INCREASING THE ACCURACY OF VEGETATION CLASSIFICATION USING GEOLOGY AND DEM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AYŞEGÜL DOMAÇ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

DECEMBER 2004

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Canan ÖZGEN
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Zuhal Akyürek
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. C. Can Bilgin	Assist. Prof. Dr. M. Lütfi Süzen
Co-Supervisor	Supervisor

Examining Committee Members

Prof. Dr. Vedat Toprak	(METU, GEOE)	
Assist. Prof. Dr. M.Lütfi Süzen	(METU, GEOE)	
Assoc. Prof. Dr. C. Can Bilgin	(METU, BIOL)	
Assist. Prof. Dr. Zuhal Akyürek	(METU, GGIT)	
Assist. Prof. Dr. Mustafa Türker	(METU, GGIT)	

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Ayşegül DOMAÇ

Signature :

ABSTRACT

INCREASING THE ACCURACY OF VEGETATION CLASSIFICATION USING GEOLOGY AND DEM

Domaç, Ayşegül

M. Sc., Department of Geodetic and Geographic Information Technologies
Supervisor: Assist. Prof. Dr. M. Lütfi Süzen
Co-Supervisor: Assoc. Prof. Dr. C. Can Bilgin

December 2004, 80 pages

The difficulty of gathering information on field and coarse resolution of Landsat images forced to use ancillary data in vegetation mapping. The aim of this study is to increase the accuracy of species level vegetation classification incorporating environmental variables in the Amanos region. In the first part of the study, coarse vegetation classification is attained by using maximum likelihood method with the help of forest management maps. Canonical Correspondence analysis is used to explore the relationships among the environmental variables and vegetation classes. Discriminant Analysis is used in the second part of the study in two different stages. Firstly Fisher's linear equations for each of the previously defined nine groups calculated and the pixels are included in one of these groups by looking at the probability of that pixel being in that group. In the second stage Distance raster value of maximum likelihood classification is used. Distance raster pixels having a value less than one is accepted as misclassified and replaced with a value of first stage result of that pixel. As a result of this study 19.6 % increase in the overall accuracy is obtained by using the relationships between environmental variables and vegetation distribution.

Keywords: Vegetation Classification, Amanos Region, Discriminant Analysis, Ancillary Data

ÖZ

BİTKİ ÖRTÜSÜ SINIFLANDIRMASININ DOĞRULUĞUNU JEOLOJİ VE SAYISAL YÜKSEKLİK MODELİ KULLANARAK ARTTIRMA

Domaç, Ayşegül

Yüksek Lisans, Jeodezi ve Coğrafi Bilgi Teknolojileri
Tez Yöneticisi: Yrd. Doç. Dr. M. Lütfi Süzen
Ortak Tez Yöneticisi: Doç. Dr. C. Can Bilgin

Aralık 2004, 80 sayfa

Araziden veri toplamanın zorluğu ve Landsat görüntülerinin çözünürlüğünün yetersiz olması, vejetasyon haritalaması çalışmalarında yardımcı katmanların kullanılmasını gerekli kılmıştır. Bu çalışmanın amacı çevresel faktörleri kullanarak Amanos bölgesinde yapılan tür bazlı vejetasyon sınıflandırmasının doğruluğunun arttırılmasıdır. Çalışmanın ilk kısmında Maksimum Benzerlik metodu kullanılarak, orman işletme haritalarının yardımıyla kaba ölçekte bir vejetasyon haritası yapmıştır. Kanonik Karşılık Analizi ile çevresel değişkenlerin vejetasyon ile ilişkileri araştrılmıştır. İkinci kısımda ise Diskriminant analizi iki ayrı aşamada kullanılarak daha detaylı vejetasyon sınıflandırma haritaları elde edilmiştir. Önceden belirlenmis dokuz grup için Fisher doğrusal denklemleri hesaplanmış ve hücreler ait olma olasılıkları yüksek olan sınıflara atanmışlardır. İkinci aşamada ise Maksimum Benzerlik sonucunda elde edilen Uzaklık Matrisi kullanılmıştır. Uzaklık Matrisi değerleri birden küçük olan hücrelerin yanlış sınıflandırıldığı kabul edilmiş ve bu hücreler ilk aşama sonucundaki sınıf değeriyle değiştirilmiştir. Bu çalışmanın sonucunda, çevresel faktörlerle vejetasyon dağılımı arasındaki ilişkiyi kullanarak genel sınıflandırmanın doğruluğu 19.6% oranında arttırılmıştır.

Anahtar kelimeler: Vejetasyon Sınıflandırması, Amanos Bölgesi, Diskriminant Analizi, Yardımcı Veri

To My Family

ACKNOWLEDGMENTS

First of all I would like to express my gratitude to my supervisor Assist. Prof. Dr. Mehmet Lütfi Süzen, for his continuous support, invaluable guidance and endless trust. His unbelievable patience and motivation provided me to overcome all the problems at every stage of the thesis.

I would like to thank my co-supervisor Assoc. Prof. Dr. C. Can Bilgin for his guidance and valuable suggestions at the stages where I really got stuck.

I would like to thank Prof. Dr. Vedat Toprak for becoming my shadow co-supervisor. He brought the peace of mind by knowing that I could ask for his help whenever I found myself in trouble.

I would like to thank Assoc. Prof. Dr. Bora Rojay, for sharing his endless geological knowledge with me.

I would like to express my gratefulness to S. Uğur Zeydanlı, for believing on me more than I did; for being my dear friend, teacher, and close colleague.

I would like to thank my close friend Çağıl Kolat for not leaving me alone in the long nights, sharing her knowledge with me, and making me laugh when I get upset. Also I would like to include my dearest friends Mine Ulusoy and Nilhan Çiftçi for the never ending support on a personal basis in everything I involve myself into.

I would like to thank WWF-Turkey, especially Yıldıray Lise for supplying the vital data to initiate my thesis and for sharing his knowledge about the study area.

I would like to thank GGIT Department, for their contribution throughout my graduate studies and during the thesis.

I would like to thank to RS-GIS laboratory team at the Geological Engineering Department for their priceless friendship and the support they offered me during the development of my thesis.

I would like to thank to Ecology laboratory team at METU, Biology Department for encouraging me to carry on in the difficult moments and for their support, motivation and happiness at the vital stages throughout this study.

I would also like to thank to my parents, my brother and my Nino, for their endless support and belief, not only in this study but also throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1. Purpose and Scope

Throughout the prolonged use of satellite images, mapping earth features and infrastructure, researching environmental changes, gathering information about land use, land cover and vegetation information at different spatial scales and temporal resolutions became easier in the last few decades. Amongst these macro activities, thoroughly classification processes seek to identify land cover classes range from broad life-form categories to narrow floristic classes (Carpenter et al., 1997).

Remote sensing has been used in vegetation research for many years to map the vegetation cover, forest fires, and detecting the change in the vegetation through different periods (Janssen et. al., 1990; Dorren et. al., 2003; Pfeffer et. al., 2003). Decrease in the prices of satellite images, easy access to current images, improvements in image processing software's, and increased CPU speed of the computers flourished the use of remote sensing in ecological studies not very different from the rest of the earth science related research. Mapping vegetation by using remote sensing is a widely used technique in ecological research since it could determine the distribution, formation, and change of vegetation for very large areas in a short time, moreover this offers the possibility to extrapolate results of mapping, especially in large and hardly accessible areas (Hoersch et. al., 2002).

Among various satellite images, Landsat TM or Landsat ETM images have some advantages over the rest of the satellites in vegetation classification. Such that 30 m ground resolution accepted as a convenient resolution for regional vegetation mapping studies with a minimum mapping unit of 100 ha; the spectral coverage fits well to the vegetation spectra; and the swath width yields in less number of images to process which maintains the coherence of the imagery. Although the above is concentrated on the fitness of Landsat system for vegetation mapping, the system is still inadequate as the results of the classification studies are not sufficient to identify the individual plant species other than groups of trees (Franklin, 1995; Nagendra, 2001; Pfeffer et. al., 2003). For the regional scale vegetation mapping studies, spectral resolutions of current images might seem to be sufficient, however, when leveling up the regional mapping campaign into a local scale, either the spectral or spatial resolution of data retrieved from common commercial satellite images and aerial photographs are insufficient. Although recent innovative approaches in remote sensing society such as the usage of LIDAR and Hyperspectral sensors theoretically solves this inadequacy, the inherent problems of vegetation classification studies such as the closeness and different crown sizes still hamper the regional vegetation mapping problem (Dorren et. al., 2003; Pfeffer et. al., 2003).

At this point, addition of some extra information is vital to provide some insight in the classification of heterogeneous vegetation cover. Incorporation of the relevant information into the classification process can increase the discrimination possibilities for cover types with irregular, overlapping spectral signatures (Maselli et. al.; 1995). Pre-classification scene stratification, post-classification class sorting and classification modifications are the common approaches used to combine ancillary data for classification to improve the accuracy (Hutchinson, 1982). Information from ancillary data sources has been widely shown to aid discrimination of classes that are difficult to classify using remotely sensed data (Strahler, 1980).

Studies related with the relationship of species composition and forest structure to topographic and geological features are quite sparse in literature (Pinder et. al., 1997; Fahsi et. al., 2000; Tayku et. al., 2002). Correlations of species with environmental variables are used to get information about the distributions of vegetation in different scales, furthermore it is quite wise to incorporate the terrain attributes (e.g., elevation, gradient, local relief) into the statistical decision rules to classify ecological units using remote sensing data (Moore et. al., 1991).

Despite all of these issues large scale vegetation mapping studies are made available for some regions of Turkey, both by governmental and non governmental organizations. The vegetation classification of Aegean region is carried out by Ministry of Forestry and Environment while for Mediterranean and South Eastern Anatolia region WWF-Turkey had completed the task (Zeydanlı and Domaç, 2004).

The main objective of this study is to increase the accuracy of vegetation classification in the alliance level with Landsat 7 ETM by using ancillary data with the scope of evaluating the use of environmental variables and raw Enhanced Thematic Mapper

(ETM) band combinations to find out discriminating functions to determine nine vegetation classes in the postclassification sorting process.

1.2. Study Area

The study area is located in the southern portion of the Amanos Mountains (Figure 1.1), ranging from E 31° 05′ to E 36° 35′ and N 36°30′ to N 36° 58′, having a coastal zone to Mediterranean Sea. The main settlements of the study area are İskenderun, Dörtyol and Hassa.

Amanos Mountains are significantly important area in terms of plant diversity, as it harbors Black Sea enclave, which is the heritage of Ice Age, intact forest cover with deep and protected valleys, high endemism ratio, and typical representative of the diverse Eastern Mediterranean flora (http://www.wwf.org.tr/en/ormanlar_dunya_sn_ad.asp (visited on 23. 10.2004)).

In aspect of Geology, Kahramanmaraş and surrounding areas may be regarded as a key region to show the geological evolution of the South-east Anatolia, because it is a place where different tectonic units with a diversity of lithological units together with their relationships observed (Yilmaz et al, 1988).

1.3. Organization of the Thesis

In the first chapter of the thesis, the research problem was defined, the purpose and scope of this research was presented, the geographical setting of the project area was defined and the organization of this thesis was presented.

The second chapter includes the previously performed studies related with the use of ancillary data and statistical methods in the vegetation classification.

Third chapter explains the data used in this study. Related information is given about satellite images, geology, elevation, and slope and aspect layers.

