


Fig. 7: Classification map of image 1987.01.17

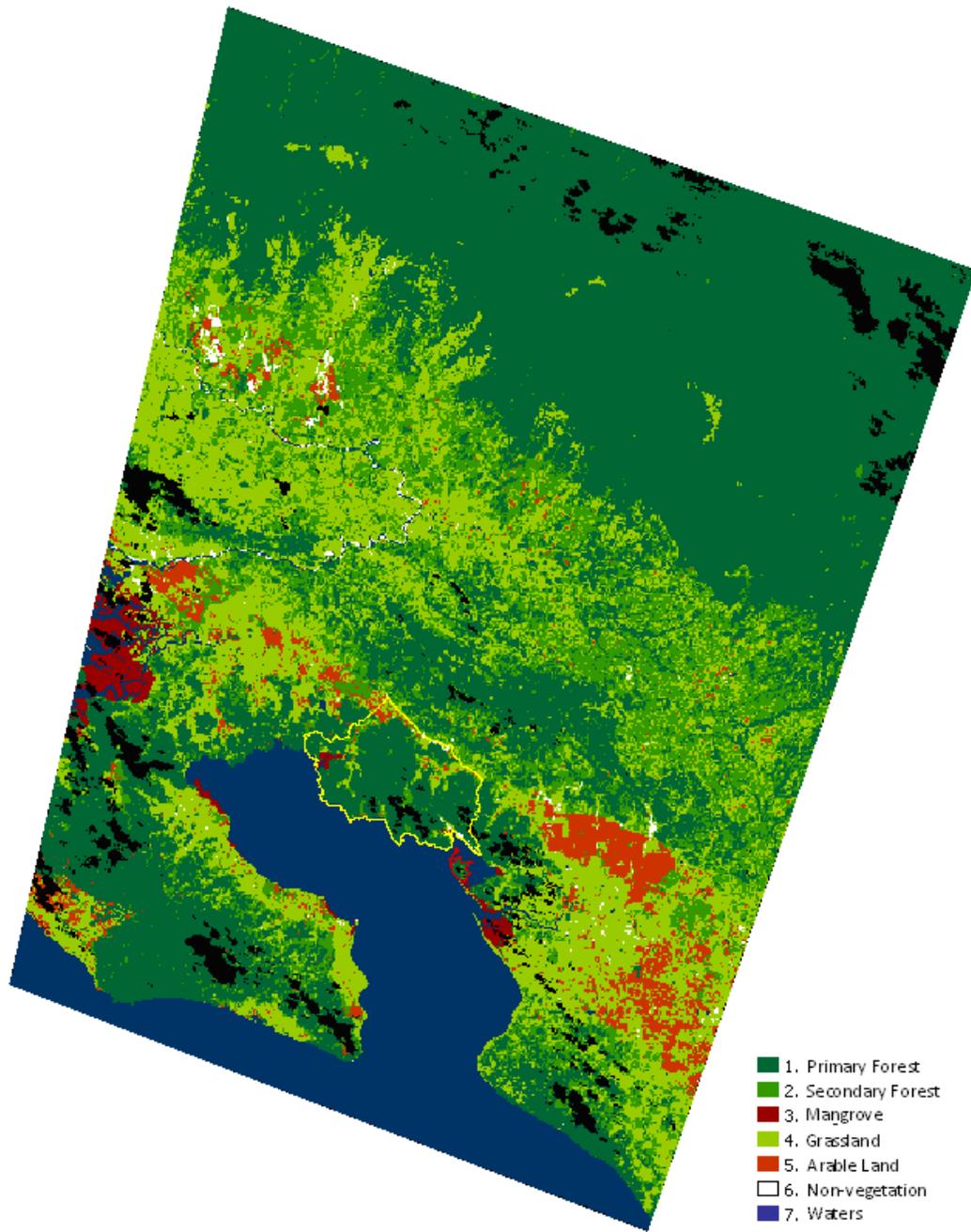


Fig. 8: Classification map of image 1998.02.16

For instance, the south-west of the peninsula with the lagoon should be pre-classified with 12 classes as wet-land, which is included into the grassland class afterwards. But it is pre-classified as oil palm. Image 1987.01.17 also has this problem. Some residuals from this class can still be seen, although roughly 20 extra reference points were generated in this swampy area.

Deforested areas were identified by comparing the primary forest, secondary forest and grassland in both images. In general, comparing the two classification results, more primary forests certainly existed especially in the northern high elevation areas in the former image. Premontane and montane cloud forests were usually less accessible and less desirable for agriculture because of rainfall, rugged terrain and soil fertility limitations (Rosero-Bixby et al., 1996). This could explain that most of the primary forests were located in the northern part of the study area. In this area, the elevation is higher, more than 2000 meters, and it is relatively remote. It is also near other forest reserves, Talamanca range and La Amistad National Park. Drier life zones are relatively accessible and have more favorable climate and soil conditions for agriculture and pasture use, which, consequently, put them at higher risk of deforestation (Rosero-Bixby et al., 1996).