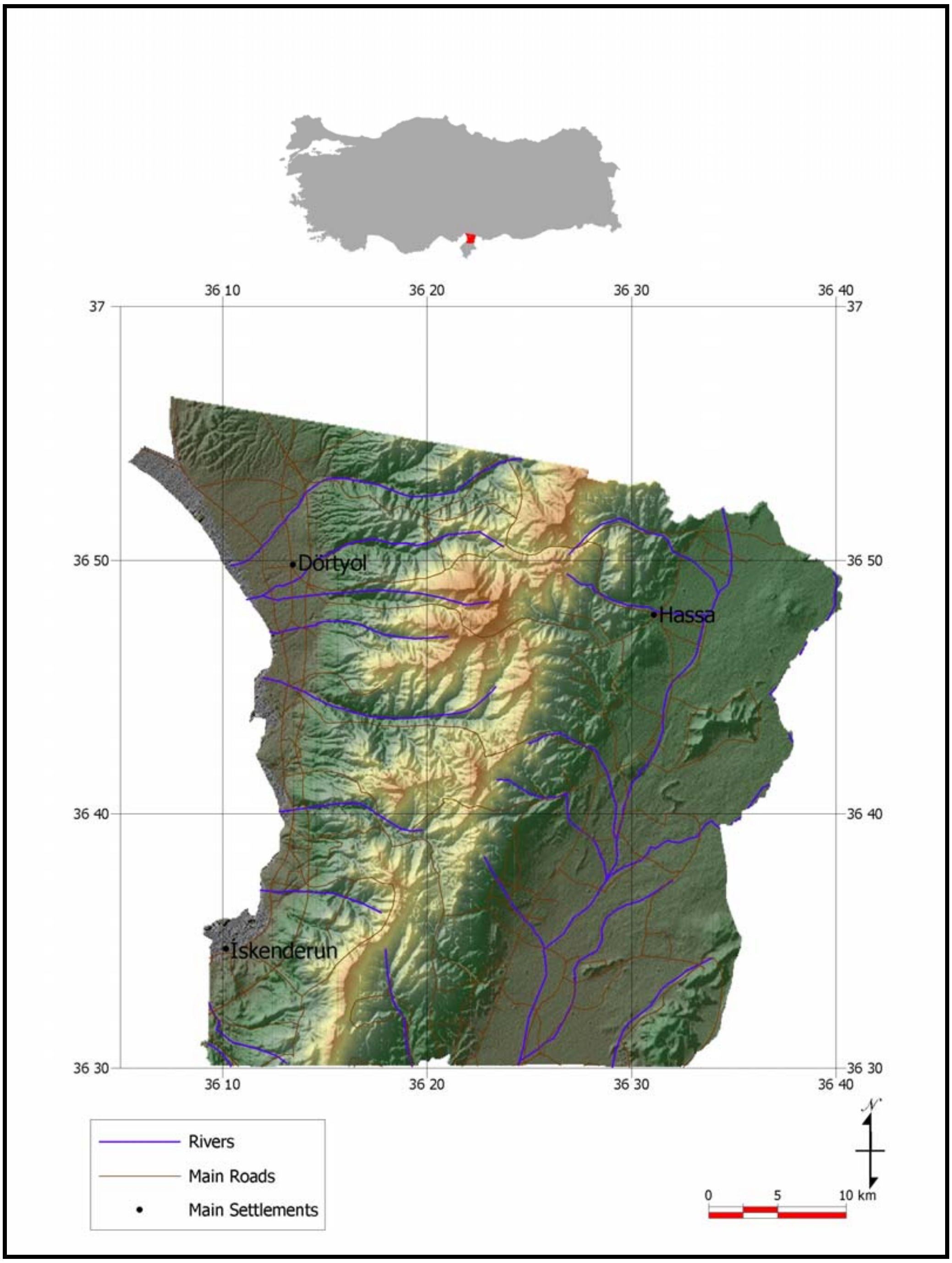

Figure 1.1. Location of the study area

Fourth chapter was composed of methods and analysis where the brief information was given in methodological snapshot. Details and results of coarse classification, fine classification and integration steps were presented with accuracy assessment values. The multivariate statistical analyses (i.e. Canonical Correspondence and Discriminant Analysis) used in statistical classification was explained and the relationships between species and variables was discussed. Finally discriminating functions extracted from the results of the analysis explained and presented.

In fifth chapter the results obtained from chapter four was discussed, while in sixth chapter the conclusions of this research thesis were presented.

CHAPTER 2

BACKGROUND INFORMATION

This chapter includes the selected previous researches related with the subject of this thesis. These researches were grouped under the headings of vegetation classification, use of ancillary data in the classification process and statistical methods used in the classification process.

2.1. Vegetation Classification with Remote Sensing

Remote sensing is a technique used to collect data from the Earth surface by measuring the reflected energy of the features on each part of the electromagnetic spectrum. Recent developments in remote sensing technology allow the use of images in land-cover and land-use changes, city planning, mapping of soil, geology, forestry and conservation planning moreover it has the potential to be used for continuous surveying of the Earth surface.

Digital image classification is the process of assigning a pixel (or groups of pixels) of remote sensing image to one of the previously defined land cover classes. This could be performed by various algorithms based on the spectral analysis of individual pixels; hence by using the spectral reflectance curves of earth materials. The maximum likelihood classifier is one of the most popular methods of classification in remote sensing. This method is preferred in classification process because unlike the minimum distance and the parallelepiped classifiers this technique takes into account the spectral variability both within and between classes (Fahsi et. al., 2000).

Since 1970's remote sensing have been used for the contribution to the forest mapping and among the satellites Landsat images are the most suitable one. However the spatial resolution of Landsat group imagery in many cases is low to derive a useful forest map, which yields in a low accuracy result by using per-pixel classification methods.

At this point, Dorren et. al. (2003) suggests two different methods to improve the accuracy of vegetation classification in steep mountainous terrain; the first is to assess topographic correction and to add DEM as an additional band, and the other is to use object-based classification instead of pixel-based classification. Among various topographic correction methods, Sun-Canopy-Sensor-correction (SCS-correction) performed the best in reducing the correlation between incidence angle and reflectance value. When the classification results of his trials were compared, the one with topographic correction and additional band used yielded the highest accuracy among them. In object based trial, first the image objects are created by means of segmentation. Then the forest mask, which was obtained by classification, was segmented into forest and non-forest objects. Classification was performed on two different levels. In level 2 large objects were identified, and in the first level forest types are separated into four different classes. In the object-based classification, smaller and larger objects produced lower accuracies whereas the best object-based was based on the segmented objects with an average size of 21.6 pixels.

Dymond et. al. (2002), compares the effectiveness of image differencing and vegetation indices to improve the forest classification with the input set of phenologically significant TM scenes. NDVI and Tasseled Cap indices (Brightness (B), Greenness (G), Wetness (W)) were computed using the TM image for each phenological period to test the effectiveness of indices to improve the forest classification. Besides, the changes in TM color composite 3-4-5 and each of the four indices values were subtracted from one phenological period to the next. The area was subdivided into smaller lands which reduced the number of categories and variation within each class. By using hybrid classification, vegetation type map was composed; whereas maximum likelihood method was used for the genera level classification. These procedures were repeated for 6 different input data sets. According to the results of this study the image differencing of the Tasseled Cap indices may have produce the best vegetation classification.

Liu et. al. (2002) tests the new integrated approaches like consensus builder system (CSB) and a combined expert system (CES) and neural network system (NNC) to improve the classification accuracy. First of the classifiers is Maximum Likelihood Classifier (MLC), in which each pixel is assigned to the class with the shortest modified "Mahalanobis distance" from the pixel to the class mean. The second classifier used in this study is NNC which is composed of two stages; training stage and classification stage. Once the training system is complete, the trained system is used for classification. The third classification system is the expert system classifier (ESC). The structure of this system composed of two parts. First part is the "knowledge base" to store expert knowledge, and rules, and the "inference engine" for system processing. Classification was performed by the tree individual classifier and two new integrated classifiers using the same training set. An integrated classifier called ESNNC produced the highest accuracy of 80% when compared with the individual classifiers.

Wang et. al. (2001), aims to generate a method to determine appropriate plot size and spatial resolution for mapping multiple vegetation using remote sensing data for large areas. There were six vegetation cover types which are different in spatial variability. The appropriate plot size and spatial resolution were studied for each vegetation type in order to capture the structures of spatial variability and to improve map accuracy. Semi-variogram method was used to model spatial variability. If there is a high correlation between field and image data, the appropriate plot size obtained using the field data will be consistent with the appropriate spatial resolution using the images. The comparison of the vegetation classification at different plot and image sizes by cross validation further proved the appropriate spatial resolution. The appropriate plot size was about 60 m for grass and shrub, 70 m for forbs, and 80 m for tree and half-shrub, and would not be less than 80 m for wood; and the TM images led to an appropriate spatial resolution of 90 m in this study.

Treitz and Howarth (2000) determined an extent to which texture and terrain variables could improve the forest ecosystem mapping within the boreal forest of north-western Ontario. Linear Discriminant analysis (LDA) was performed within a program to explore the discriminatory power of (1) spectral-spatial, (2) texture, (3) terrain, (4) combination of these features. This method first derives a transform that minimizes the ratio of difference between group multivariate means and their within group multivariate means and also this is used to find a discriminant function. Discriminant analysis as a classification technique has been shown to be less sensitive to the number of variables and deviations from the normal (Gaussian) distribution as opposed to other methods such as maximum likelihood.

2.2. Ancillary Data in Classification Process

Shrestha and Zinc (2001), assumed that the amount of energy received by the sensor depends on slope gradient, and exposition with respect to sun elevation. In this study intensity normalization for the spectral bands for removing the variations in the solar illumination angle is implemented. Because of the topographic effects, training samples results in elongated clusters. After normalization of the bands elongation disappears and the training samples could be assumed to be approximately normally distributed. Maximum likelihood method was performed on the bands both before and after normalization. Overall accuracy of the final classification was increased to 94 % from 67 %. It is concluded that the variation on the solar illumination angle could be easily corrected by the normalization of the bands by using total intensity.

De Bruin and Gorte (2002) used geologic map units to improve land cover classification. In classification process, use of prior probabilities related to slope and aspect classes improves the classification accuracy, over the obtained with spectral data alone (Strahler, 1980). The objective of their study was to demonstrate: a) The use of stratification combined with iterative estimation of prior probabilities to improve classification accuracy; b) The use of posterior probability vectors to represent uncertainty in image classifications and in the result of subsequent analysis and c) To reduce the number of geological units, increase the minimum size, and to harmonize the level of detail while the original map was generalized. Two Landsat images were classified independently using iteratively estimated prior probabilities per geologic unit. Two main results of this study is as follows: a) Iterative estimation of prior probabilities after stratification according to map units carrying information relevant to the classification theme provides a practical approach to improve classification accuracy and b) Posterior probability of class membership provides useful estimates of the magnitude and spatial distribution of local uncertainty in the classification results.

Fahsi et. al. (2000), present that, DEM could be used to improve the classification accuracy by reducing the effect of relief on satellite images. In higher relief, the pixels of same cover type could have different spectral value and the pixels of different cover types could have a same spectral value because of the effect of the topography. Three components are affecting the spectral radiance of the surface yielding in direct radiation, diffuse radiation, and the radiation reflected from the adjacent surfaces. Topographical shapes of the surfaces are the main factor and have influences on these components. In order to normalize the topographic effect on each pixel a model proposed by Yang et. al. (1993) were used in the study. Statistically, removal of the topographic effect is accepted as successful if, the spectral variability within the same cover type extending over varying terrain is substantially reduced after image correction. Principal Components, vegetation index, and the brightness index were used in the classification process. As a result, integration of DEM with the satellite images reduced the shadow effect by decreasing the brightness values of surfaces facing the sun and increasing those of the surfaces facing away from the sun. The increase in the accuracy is more noticeable in rugged areas, which are more affected by the topographical variation. Hence by using topographic correction method, the accuracy of the classification was increased from 64% to 82%.

Maselli et al. (2000) considered that different vegetation types are affecting the risk of fires. Supervised statistical classification was carried out using TM scenes and ancillary data layers. The effects of conventional processing of data were investigated on a different image acquisition periods of spring and summer season. First four principal components of the two scenes were used in the classification. The algorithm was first applied in a conventional method using the TM scenes and scenes plus ancillary data. Four classification results were gathered with 16 land cover classes. Improvement of accuracy is observed with the addition of ancillary data so TM plus ancillary data were used for the rest of the study. The results of the study showed that the addition of ancillary data yielded in Kappa improvements, and the result of classification can be used to estimate the risk of fire by assigning risk values to each vegetation class.

Lewis (1998), demonstrates the vegetation mapping methodology by using multispectral imagery. Two contrasting Landsat scenes of summer and winter were used in the study. The effects of atmospheric haze removed from the imagery. One hundred sample points were chosen to be representative of the range of vegetation types. Half of them were used as training samples and the remaining half were used as an validation sites to test the spectral classification. The biotic and physical ground cover components of the 100 points were estimated by using wheelpoint apparatus to detect the objective groupings of the sites. 50 training sites were labeled according to their corresponding groudcover class. Canonical variate analysis used to develop a spectral discrimination functions that would separate the classes. Then these training sites were used in maximum likelihood classification resulting in 9 map classes. Accuracy of vegetation mapping was tested in three different way; by testing the discriminant function developed on the training sites, by comparing image class with field class using 50 validation sites, and by comparing the image classification with an existing land system map.

Pinder et. al. (1997), studied the relationships between the relative distribution of vegetation types mapped from satellite images with elevation, slope and aspect parameters. Topographic Relative Moisture Index (TRMI) was computed by using aspect, slope steepness, slope configuration, and slope position values which are changing for xeric (being deficient in moisture) to mesic (having well-balanced supply of moisture) conditions. The TRMI values ranges between 0-60 and divided into quarters which represented 4 classes denoted as most xeric (TRMI<15), xeric (TRMI of 16 to 30), mesic (TRMI of 30 to 44) and most mesic (TRMI>45). The classification was done using Landsat TM images and five vegetation classes were identified. The statistical significance of the relationships between vegetation types and topographic variables were tested using χ^2 contingency tables. Statistical significances of χ^2 for the Linear test and Residual test determines a trend of either decreasing or decreasing abundance of vegetation related to the moisture gradient. The results showed that the topographic positions affect the distribution of all the vegetation types in all elevation intervals.

Maselli et. al. (1995) aims to find relative frequencies of the training sets on the ancillary layers can be transformed into modified priors by a strategy similar to that utilized for the spectral data. In this work an extension of the Maselli classifier is presented which allows the statistical integration of ancillary data layers into a classification process of remotely sensed scenes. Being nonparametric in nature, Mutual Information Analysis (MIA) was applied with this objective. MIA uses the concept of common entropy to estimate the statistical information shared by the different variables and can be applied to the processing of remotely sensed images for ecological investigations (Davis and Dozier, 1990). The first classification was based on the spectral data and its accuracy was measured by means of an error matrix compared to the test pixels. And then a new classification performed based on the same spectral data plus the ancillary information. In this situation, the inclusion of the ancillary information of the data layers yields a notable improvement in the classification accuracy. Among the layers considered separately, the incorporation of the soil map produces the maximum increase in accuracy.

Janssen et. al. (1990), aims to derive a more accurate land-cover classification using geographical data from a GIS and enable feed back of remote sensing derived information to a GIS. For this purpose two different test areas were selected, both representing agricultural regions. In the first test area, the agricultural fields were small and irregular, whereas the other test area was having more regular and larger agricultural areas. In an object based classification, the geometry of the objects were digitized and stored in the GIS. Then per-pixel classification was performed on the images. Incorrectly classified pixels are corrected by using the label of largest frequency in the object and assigning that label to all of the pixels within that boundary. In both of the test areas the accuracy of object based classification is higher than the per-pixel classification. Since the objects in the first area were small and with irregular shapes, the increase in the overall accuracy of this test site is less than the other site.

Hutchinson (1982), suggested the use of ancillary data in classification in three different stages. The incorporation of data could be either before, during or after the classification process.

Use of ancillary data before classification could be utilized to divide the area into smaller areas or strata, so each stratum could be processed independently, which is named as pre-classification scene stratification.

Ancillary data could also be included in a process during the classification stage. Most widely used approaches are increasing the number of channels of observation, or the modification of maximum likelihood decision rule by using prior probabilities, or generating some new feature components (Domaç et. al., 2004).

The use of ancillary data after multispectral classification is based on the discrimination of spectral confusion between different classes. In post classification, problematic spectral classes are treated as separate special cases. If there exist these special classes are treated via some other intrinsic decision rules that have been derived from other ancillary data.

2.3. Statistical Methods Used in the Classification Process

Cingolani et. al. (2004), proposed a method for defining discrete landcover units discernible by satellite. Firstly they performed unsupervised classification with a 10 class legend and select 251 patches to be used in the field for sampling. By examining these patches in the field 13 ecologically meaningful structural types were identified. Also the topographic position, slope, aspect, and altitude values for each stand were recorded.

In order to extract the signatures for statistical analysis and landcover mapping, another image of the area was used. Spectral signatures for 3x3 pixels surrounding each stand for each band are produced, which is named as 'brightness matrix'. This matrix is subjected to Principal Component Analysis to summarize the brightness variation in two axes. To find out the best combination of terrain attributes predicting the spectral characteristics of bands multiple regression analysis is performed on PCA axis and 14 structural variables. Newly generated groups are defined considering the knowledge of the area. Then new training areas are defined according to new generated groups. Maximum likelihood and Discriminant Analysis (DA) is performed on the images. The comparison through the field validation showed that the DA classification produced better results than the traditional Maximum Likelihood method.

Hietel et al. (2004), developed a method to characterize the major spatial-temporal processes of land-cover changes, identify a correlation between attributes and land-cover changes, and to derive concept of environmental factors of land-cover changes between 1945 and 1998. Land cover changes between these years were interpreted from aerial photographs and six different system included. Eight temporal images were used to define 'trajectories of change' that is the temporal sequences of eight successive land cover types of the patches. In order to group the trajectories cluster analysis were performed by coding the land cover data of the patch trajectories by their frequencies. Then the trajectories are grouped into land-cover types. To find the general land-cover pattern cluster analysis were used. Canonical Correspondence Analysis (CCA) was performed on the land cover data and environmental variables to find the relationships between them. Results of CCA show clear spatial and temporal correlations between environmental variables and land-cover data. From 1945 to 1998 CCA differentiated tree periods with a stable correlation between variables. Also CCA diagram reflects the typical stable sites of the land-cover classes.

Intraset correlations of environmental variables are used for which environmental variables were more important in structuring the ordination of land cover trajectories. A joint plot showed the general land-cover trajectory types over the whole observation time period and their correlations with environmental variables. Single joint plots at individual time intervals showed correlations between land-cover transitions and environmental variables.

Domaç et al. (2004), studied to improve the accuracy of vegetation classification by using future components which were constituted by using raw bands and various vegetation indices. The main inputs of the feature components are the indices. Two sets of indices were used; the first set includes the vegetation indices which directly give the spectral response of chlorophyll by using the ratio between red and NIR bands. The second set was used to remove the soil noise by changing slope value of red and NIR bands. The relationship between vegetation cover and the indices appears to change over the area according to the certain conditions such as soil cover type. To minimize the effect of soil on vegetation reflectance, second set of indices were used. Besides these feature extraction oriented indices, PCA were performed on raw bands in order to find if vegetation related information could be collected in few explanatory bands. At the end of these analyses it is assumed that; selecting PC1 and PC2 of vegetation indices, PC1 of soil indices, PC2 and PC3 of raw bands and DC3 and DC4 as feature components will remove the redundant data among multivariate datasets, such as multispectral remote sensing images and increase the accuracy of the classification. By using these feature components overall accuracy was increased to 76.92 % from 62.96 %. This rise shows that the new formed bands were successful in the discrimination of vegetation classes with very similar spectral reflectance values.

Dirnböck et. al. (2003), evaluated the use of correlation between plant communities and topography for vegetation mapping in Alpine regions. Two main techniques were employed in this study; 1) the gradient analysis by means of Canonical Correspondence Analysis (CCA) as a prediction tool, 2) image segmentation as a filter for reducing the number of incorrect predictions. The positions of each sample pixel in the CCA space were calculated, since its spectral and topographical variables are known. Extrapolation algorithm is constructed by using canonical coefficients and each pixel is assigned to a vegetation class which is in the nearest neighbor to this pixel in the ordination space. Region growing multiple pass segmentation algorithms was used in the image segmentation part to define regions in an image that correspond to objects, plant communities on the ground. The advantage of this method is to control the minimum and maximum number of polygons especially heterogeneous and patchy environments. Total area was divided into 162,000 regions by using orthophoto images. After overlaying the vegetation types derived from CCA results, 24 vegetation types were determined. In this study the accuracy of the result of CCA classification is calculates as 57 %. Combination of image segmentation and post-processing stages, overall accuracy increased by 12 %.

Pfeffer et. al. (2003), developed a procedure for moderate to high resolution vegetation mapping, where access is difficult and field data collection is expensive. In this study the topographic attributes were derived from a digital elevation model and nominal vegetation data were reduced to normalized scores by Detrended Correspondence Analysis (DCA). In the first step of the study, ecologically important derivatives of DEM were computed. DCA was used to reduce the vegetation scores to a limited number of major axis. Then DCA axis scores were related with the topographical attributes using multiple linear regressions. Spatial correlation structures of the regression residuals were examined using semi-variograms. In the final stage the semi-variograms and regression models were combined using kriging interpolation. DCA scores are classified into vegetation classes. Topographic attributes; altitude, slope, planform curvature, solar radiation, distance to ridges, mean wetness index and mean sediment transport were derived from the digital elevation model. The result of the field studies identified 147 species; most of them are common among the study area. Although all species has its own preferences, some of them are rather tolerant, which makes it difficult to identify correlation between these species and topographic attributes.

Takyu et. al. (2002), examined the differences in soil nutrient and moisture conditions along topographical gradient on different substrates, and compared the magnitude of changes among the substrates. Each substratum is divided into three topographical units (ridge, middle- and lower-slope) and the core samples were taken on the chosen plot areas. The mineral content of these samples were tested. The supply of soil is decreasing in the upslope direction with a decreasing order of Quaternary sedimentary, Tertiary sedimentary, ultrabasic rocks. To compare the composition of species among the nine plots, Detrended Correspondence Analysis (DCA) was performed to ordinate the plots based on the relative basal area of the species. Relationships between the number of species and sample area were compared among the nine plots by ANOVA.