It can be recognized that grassland and secondary forest stretching from the north-west to the south-east expand to the perpendicular direction. More secondary forest developed probably due to the reforestation and forest law in the 1990s and appears in the later image. Obviously, a considerable amount of the primary forest was felled before acquisition of the second image. This phenomenon may be especially seen in the more remote areas north of the Piedras Blancas National Park. It could be explained by the population migration from North America to the south and the resulting population pressure in this area. It may not easily be seen directly from these two classified maps how much area changed from primary forest to other land use types. These changes are shown in figure 9. This map shows 3 classes of transition. Dark green indicates the areas where primary forests existed in 1987 and still remained as primary forests in 1998. Light green represents the primary forest

changed to grassland and secondary forest in the later image, which might result from pasture land extension. Orange signifies primary forest altered to other land use types, including arable land, settlements and so forth. The total area altered to grassland and secondary forest, light green, is 86,733 hectare, and the orange area, primary forest converted to other land use types, occupied 19,877 hectare. This was calculated from the number of pixels which have changed the class from primary to other land use types.

From this map, some information may be extracted. The light green and orange parts represent the total deforestation areas of primary forest. The areas with orange color, where the forests were converted due to anthropogenic activities, are smaller than the areas with light green color which signifies the pixels from primary forest in the image 1987.01.17 to grassland or secondary forest in the later image. Most of the converted areas belong to grassland or secondary forest. These areas extend along the developed areas from north-west to south-east. It might result from the population migration pressure and trend from North America to the South and the economic incentives converting forest to pasture land.

The primary forest still remains in the northern remote areas and in two national parks and the forest reserve in ACOSA. The inaccessibility of these remote areas due to high elevation and the legislation of forest law conserved the precious natural resources. Corcovado National Park, Golfo Dulce forest reserve, and Piedras Blancas National Park form a belt-like continuous primary forest from the peninsula around Golfo Dulce bay. It would be desirable if this belt-like primary forest would be connected to the primary forest area in the north. Unfortunately, the middle area, where the primary forests were mostly converted to grassland, has only residual fragmented primary forest existing between the northern more remote undisturbed part and the southern national parks.

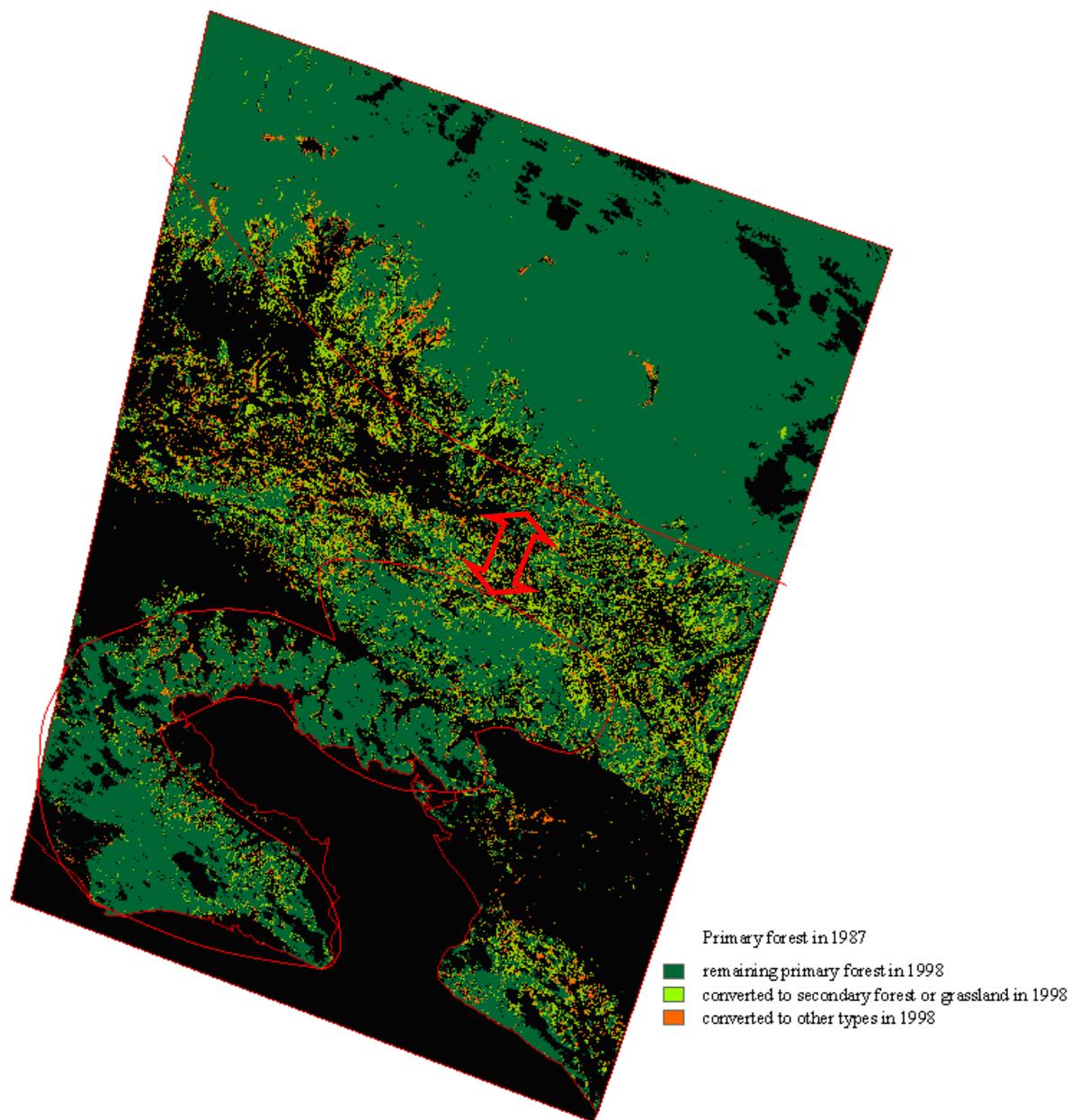


Fig. 9: Conversion of primary forest between 1987.01.17 and 1998.02.16

6 DISCUSSION

6.1 Sources of Errors

Several errors of these classified maps due to the complexity of the problem and due to the limited information content of the available reference data and of the remote sensing images can be mentioned.