Hoersch et. al. (2002) analyzed the overall influence of topography and landform on vegetation distribution using contingency tables and principal component analysis. In high mountain areas, site information is generally lacking, so DEM is an invaluable potential substitute for use in vegetation analyses. In the study appropriate landform parameters have been derived, indicating temperature and moisture distribution, exposure towards wind, snow and etc. Using contingency tables and principal component analysis the overall influence of topography and landform on vegetation distribution was analyzed. Throughout the study analysis of the correlation between vegetation types, distribution patterns and landform characteristics in an alpine environment is stressed. Besides primary parameters like elevation like slope amount and curvature indices, its first and second derivatives were created. As radiation and moisture supply can impose major restrictions for vegetation growth, some combined parameters were developed. Band wise topographic normalization was performed to minimize the problems arose because of the illumination effects and topographic shading. To find out the correlation between the landform parameters and vegetation types three main steps were performed within statistical analysis: a) Qualitative Analysis, b) Contingency tables and c) Principal Component Analysis. Landform variables showing high correlation with the first principal components are likely to be most important for the relevant vegetation class. For the whole vegetation dataset nearly all the landform parameters show high correlation with contingency coefficients. The PCA results show that in order to explain all of the information within the original landform dataset there has to be 8-9 principal components; this means there is not much redundancy in the data.

CHAPTER 3

DATA

Throughout the analyses of the thesis, six main data sets were used. The first one was the satellite images that were used to extract the vegetation classification map and the rest were topographic and geologic data set, which were used as an ancillary data in the post classification sorting step.

3.1. Satellite Data

The image used in this study was acquired from Landsat 7 Enhanced Thematic Mapper (ETM), which is the latest in a series of Landsat earth observation satellites. The acquisition date of the 174/35 image used in this study was 2000/8/7. Landsat ETM is composed of nine bands, six of them having the spatial resolution of 30 m and detecting the reflectance values both in the visible and infrared portions of the electromagnetic spectrum, one of them is panchromatic band and has 15 m resolution, and two of them are thermal bands with 60 m resolution.

Digital images collected from airborne or space borne sensors often contain systematic and unsystematic geometric errors. Some of these errors can be corrected systematically by giving calibrations data or parameter for correction. Other errors can only be corrected by matching image coordinates of physical features recorded by the image to the geographic coordinates of the same features collected from a map or from the field by using Global Positioning System.

In this study the scene was registered to the earth coordinates by using 1/25.000 scaled topographical base maps. The geometric correction was performed using a second-degree polynomial with a Root Mean Square Error (RMSE) of 0.5 pixels. Since the original DN values are not preserved during the resampling process, classification process was performed on uncorrected image. Geometrically corrected coordinates were given to the resulting classification map. ETM 5/4/3 color-composite of the study area is given in Figure 3.1.

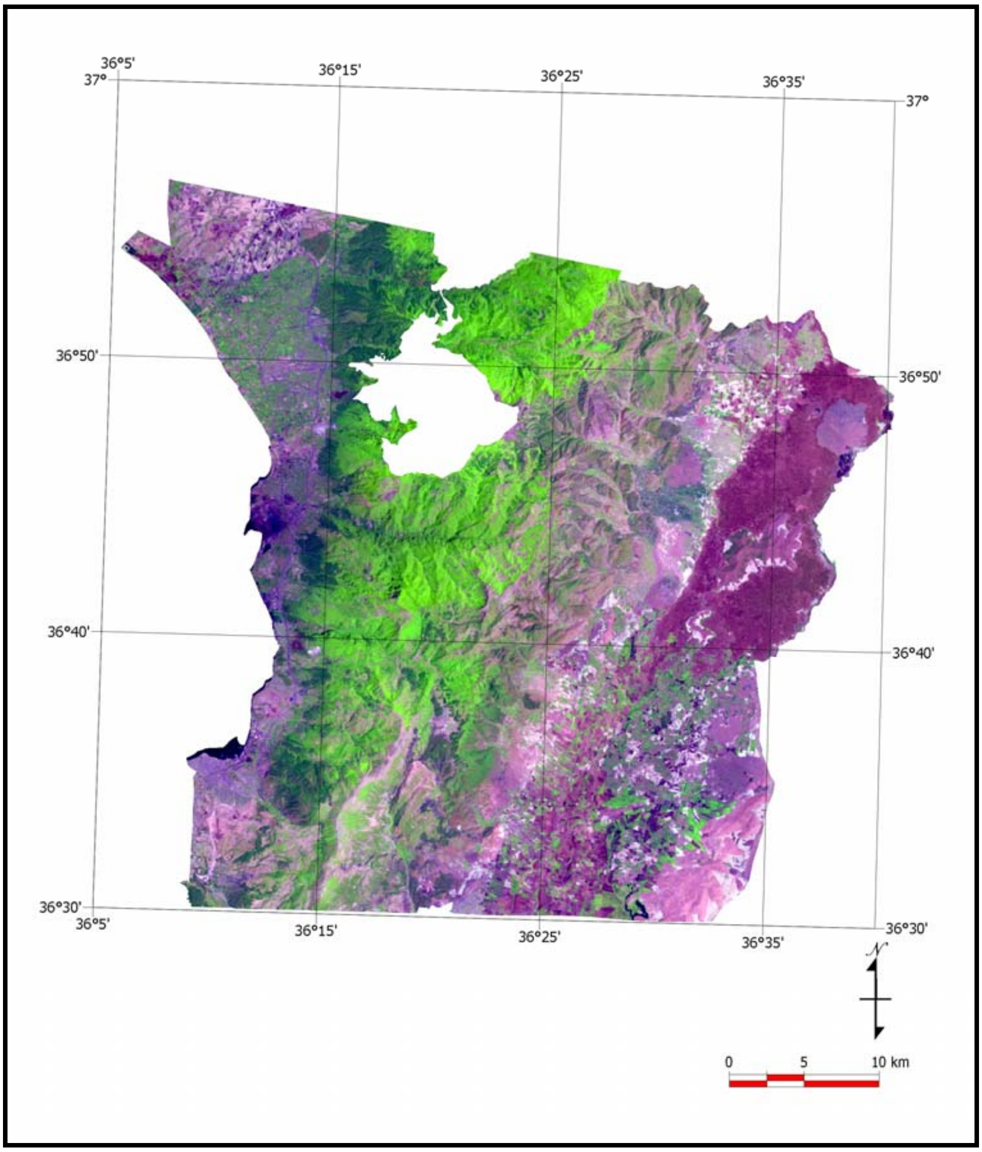

Figure 3.1. Landsat ETM 5/4/3 color-composite

3.2. Forest Data

The study area was covering 163,263 ha of areas and this corresponds to ten 1/25.000 scaled forest management maps which were gathered from the Ministry of Forest and Environment, Department of Mapping and Photogrammetry. Since middle upper portion forest management map of the study area could not be found, this part was not included in the thesis. These maps were then scanned and registered to the exact earth coordinates. Digitization process was completed in TNTmips software and the polygons were digitized according to the type, age, and canopy of forests. At the end of digitization process 26 different classes were determined, including settlement and agriculture areas (Figure 3.2). Following this, 26 different raw legend classes were merged into 9 modified uniform classes according to type, canopy and degradation of the vegetation which were given in Table 3.1 and Table 3.2 respectively (Figure 3.3).

This process is important for the ecological validity of the classification. Forest management plans are prepared for the timber production purposes and they do not always bear ecologically meaningful classes. Classes such as degraded maquis, coppice can be considered as an example to this situation. Based on this fact some of the classes have to be revised and converted or merged to acquire ecologically meaningful classes as discussed in the previous paragraph.

In the legend of the forest management maps some polygons are labeled as degraded forests. This class mainly determined according to the canopy closure; if the forest has open canopy with less than 40 % crown closure, it is called degraded. However this degradation could be result of two reasons, the areas could be either under the effect of severe human impact or the physical conditions of these areas were not suitable for the recovery of the formation prior to any sort of impact (i.e. natural or human induced). In order to carry out ecologically meaningful classification it was important to separate naturally open canopy forest types from the human induced open canopy. For finding the reason of degradation, these polygons were examined one by one and the new codes were assigned according to Table 3.2.

For ÇBÇz, (Highly Degraded Callabrian Pine) the polygons in the Eastern part of the study area are seems to be suitable for the growth of CP (Callabrian Pine) forest even though these areas classified as degraded in the forest management maps. These classes were added to CP in the second table considering the neighborhood of these polygons to CP forests, the suitability of elevation values and proximity of these polygons to the settlement points. Using the same logic with the CP species, the degraded OB, BP and O polygons were searched one by one to be included in BP, OB and O classes.

Table3.1. Legend of digitized raw forest management maps.

Raw legend Classes and ID numbers			
1	Forestation Area	**14**	Settlement and Agriculture Areas
2	Highly Degraded Black Pine	**15**	Oriental Beech
3	Highly Degraded Callabrian Pine	**16**	Oriental Beech Black Pine
4	Highly Degraded Other Broad-Leaved Trees	**17**	Oriental Beech Other Broad Leaved Trees
5	Highly Degraded Coppice	**18**	Oriental Beech Oak
6	Highly Degraded Oriental Beech	**19**	Oak
7	Highly Degraded Oak	**20**	Oak Black Pine
8	Highly Degraded Taurus Cedar	**21**	Oak Other Broad Leaved Trees
9	Black Pine	**22**	Oak Oriental Beech Other Broad Leaved Trees
10	Black Pine (Forestation Areas)	**23**	Normal Oak Coppice
11	Callabrian Pine	**24**	Forest Soil
12	Callabrian Pine(Forestation Areas)	**25**	Taurus Cedar
13	Callabrian Pine-Oak	**26**	Taurus Cedar Black Pine

Table 3.2. Modified uniform classes of forest management maps.

Class Code	Classes	Definition	Original class ID's
OB-O-C	Oriental Beech- Oak-Carpinus	Mixed Broad-Leaved Forest	17,18, 21, 22, 23
O	Oak	Broad-Leaved Forest	19
CP	Callabrian Pine	Needle-Leaved Forest	11, 12
FS	Forest Soil	Empty area within a forest	24
HD-OB-O-BP-C-Copp	Highly Degraded-Oriental Beech-Oak-Black Pine-Carpinus-Coppice	Mixed Needle and Broad-Leaved Forest	4, 5, 7
TC-O-OB-BP-CP	Taurus Cedar-Oak-Oriental Beech-Black Pine-Callabrian Pine	Mixed Needle and Broad-Leaved Forest	13, 16, 20
OB	Oriental Beech	Broad-Leaved Forest	6, 15
HD-TC-CP	Highly Degraded-Taurus Cedar- Callabrian Pine	Mixed Needle-Leaved Forest	3, 8
TC-BP	Taurus Cedar-Black Pine	Mixed Needle-Leaved Forest	2, 9, 10, ,25, 26
S&A	Settlement and Agriculture Areas	Settlement and Agriculture Areas	14