There are several limitations of any classification of a tropical ecosystem due to the abundant biodiversity. The mixture of tropical tree species leads to inaccurate results. Tree age and stress can also cause variations of crown characteristics and lead to misclassification (Rahman, 2004). Even a pastureland site with intensive land use shows significant plant diversity (Oosterhoorn & Kappelle, 2000). Gjertsen (2007) mentioned that the kNN-method was strongly sensitive to changes of the proportions within the different classes in the reference dataset. With different forest stand compositions, e.g. 70 % for spruce or 50 % of deciduous trees, the pixels are not always assigned to the correct class. Our study area includes large areas and covers two conservation areas, **ACLA-P** and **ACOSA**. Even in the ACOSA conservation area, two national parks, Piedras Blancas and Corcovado National Parks, and one forest reserve, Golfo Dulce, also comprise many significant vegetation zones. The vegetation composition is complex even within the same class, which may show spatially continuous change, and may influence the accuracy of classification leading to misclassification.

The detailed vegetation map, which is employed as reference dataset presents the vegetation condition in 2003 and partially in 1998 in areas, which were covered by clouds in 2003 imagery. This means that there is a time difference between the vegetation map and both images, especially the image 1987.01.17. The land cover types change rapidly in tropical ecosystem, due to the natural dynamic succession.

Also, the forest law was changed in the 1990s. This can be considered as one factor impacting the vegetation composition. Unfortunately, this vegetation map was the only source for local land use reference information.

The SRTM elevation model, which was downloaded from GLCF, had spatial resolution 90m x 90m. Although the pixels of this model were resampled to the Landsat pixel size, the information content remains moderate. This problem was especially noticed in radiometric-topographic normalization using SCS correction, due to SRTM as reference data. Normalized images didn't reveal the perfect calibrated effect. Some small areas along ridges or valleys had erroneous elevation information and showed unsatisfactorily radiometric correction. The pixel values of these areas were not perfectly calibrated, which resulted in incorrect classification. Particularly in steep terrain, this problem was accentuated.

6.2 Possibilities of Improvement

Several possibilities may be suggested in order to enhance the classification results. SRTM with higher resolution may be first considered which may result in better radiometric-topographic correction. This may influence the following classification and increase the overall accuracy and Kappa.

In the kNN program, it is most important to have trustable reference points, which have to be representative for the spectroradiometric characteristics of each class and also for their variations within each class. The reference points in this study were generated according to the detailed vegetation map in the Piedras Blancas National Park. There were mainly two restrictions. First, as aforementioned, the map was compiled in 2003 and the existing time difference did disturb the correct land use information. Unfortunately, this map was the only information for the vegetation type in this study.

The second obstacle was that this vegetation map only covered the area of the Piedras Blancas National Park. This problem may be alleviated by visual selection and interpretation of more reference points in other areas.