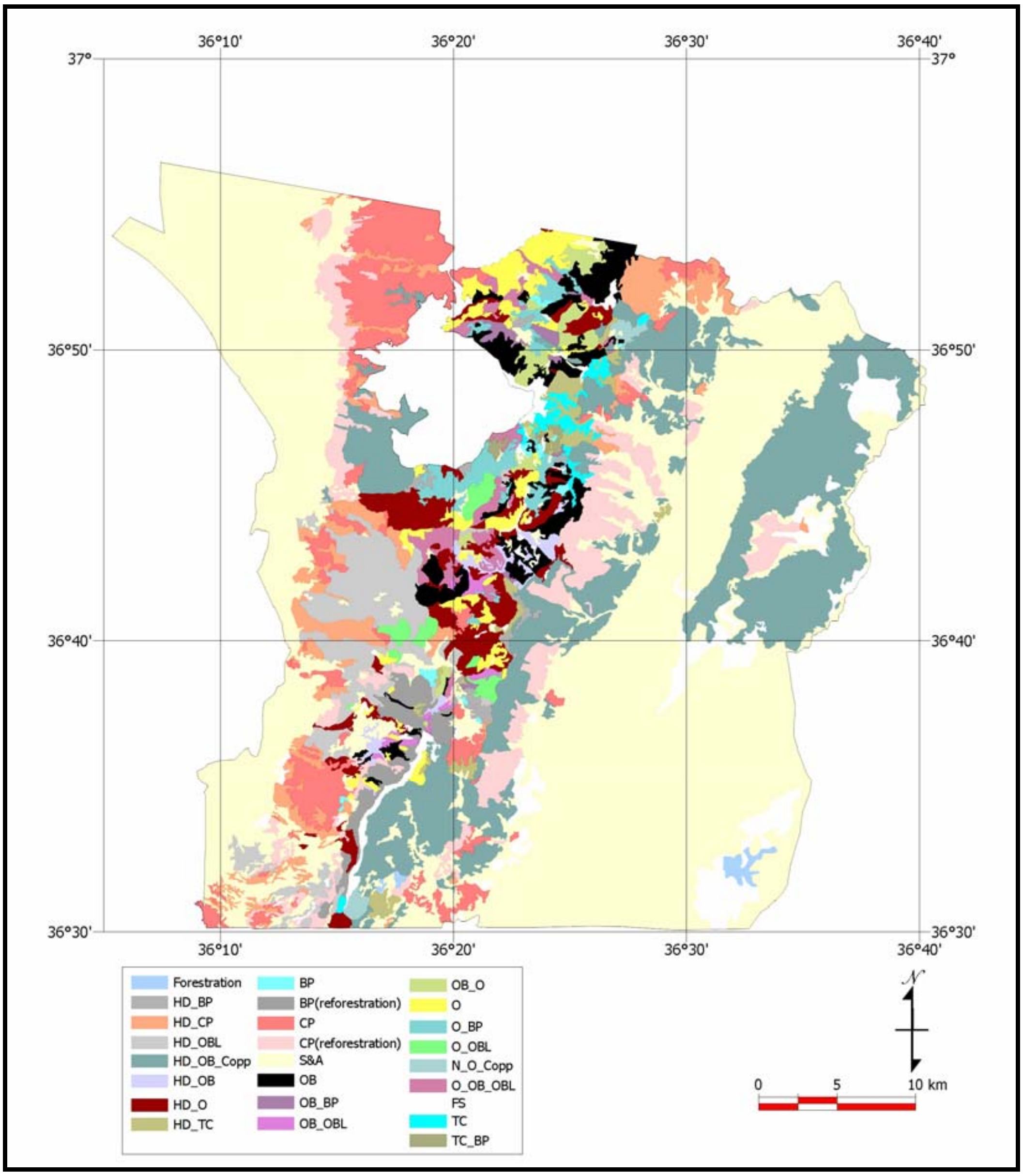

Figure 3.2. Digitized forest map of the study area

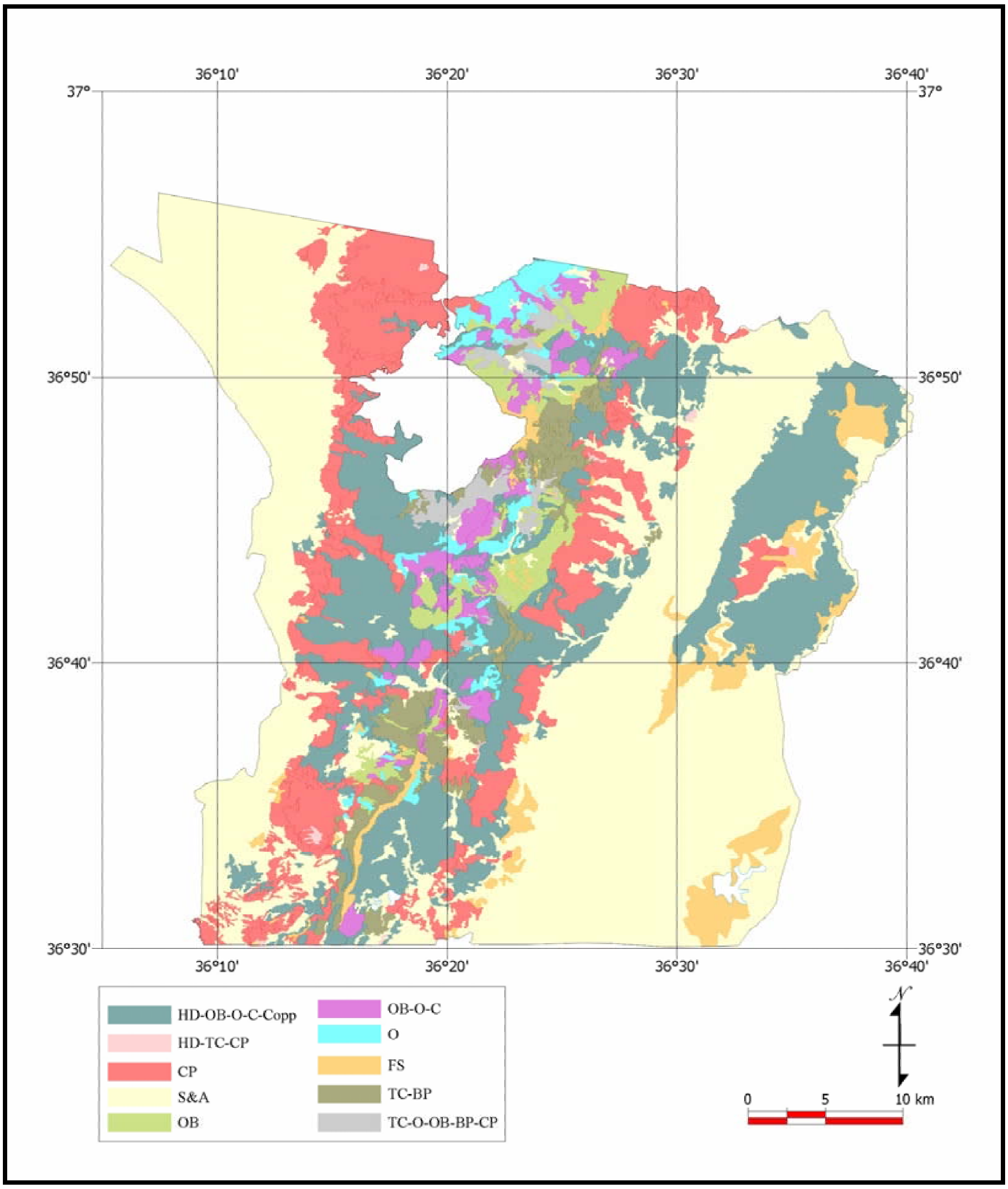

Figure 3.3. Reclassified forest Management Map of the study area

Due to its mosaic feature and the continuity it is usually difficult to identify discrete, repeatable classes to describe the vegetation cover (Whittaker, 1970 and Kimmins, 1997). Especially in the floristic level classification issue gets more complicated due to species that are physiognomically and physiologically similar to each other. Black Pine and Callabrian Pine, Oak and Hornbeam species, Oriental Beech and Hornbeam are the species of the study area that are highly similar to each other from remote sensing perspective. Figure 3.4 and Figure 3.5; give more information about the similarity of the above species. Besides above problems, forests management maps, only source of vegetation maps can be found extensively for Turkey, has many problems from the ecological classification perspective that does not help in the remote sensing classification exercises.

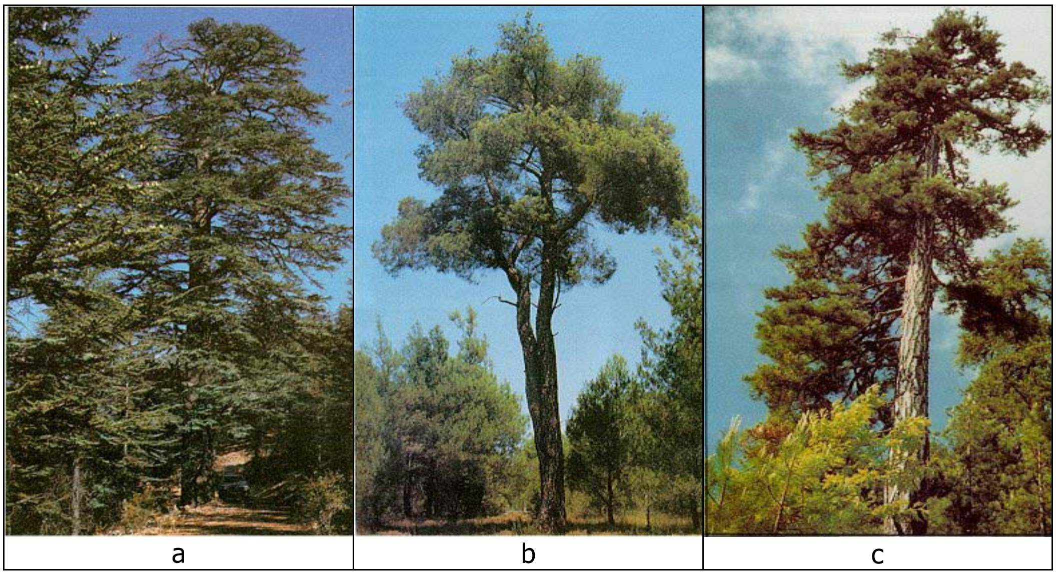

Figure 3.4. A figure showing needle-leaved species: a) Taurus Cedar, b) Callabrian Pine, c) Black Pine

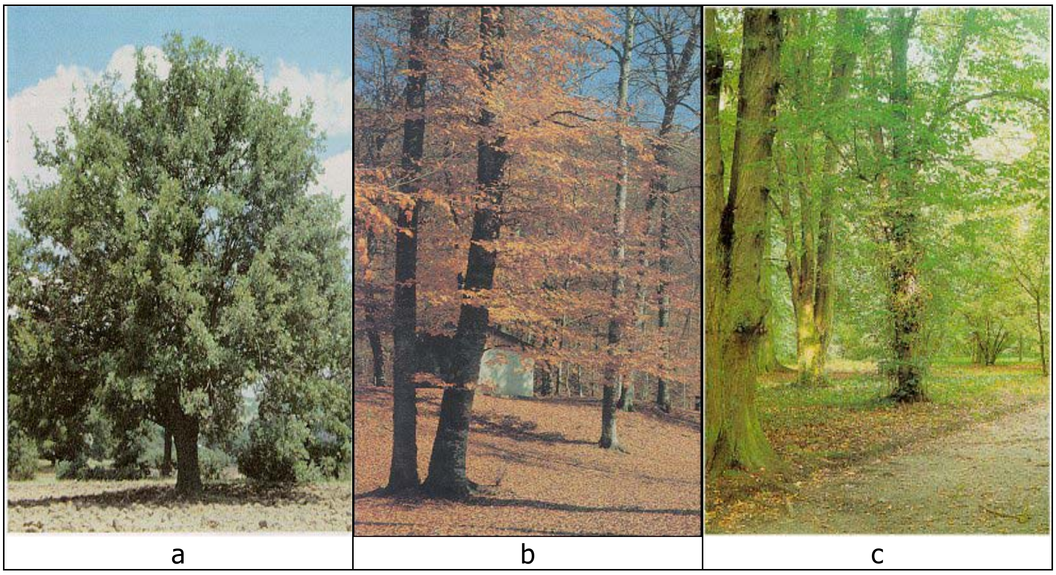

Figure 3.5. A figure showing broad-leaved species: a) Oak, b) Oriental Beech, c) Carpinus

3.3. Geologic Data

The 1/100,000 scaled geologic map of Amanos Mountains was acquired from the MTA General Directorate archive (Aksay et. al., 1988). This map then was scanned, extracted according to the boundary of the study area and registered to its exact coordinates using 1/100,000 scaled topographical maps. The extracted geologic map was digitized and the boundaries of the units obtained. Actually the original map consists of 17 distinct geological units given in Figure 3.6 (Table 3.3). This map was reclassified into a material map as some of the geological classes are made up of same rock material but different in age. All of the units were checked both by definitions from MTA reports and field based expert knowledge had been gathered from geologists who had been worked in the study area (oral communication, Dr. Bora Rojay, 2004), based on physical characteristics. Finally geological units have been merged, the final legend is decreased down to 11 classes (Table 3.4) and geological map is reclassified as a material map that is presented in Figure 3.7.

Table 3.3. Legend of geology map taken from MTA.

Code	Age and map code	Short Description
1	Lower Jurassic-Triassic Tra	Quartz Arenite, quartzite conglomerate
2	Lower Jurassic-Triassic Trk	Dolomite, dolomitic limestone
3	Eosene Ta	Cherty nodular limestone
4	Cambrian Ee	Conglomeratic quartzite, quartz arenite
5	Cambrian Ec	Shale, siltstone, sandstone, greywacke
6	Cambrian Ek	Dolomite, dolomitic limestone
7	Cambrian Et	Shale, siltstone, greywacke , argillaceous limestone
8	Cretaceous-Upper Jurassic Kjk	Limestone, dolomitic limestone
9	Quaternary Qal,Qm	Alluvium, debris
10	Miocene Te	Macro fossiliferous limestone
11	Miocene Ty	Marn, sandstone, claystone, limestone
12	Ordovisian Ok	Quartzite, sitstone, shale
13	Paleocene-Maastrichtian Tka	Limestone with basal conglomerate
14	Pliocene Ts	Conglomerate, sandstone, claystone
15	Quaternary Qb	Basalt
16	Cenonian Kol	Serpantinite, limestone, ophiolite
17	Upper Cretaceous Kof	Tectonic peridotite, gabbro, diabase

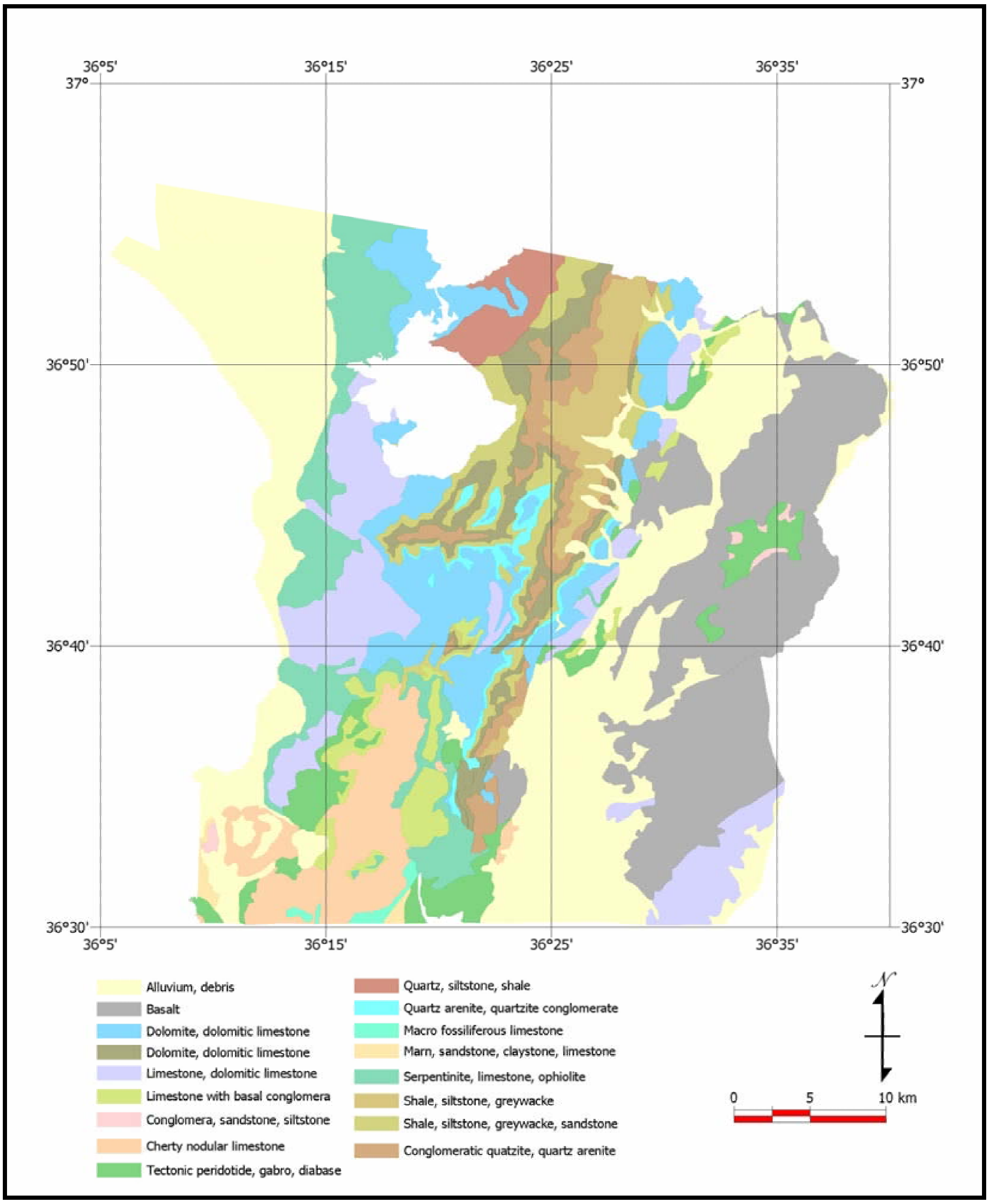

Figure 3.6. Figure showing the digitized geology map of the study area

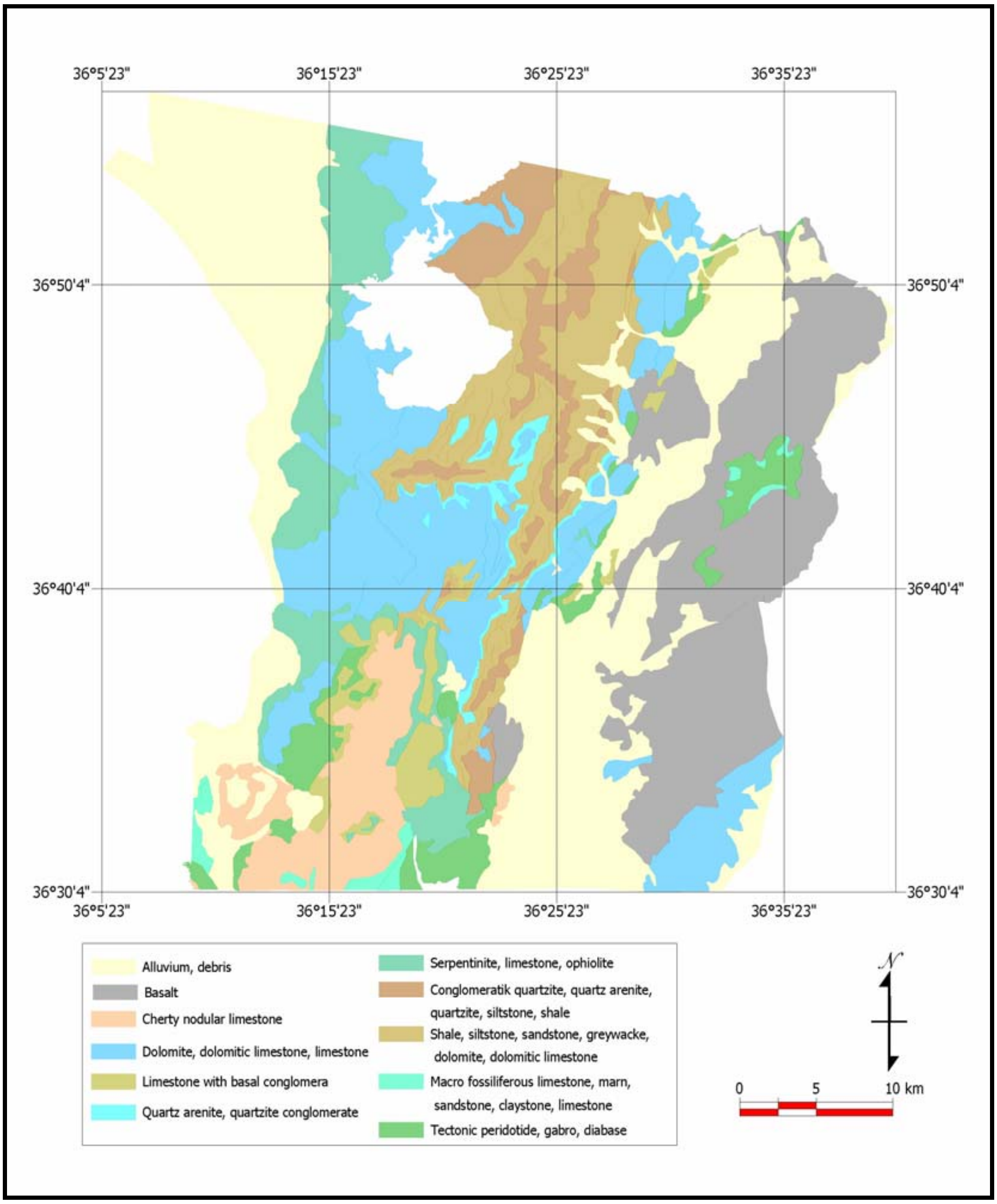

Figure 3.7. Reclassified geologic map of the study area.

Table 3.4. Legend of reclassified geology map

Original units code	Description
2,8	Dolomite, dolomitic limestone, limestone
9	Alluvium, debris
16	Serpantinite, limestone, ophiolite
4,12	Conglomeratic quartzite, quartz arenite, quartzite , siltstone, shale
17	Tectonic peridotite, gabbro, diabase
5,6,7	Shale, siltstone ,sandstone, greywacke, Dolomite, dolomitic limestone
15	Basalt
13	Limestone with basal conglomerate
3	Cherty nodular limestone
10,11,14	Macro fossiliferous limestone, Marn, sandstone, claystone, limestone
1	Quartz Arenite, quartzite conglomerate

3.4. Digital Elevation Model (DEM)

The Digital Elevation Model (DEM) used in this study was taken from the Shuttle Radar Topography Mission (SRTM) data. SRTM obtains Earth surface data by remote sensing technology utilizing synthetic aperture radar. Obtained data is converted into height data called a Digital Elevation Model (DEM), having a resample resolution of 90 meters (Figure 3.8) (http://iss.sfo.jaxa.jp/shuttle/flight/sts99/mis_srtm_e.html). Absolute vertical accuracy of SRTM data is ±16 m and relative vertical accuracy is ±6 m, horizontal positional accuracy is about ±20 m (Kääb, 2004).