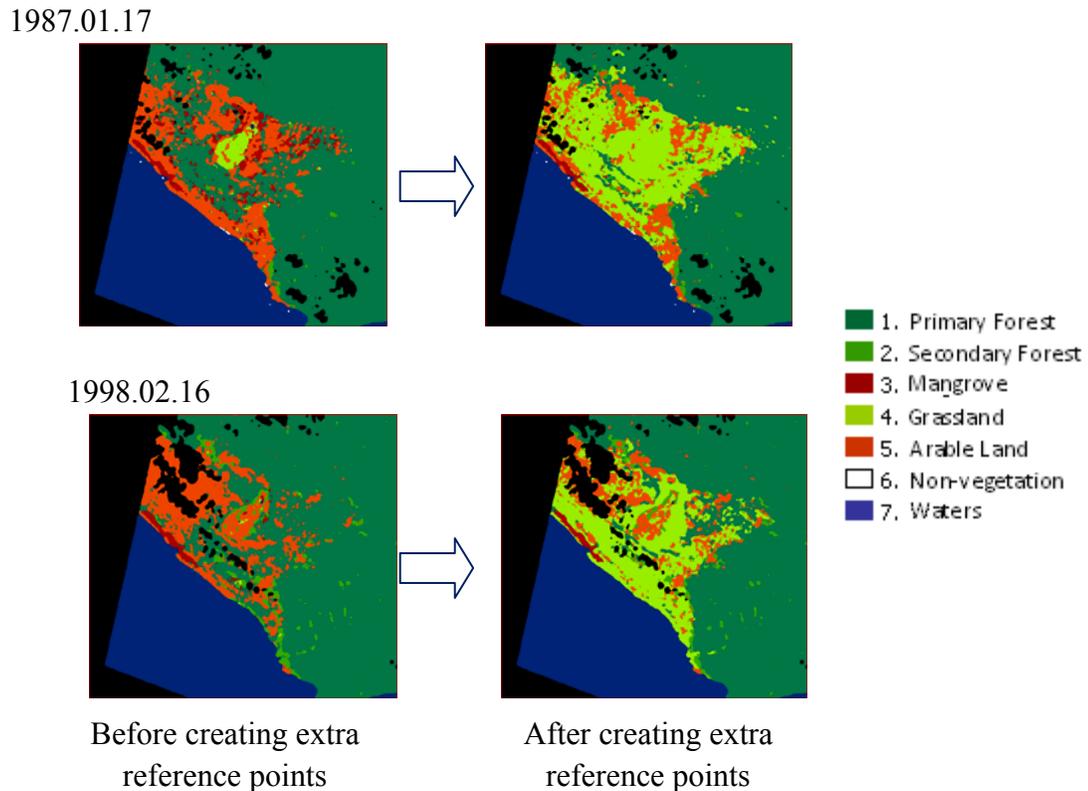


Fig. 10: The misclassification of class arable land in south-west of Peninsula, and the results after increasing additional reference points.

An example of the usefulness of additional reference points is shown in the figure 10. The problem is that large areas are erroneously classified as oil palm (in class “arable land”) in swampy areas. After hints from a botanic expert that *Raphia* palms are natively growing in these swampy areas (to be classified as “grassland”), the *Raphia* palms could be seen in high resolution images in Google Earth. Additional reference points were selected there. The classification results could be considerably improved as can be seen in the right side of figure 10.

The use of special additional reference points for verification may also help to obtain more realistic figures on the accuracy (overall accuracy, Kappa), as the sole use of training reference points for accuracy assessment produces overoptimistic accuracy figures.

In addition to the proposals made above, some further improvements may be suggested in the kNN program. Haapanen et al. in their study (2004) showed that the distance constraints (restricting maximum horizontal and vertical search radii) can improve the classification to be more effective. However, the reference points of this study area are mainly located in the Piedras Blancas National Park. The horizontal and vertical search radius functions therefore are unsatisfactory in this study. If more reference points representative for all classes and well distributed over the whole area to be mapped are available, it may be possible to effectively use horizontal and vertical search radii (as discussed in chapter 4.4) in the classification – which was not possible in this study due to the lack of well distributed reference points.

6.3 Suggestions of Wildlife Corridor

From figure 9, it can be seen that primary forest still remains in the northern remote area and in two national parks and a forest reserve in ACOSA (southern half of figure 9). Obviously, the Corcovado National Park, Golfo Dulce forest reserve, and the Piedras Blancas National Park form a belt-like continuous primary forest around Golfo Dulce bay. This belt-like primary forest without disturbance provides local wildlife natural habitats and food resources and is essential to maintain the local ecological environment and stabilize the biodiversity in this area. Unfortunately, this forest belt is not fully connected to the primary forest in the northern remote area. This route was disturbed and interrupted by the deforestation in between. It is to be considered to develop forest reserves or protection areas between the north high elevation area and the southern forest belt, to connect these two ecosystems, to provide more living space for wildlife and to thus preserve biodiversity.

7 ABSTRACT

The aim of this thesis is to develop a method for producing a vegetation map using multispectral satellite images from Landsat. Two dates of Landsat images are applied, one from 1987 and another from 1998, in order to detect the change from primary forest to other land cover classes during this period. The images are preprocessed by georeferencing and SCS radiometric-topographic normalization. They are then processed by kNN classification including outlier elimination. The method is developed, demonstrated and verified by using a vegetation map of the Piedras Blancas National Park (Costa Rica) compiled in 2003 as reference information. Information on forest change between 1987 and 1998 is deduced from the individual classifications of the two images. A deforestation trend is shown. Suggestions for designating a wildlife corridor in Middle America are also discussed.

8 REFERENCES

- CASARETTO, C., 2004, Land cover changes and fragmentation patterns in ACOSA – COSTA RICA. *NWS-I-2005-1*.
- FAO, 2005, Global Forest Resources Assessment. Available at: <http://www.fao.org/forestry/country/32185/en/cr/>. (accessed June 2010)
- FINLEY, A. O. and McRoberts, R. E., 2008, Efficient k-nearest neighbor searches for multi-source forest attribute mapping. *Remote Sensing of Environment*, 112, pp. 2203-2211.
- FINLEY, A. O., McRoberts, R. E. and EK A., 2006, Applying an efficient k-Nearest Neighbor search to forest attribute imputation. *Forest Science*, 52, pp. 130-135.
- FRANCO-LOPEZ, H., EK, A. R. and BAUER M., 2001, Estimation and mapping of forest stand density, volume, and cover type using k-nearest neighbors method. *Remote Sensing of Environment*, 77, pp. 251-274.
- GJERTSEN, A. K., 2007, Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sensing of Environment*, 110, pp. 420–430.
- GLCF (Global Land Cover Facility). Available at: <http://glcfapp.glc.f.umd.edu:8080/esdi/index.jsp>. (accessed March 2010)
- GU, D. and GILLESPIE, A., 1998, Topographic normalization of Landsat TM images of forest based on subpixel Sun-Canopy-Sensor geometry. *Remote Sensing of Environment*, 64, pp.166-175.
- HAAPANEN, R., EK, A. R., BAUER, M. E. and FINLEY, A. O., 2004, Delineation of forest/nonforest land use classes using nearest neighbor methods. *Remote Sensing of Environment*, 89, pp. 265-271.
- HAYES, D. J. and SADER S. A., 2001, Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series. *Photogrammetric Engineering & Remote Sensing*, 67, 1067-1075.
- HOLL, K. D., LOIK, M. E., LIN, E. H. V. and SAMUELS, I. A., 2000, Tropical montane forest restoration in Costa Rica: overcoming barriers to dispersal and establishment. *Restoration Ecology*, 8, pp.339-349.
- HOLMSTRÖM, H., KALLUR, H. and STAHL, G., 2003, Cost-plus-loss analyses of forest inventory strategies based on kNN assigned reference sample plot data. *Silva Fennica*, 37, pp. 381-398.
- INBio, National Biodiversity Institute of Costa Rica, Available at: http://www.inbio.ac.cr/en/biod/bio_biodiver.htm. (accessed June 2010)

- KANE, V. R., GILLESPIE, A. R., MCGAUGHEY, R., LUTZ, J. A., CEDER, K. and FRANKLIN, J. F., 2008, Interpretation and topographic compensation of conifer canopy self-shadowing. *Remote Sensing of Environment*, 112, pp. 3820-3932.
- KATILA, M. and TOMPPONEN, E., 2001, Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment*, 76, pp. 16-32.
- KOUKAL, T., SUPPAN, F., and SCHNEIDER, W., 2005, Radiometric-topographic normalization in mountainous terrain for Landsat-TM-based forest parameter assessment by the kNN method. *Oluic, New Strategies for European Remote Sensing, Proceedings of the 24th Symposium of the European Association of Remote Sensing Laboratories (EARSeL)*, Millpress, Rotterdam.
- KOUKAL T., SUPPAN F. and SCHNEIDER W., 2007, The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes. *Remote Sensing of Environment*, 110, pp. 431-437.
- KOUKAL, T., 2009, User Guide to the kNN-Program.
- LAURANCE, W. F., 2007, A New Initiative to Use Carbon Trading for Tropical Forest Conservation. *BIOTROPICA*, 39, pp. 20–24.
- LILLESAND, T., KIEFER, R. W., and CHIPMAN, J. W., 2007, Remote sensing and image interpretation, sixth edition. John Wiley & Sons, Inc.
- OOSTERHOORN, M. and KAPPELLE, M., 2000, Vegetation structure and composition along an interior-edge-exterior gradient in a Costa Rican montane cloud forest. *Forest Ecology and Management*, 126, pp. 291-307.
- PFÄFF, A. S. P. and SANCHEZ-AZOFEIFA, G. A., 2004, Deforestation pressure and biological reserve planning: a conceptual approach and an illustrative application for Costa Rica. *Resource and Energy Economics*, 26, pp. 237-254.
- PRADER, W., 2008, Tree diversity and vegetation dynamics of a one hectare forest plot census in the lowland rain forests of the Piedras Blancas National Park ("Regenwald der Österreicher"), Costa Rica. *Diplomarbeit-University of Vienna*.
- RAHMAN, M. M., CSAPLOVICS, E., KOCH, B. and KÖHL, M., 2004, Interpretation of tropical vegetation using Landsat ETM+ imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (ed. Altan, M. O.), Istanbul, Part B-YF, pp. 157-162.
- REESE, H., GRANQVIST-PAHLEN, T., EGBERTH, M., NILSSON, M. and OLSSON, H. 2005. Automated estimation of forest parameters for Sweden using Landsat data and the kNN algorithm. *The 31st International Symposium on Remote Sensing and Environment, ISRSE'05*.

- RIANO, D., CHUVIECO, E., SALAS, S. J., and AGUADO, I., 2003, Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types. *IEEE Transactions on Geoscience and Remote Sensing*, 41, pp. 1056-1061.
- ROSETO-BIXBY, L. and PALLONI, A., 1996, Population and deforestation in Costa Rica. *CDE Working Paper No. 96-19*
- SANCHEZ-AZOFEIFA, G. A., RIVARD, B., CALVO, J. AND MOORTHY, I., 2002, Dynamics of tropical deforestation around national parks: remote sensing of forest change on the Osa Peninsula of Costa Rica. *Mt. Res. Dev.*, 22, pp. 352–358.
- SANCHEZ-AZOFEIFA, G. A., DAILY, G. C., PFAFF, A. S. P. and BUSCH, C., 2003, Integrity and isolation of Costa Rica's national parks and biological reserves: examining the dynamics of land-cover change. *Biological Conservation*, 109, pp. 123-135.
- SANCHEZ-AZOFEIFA, G. A., HARRISS, R. C. and SKOLE, D. L., 2001, Deforestation in Costa Rica: A quantitative analysis using remote sensing imagery. *Biotropica*, 33, pp. 378-384.
- SOENEN, S. A., PEDDLE, D. R. and COBURN, C. A., 2005, SCS+C: A modified Sun-Canopy-Sensor topographic correction in forested terrain. *IEEE Transactions on Geoscience and Remote Sensing*, 43, pp. 2148 - 2159
- THESSLER, S., SESNIE, S., BENDANA, Z. S. R., RUOKOLAINEN, K., TOMPPA, E. and FINEGAN, B., 2008, Using k-nn and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica. *Remote Sensing of Environment*, 112, pp. 2485-2494.
- The World Bank OED, 2000, Costa Rica forest strategy and the evolution of land use-evaluation country case study series. *The World Bank*.
- TOMPPA, E. and HALME, M., 2004, Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. *Remote Sensing of Environment*, 92, pp. 1-20.
- UNFCCC (United Nations Framework Convention on Climate Change), Subsidiary body for scientific and technological advice, 2006, Issues relating to reducing emissions from deforestation in developing countries and recommendations on any further process. Twenty-fourth session, paper No.6: Costa Rica.
- VELDKAMP, A. and FRESCO, L. O., 1997, Reconstructing land use drivers and their spatial scale dependence for Costa Rica (1973 and 1984). *Agricultural Systems*, 55, pp. 19-43.
- WEISSENHOFER, A., HUBER, W., KOUKAL, T., IMMITZER, M., SCHEMBERA, E., SONTAG, S., ZAMORA, N. and WEBER, A., 2008, Ecosystem diversity in the Piedras Blancas National Park and adjacent areas (Costa Rica), with the first vegetation map of the area. *Neue Serie*, 88, pp. 65-96.