Elevation of the study area ranges from 0 to 2214 m. The lowest elevations are dominant around the eastern part of the area generally covered by the agricultural areas and the western most areas where the Mediterranean Sea exists; whereas the highest elevations can be observed around the central and central north regions of the study area which are the Amanos Mountains.

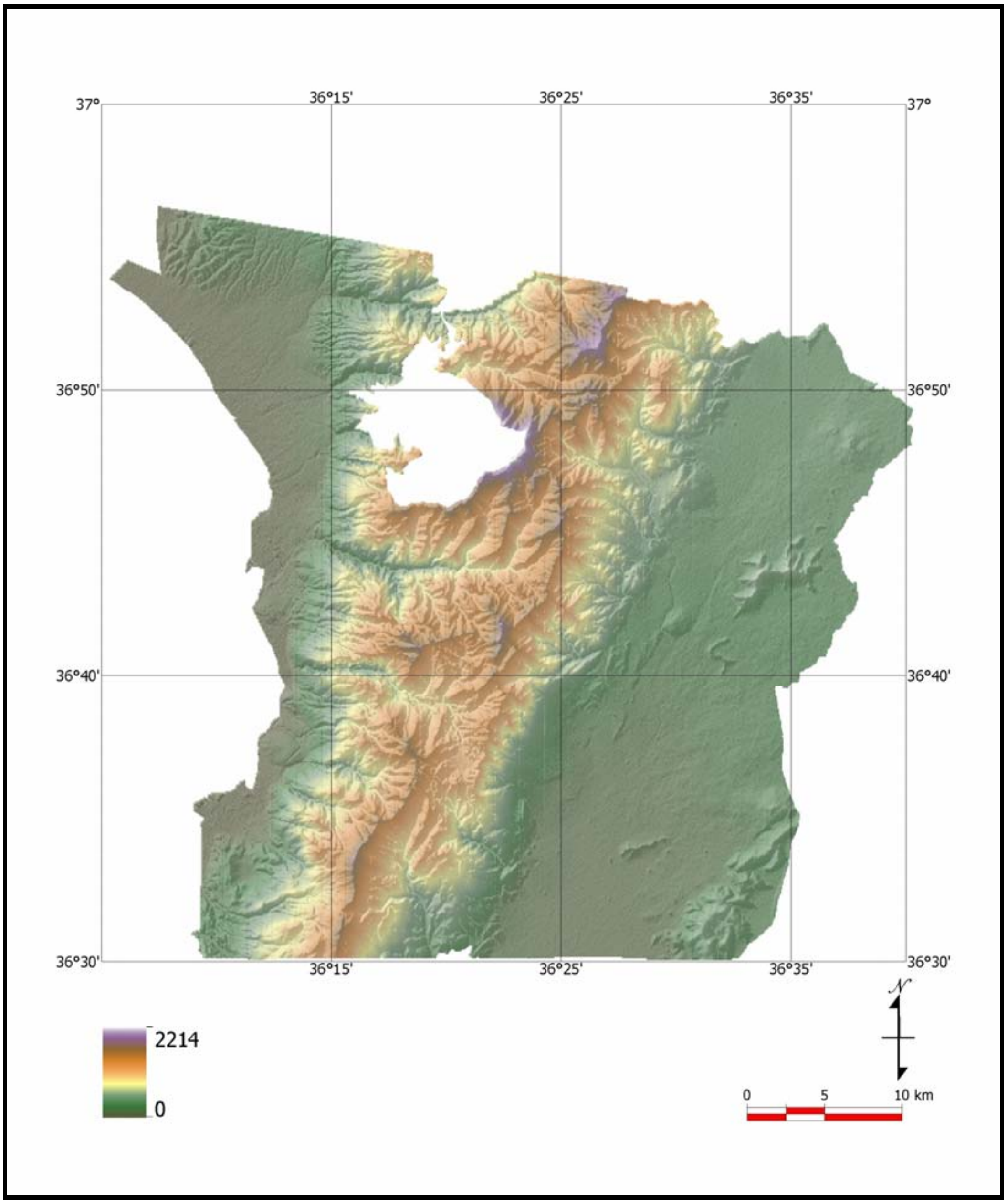

Figure 3.8. Digital Elevation Model of the study area

3.5. Slope Data

Slope data was constituted using DEM of the area. Elevation values in adjacent cells are compared for the slope output and slope value ranges between 0 to 75 degrees. Slope map of the region is given in Figure 3.9.

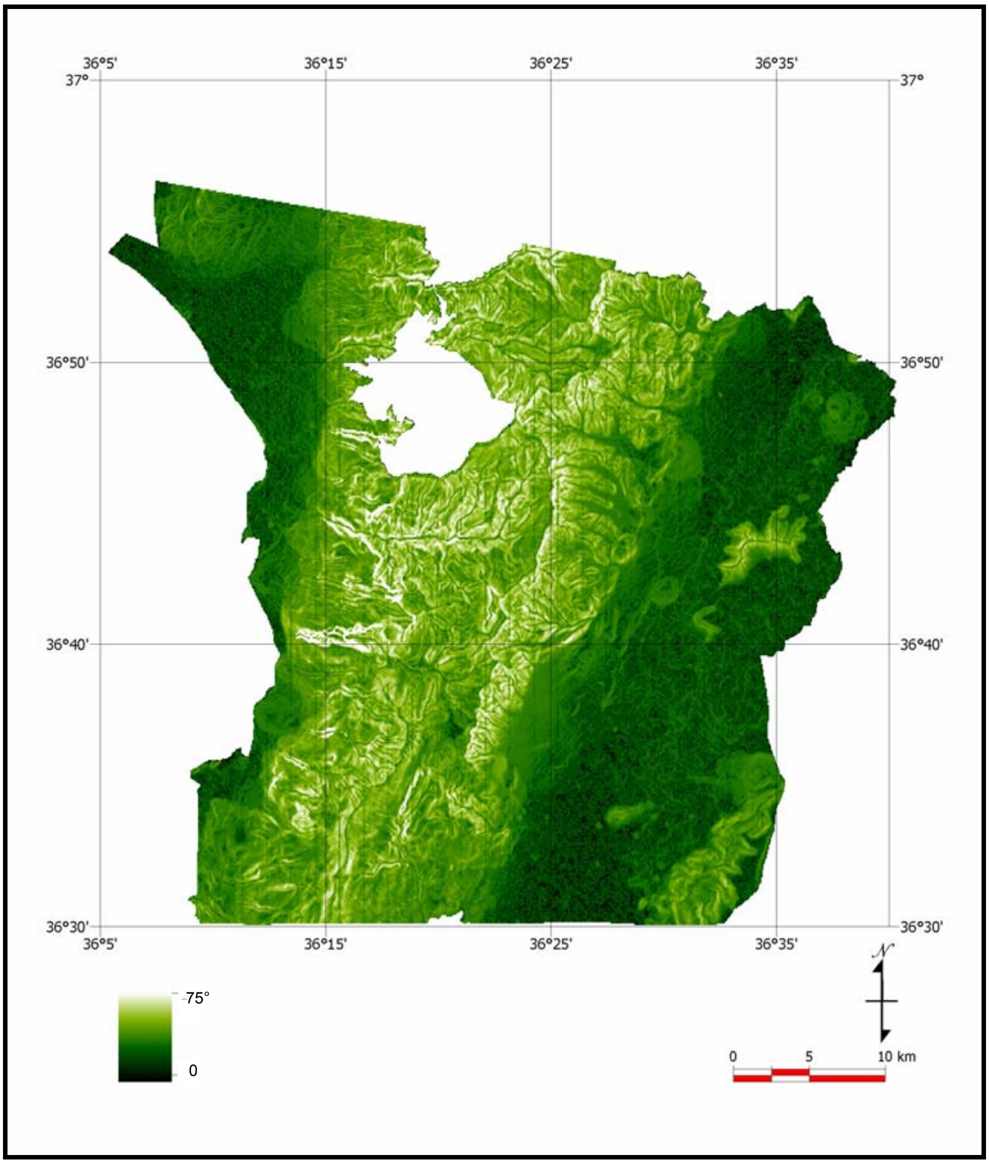

Figure 3.9. Slope map of the study area

3.6. Aspect Data

Aspect values (or compass direction of a slope) of the region vary between 1° to 360° (Figure 3.10). This circular distribution is misleading in the statistical analysis stage. For example even though 2°and 358° are very close values in real world, but in interval scale analysis stage these two values would gain totally different ranks. In order to overcome this problem aspect should be transformed by trigonometric functions (Roberts, 1986) into its components. The simplest way to do this is to create two new variables, called "northness" and "eastness". These components are calculated as:

$$\text{Northness} = \cos(\text{aspect})$$

$$\text{Eastness} = \sin(\text{aspect})$$

Northness will take values close to 1 if the aspect is generally northward, close to -1 if the aspect is southward, and close to 0 if the aspect is either east or west. On the other hand eastness behaves similarly, except that values close to 1 represent east-facing slopes. Eastness and northness maps of the study area was given in Figure 3.11.

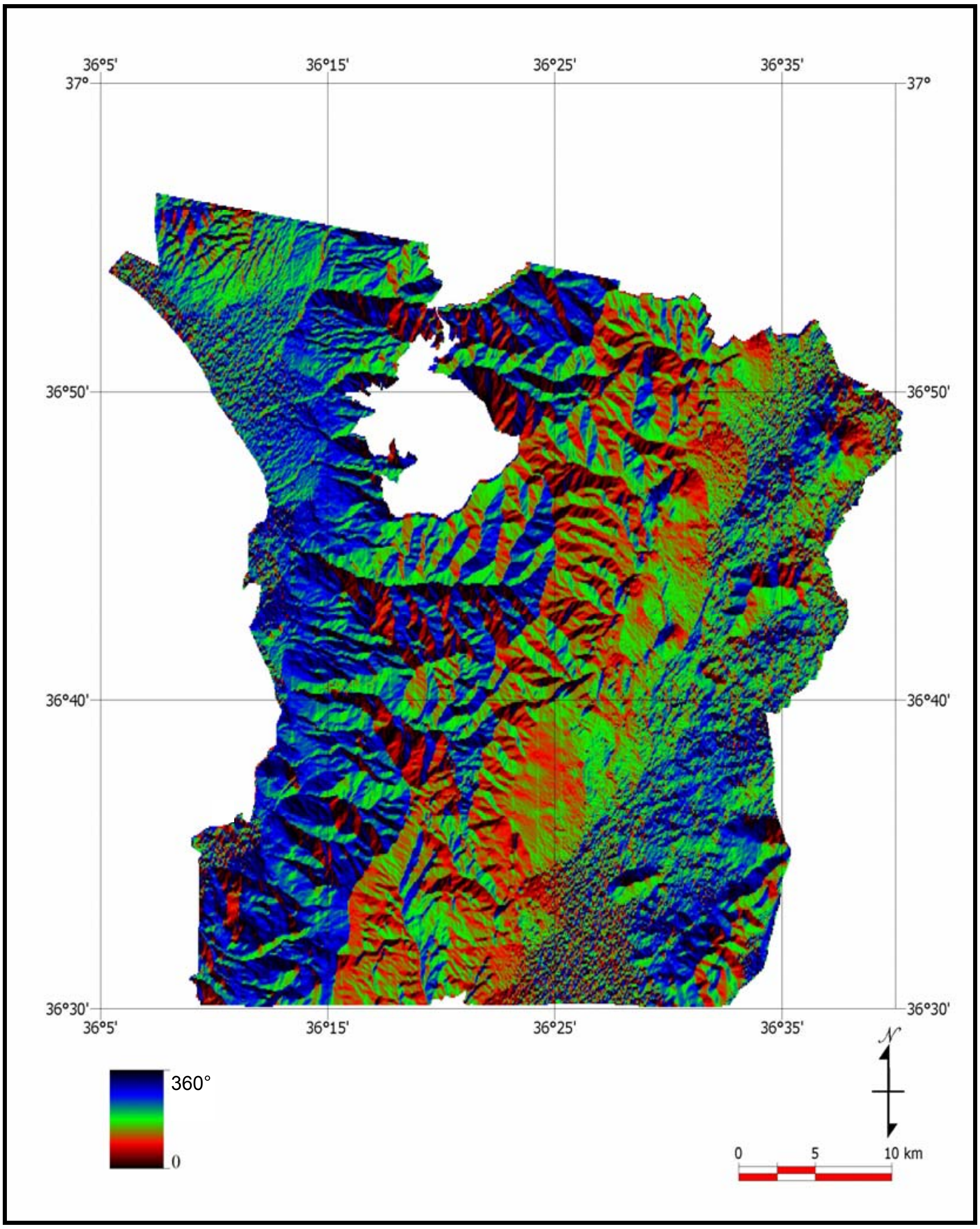

Figure 3.10. Aspect map of the study area.