9 APPENDICES

Table 4: Crossvalidation results of image1987.01.17 with 12 pre-classes, k=6

CONFUSION MATRIX CATEGORICAL VARIABLE 1

	Cat1	Cat2	Cat3	Cat4	Cat5	Cat6	Cat7	Cat8	Cat9	Cat10	Cat11	Cat12	Sum	user's Acc.
Cat1	830	116	7	2	3	0	3	1	2	2	0	0	966	0.859
Cat2	59	163	1	8	3	4	2	6	9	3	0	0	258	0.632
Cat3	3	2	136	0	6	0	0	0	0	0	0	0	147	0.925
Cat4	4	5	0	53	17	13	8	4	15	11	0	0	130	0.408
Cat5	8	4	4	20	60	4	0	6	3	5	0	0	114	0.526
Cat6	1	4	0	6	1	156	1	1	2	7	0	2	181	0.862
Cat7	0	2	0	0	1	0	0	0	0	0	0	0	3	0
Cat8	12	30	0	8	45	3	5	386	8	4	0	0	501	0.770
Cat9	1	3	0	13	0	3	1	1	53	1	0	0	76	0.697
Cat10	0	0	0	2	3	1	0	0	2	44	1	11	64	0.688
Cat11	0	0	0	0	1	0	0	0	0	0	28	0	29	0.966
Cat12	0	0	0	0	0	0	0	0	0	2	0	25	27	0.926
Sum	918	329	148	112	140	184	20	405	94	79	29	38	2496	
producer's Acc.	0.9	0.5	0.9	0.5	0.4	0.8	0	1	0.6	0.56	1	0.66		

Kappa = 0.717 Overall Accuracy = 0.775

Table 5: Crossvalidation results of image 1998.02.16 with 12 pre-classes, k=6

CONFUSION MATRIX CATEGORICAL VARIABLE 1

	Cat1	Cat2	Cat3	Cat4	Cat5	Cat6	Cat7	Cat8	Cat9	Cat10	Cat11	Cat12	Sum	user's Acc.
Cat1	811	89	2	0	11	0	2	1	2	0	0	0	918	0.883
Cat2	40	142	0	3	3	7	3	7	4	0	0	0	209	0.679
Cat3	3	0	57	0	1	0	0	0	1	0	0	0	62	0.919
Cat4	0	0	0	107	12	15	0	0	8	10	0	0	152	0.704
Cat5	6	2	3	13	50	1	0	2	1	1	0	0	79	0.633
Cat6	0	1	0	8	1	119	1	1	3	4	0	0	138	0.862
Cat7	0	1	0	1	2	1	2	0	3	0	0	0	10	0.2
Cat8	10	15	0	0	9	4	2	208	2	0	0	0	250	0.832
Cat9	0	2	0	4	7	4	5	0	53	0	0	0	75	0.707
Cat10	0	0	0	1	1	0	0	0	0	43	0	3	48	0.896
Cat11	0	0	0	0	1	0	0	0	1	0	28	0	30	0.933
Cat12	0	0	0	0	0	0	0	0	0	3	0	20	23	0.870
Sum	870	252	62	137	98	151	15	219	78	61	28	23	1994	
producer's Acc.	0.9	0.6	0.9	0.8	0.5	0.79	0.1	0.9	0.7	0.7	1	0.87		

Kappa =
0.765

Overall Accuracy =
0.822

Table 6: Crossvalidation results of image 1987.01.17 with 7 classes, k=6

CONFUSION MATRIX CATEGORICAL VARIABLE 1

	Cat1	Cat2	Cat3	Cat4	Cat5	Cat6	Cat7	Sum	user's Acc.
Cat1	830	116	7	5	5	2	0	965	0.860
Cat2	59	160	1	11	14	3	0	248	0.645
Cat3	3	2	136	6	0	0	0	147	0.925
Cat4	14	16	4	343	42	27	2	448	0.766
Cat5	12	35	0	70	456	5	0	578	0.789
Cat6	0	0	0	1	2	80	1	84	0.952
Cat7	0	0	0	0	0	0	26	26	1
Sum	918	329	148	436	519	117	29	2496	
producer's Acc.	0.904	0.486	0.919	0.787	0.879	0.684	0.897		

Kappa = 0.755 Overall Accuracy = 0.814

Table 7: Crossvalidation results of image 1998.02.16 with 7 classes, k=8

CONFUSION MATRIX CATEGORICAL VARIABLE 1

	Cat1	Cat2	Cat3	Cat4	Cat5	Cat6	Cat7	Sum	user's Acc.
Cat1	818	96	2	10	4	0	0	930	0.880
Cat2	31	135	0	11	9	0	0	186	0.726
Cat3	2	0	57	1	1	0	0	61	0.934
Cat4	6	4	3	332	21	15	0	381	0.871
Cat5	13	17	0	31	277	0	0	338	0.820
Cat6	0	0	0	0	0	69	0	69	1
Cat7	0	0	0	1	0	0	28	29	0.966
Sum	870	252	62	386	312	84	28	1994	
producer's Acc.	0.940	0.536	0.919	0.860	0.888	0.821	1		

Kappa = 0.806 Overall Accuracy = 0.861

Fig.11: Pre-classified vegetation map of image 1987.01.17

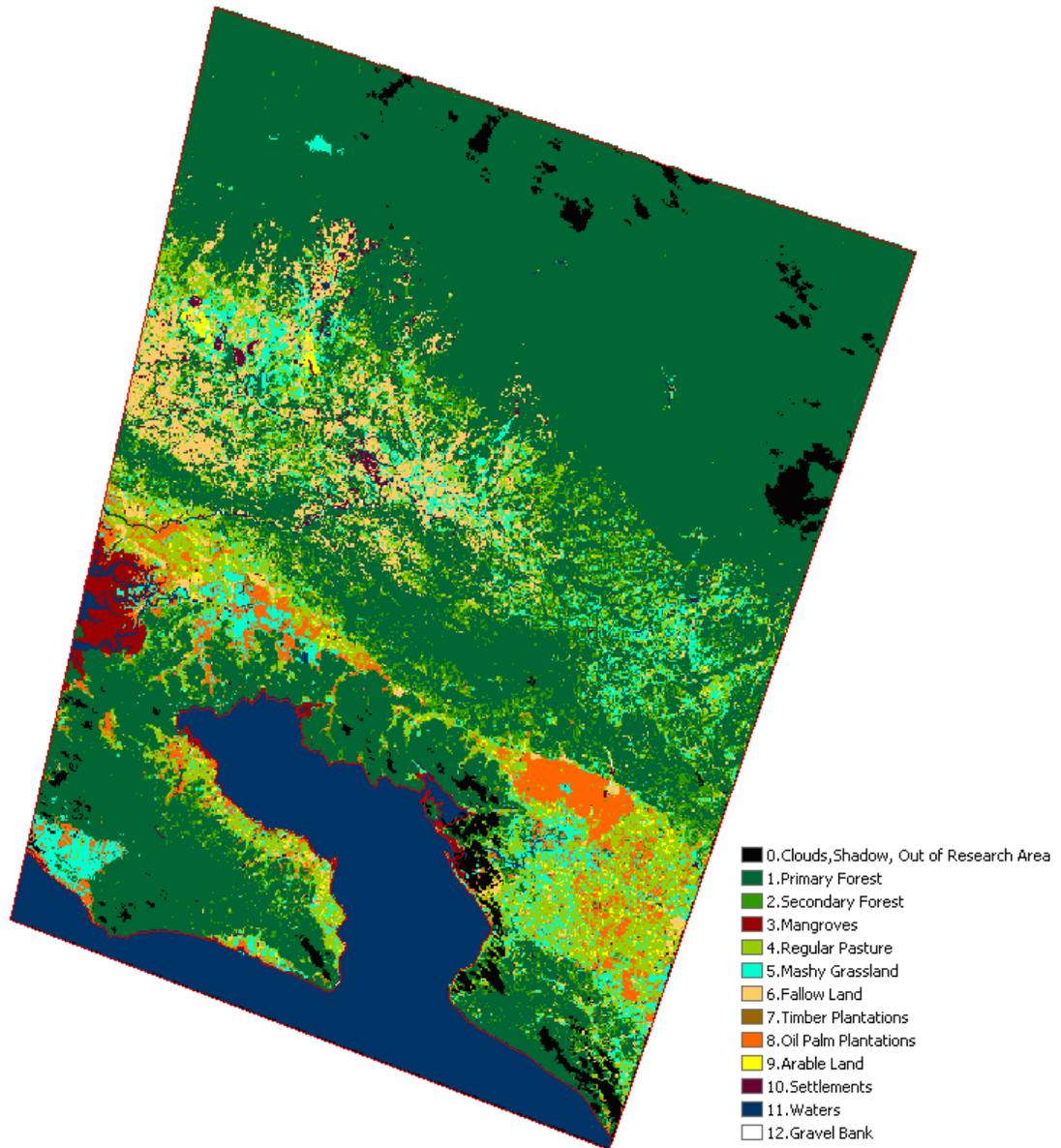
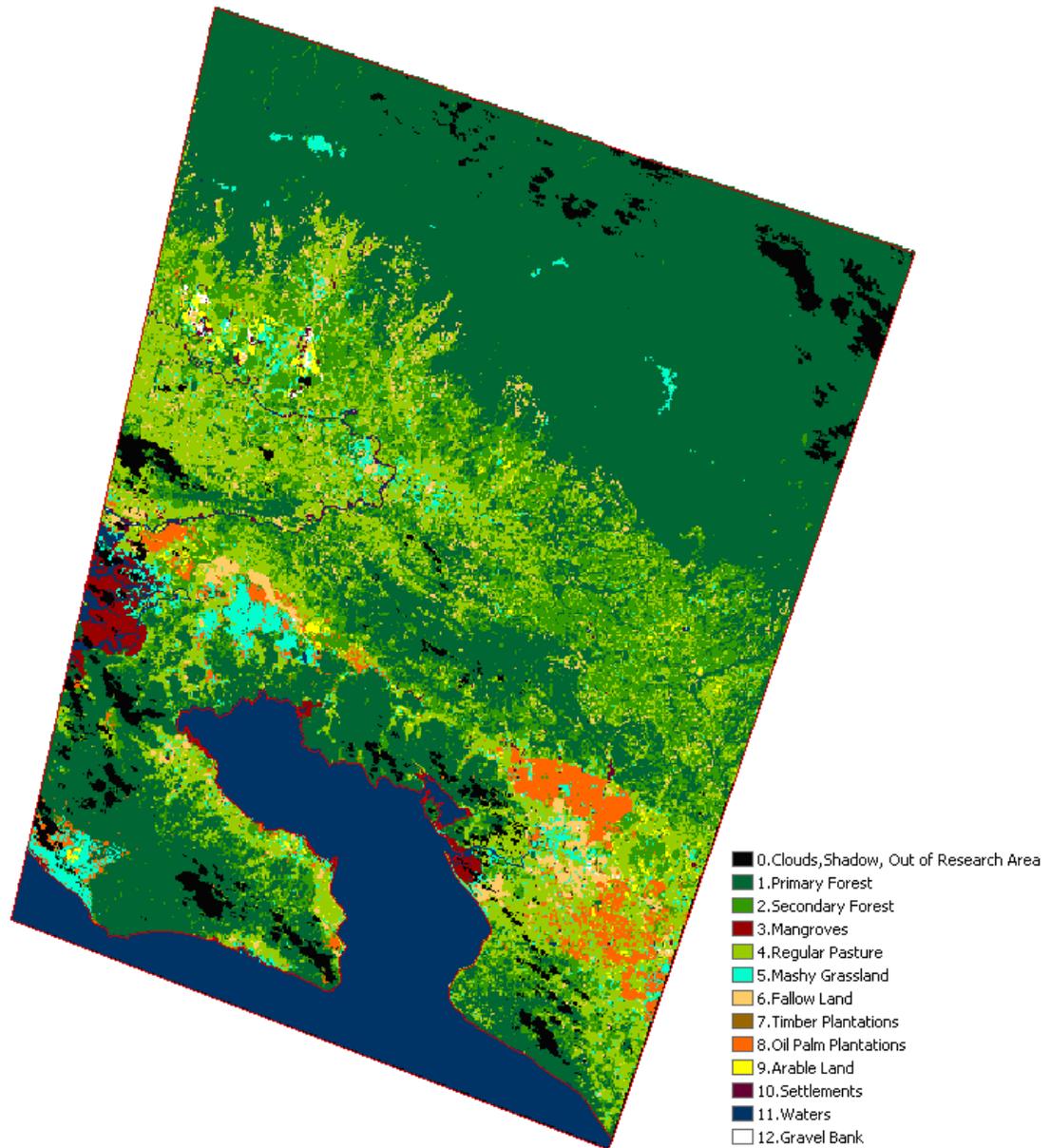


Fig.12: Pre-classified vegetation map of image 1998.02.16



10 TABLE OF FIGURES

Figure 1:Deforestation in Costa Rica (UNFCCC, 2006)	6
Figure 2:11 conservation areas in Costa Rica and study area.....	7
Figure 3:Flow diagram of processing	13
Figure 4:Diagram of kNN Algorithm in Feature Space.....	18
Figure 5: kNN-program flow chart (Koukal, 2009).....	23
Figure 6:Results of vegetation classification in the Piedras Blancas National Park...35	
Figure 7:Classification map of image 1987.01.17	37
Figure 8:Classification map of image 1998.02.16	38
Figure 9:Conversion of primary forest between 1987.01.17 and 1998.02.16.....	41
Figure 10:The The misclassification of class arable land in south-west of Peninsula, and the results after increasing additional reference points	42
Figure 11:Pre-classified vegetation map of image 1987.01.17.....	53
Figure 12:Pre-classified vegetation map of image 1998.02.16.....	54

11 TABLE OF TABLES

Table 1: Image data.....	10
Table 2: Symbols used in SCS correction models	16
Table 3: Categories of the classification	27
Table 4: Crossvalidation results of image 1987.01.17 with 12 pre-classes, k=6.....	50
Table 5: Crossvalidation results of image 1998.02.16 with 12 pre-classes, k=6.....	51
Table 6: Crossvalidation results of image 1987.01.17 with 7 classes, k=6	52
Table 7: Crossvalidation results of image 1998.02.16 with 7 classes, k=8	52