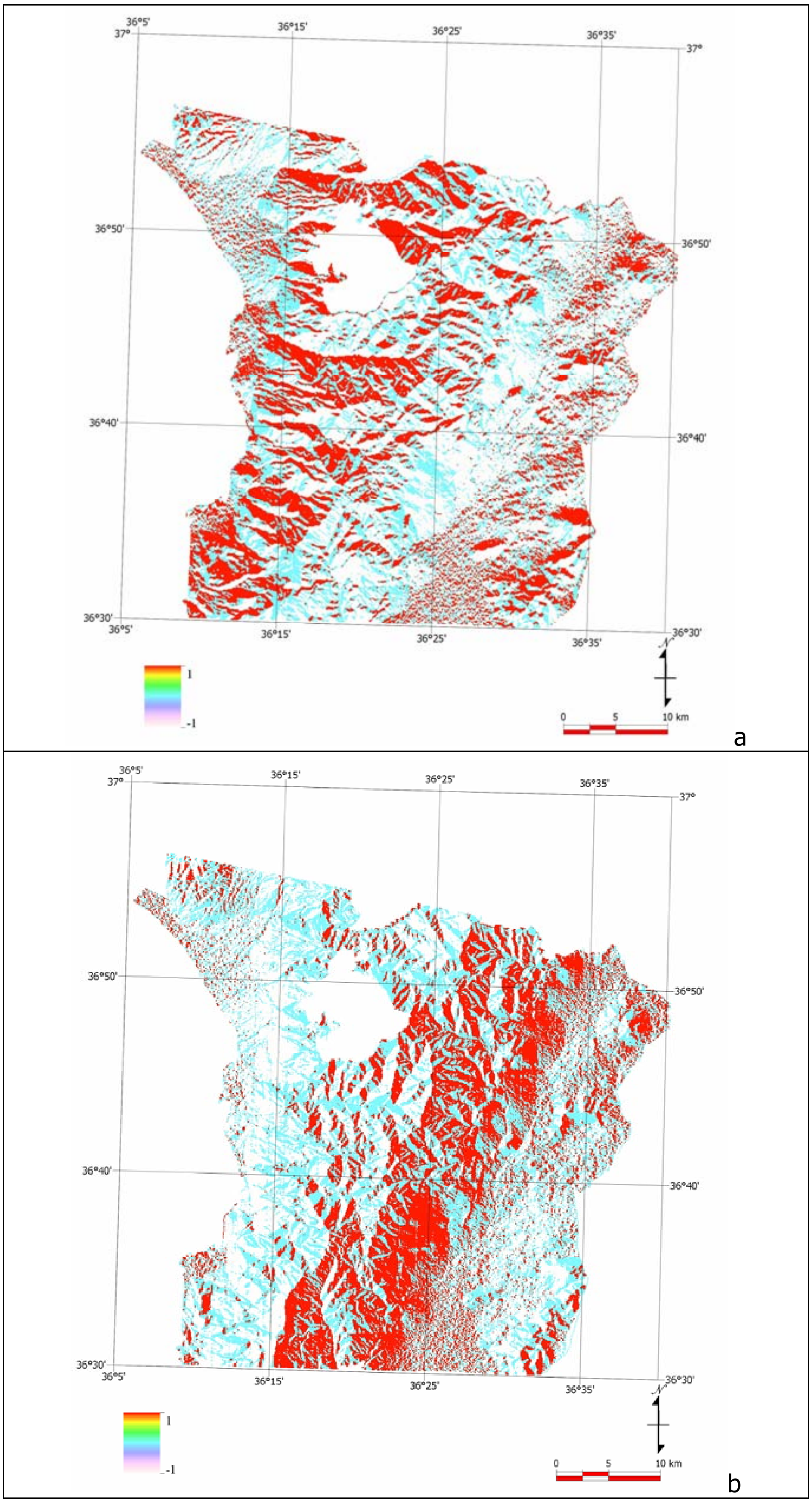

Figure 3.11. a) Northness component of aspect map, b) Eastness component of aspect map

CHAPTER 4

ANALYSIS

4.1. Methodological Snapshot

The flowchart of the study explains the structure of the thesis with a graphical representation (Figure 4.1.).

In the data preparation part of the study, two main sets of data are used. The first set is the Landsat ETM images and SRTM DEM, slope, geology, forest, northness, and eastness are the second set of data used as ancillary information in the study.

Coarse classification of the Landsat images is performed using maximum likelihood classification algorithm in the second part. The class raster which is the result of this process is the taken as the coarse classification result of the study and named as *stage 1*. A bi-product was also produced called distance raster showing the probability of each class.

For the statistical analysis, tree sets of data are prepared, two of which are sampled by random stratified method and the other is prepared with an expert knowledge. These data sets are evaluated by canonical correspondence analysis to find out the general relationships between environmental variables and species distribution.

In the second part discriminant analysis is used to determine which environmental variables differentiate the vegetation classes. Determined Fisher's coefficients are used to classify the vegetation cover and named as *stage 2* classification.

Integration of maximum likelihood classification and the results of the statistical analysis are applied in part three by using the distance raster values. The pixels with a distance raster value of 1 are considered as correctly classified by maximum likelihood classification and taken from *stage 1* result. Pixels having a distance raster value less than 1 are accepted as not correctly classified with a maximum likelihood classification and result of *stage 2* is used instead of those pixels and *stage 3* is formed. Accuracy assessment of all stages results are calculated for each data set, and discussed in this chapter.

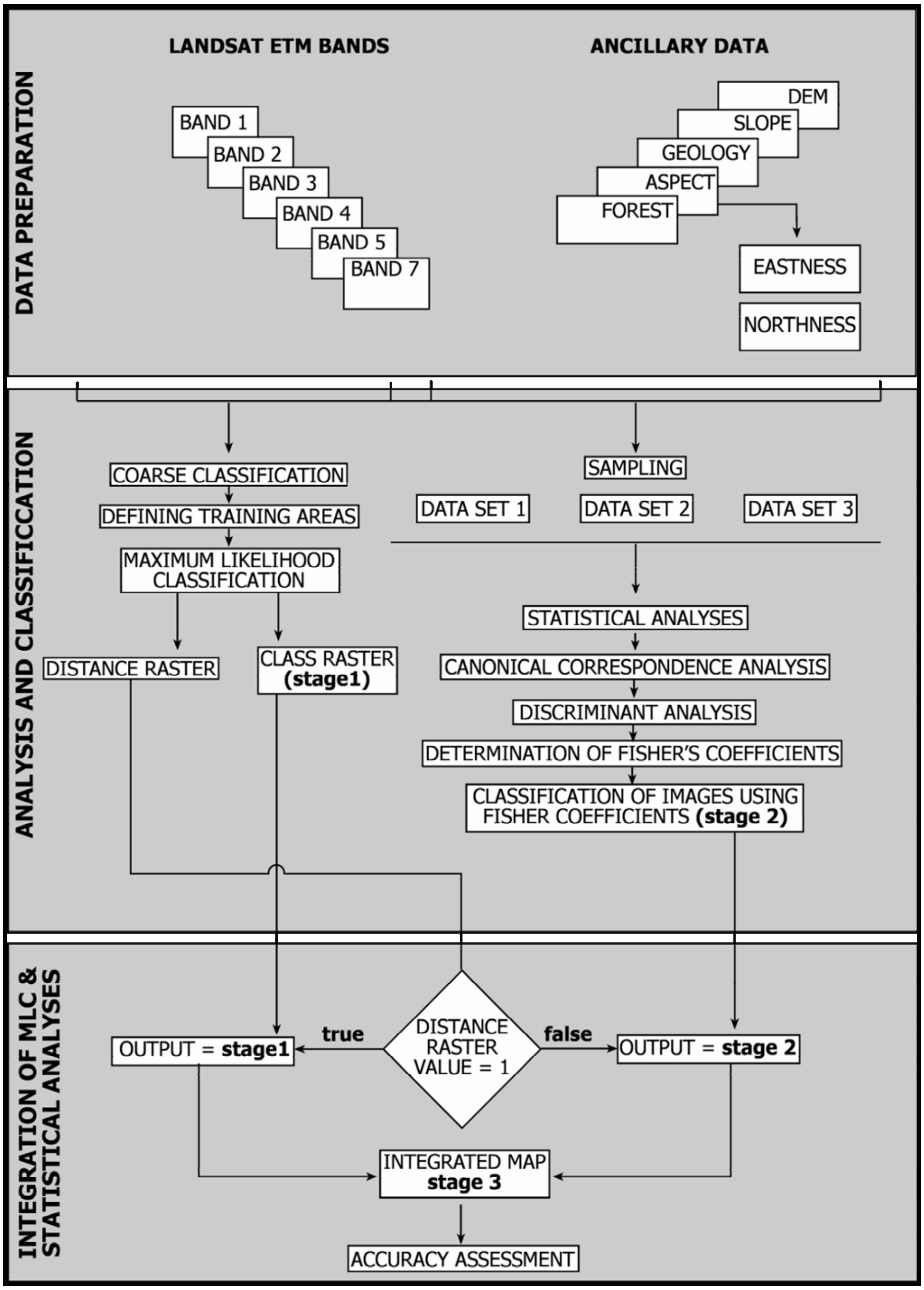

Figure 4.1. Flowchart of the study.

4.2. Coarse Classification

In this study the vegetation classification is performed using the subscene of Landsat 7 ETM, extracted from the 174/35 scene. TM1, TM2, TM3, TM4, TM5, and TM7 bands of the Landsat image with a spatial resolution of 30 meters were used in the classification of the image.

4.2.1. Training Set

Supervised classification method performs the classification using the spectral values of the training areas. Known points are signed to train and the method uses these known pixels to classify the unknown areas into one of the previously defined classes. The aim of defining a training set is to gather the statistical information of the spectral responses of the vegetation classes. Despite Jensen, (2000) recommended that the minimum number of the training pixels should be ten times the number of the used bands, to get the statistical information of the spectral bands it would be useful to take the training pixels ten or hundred times the number of bands.

Training set of the coarse classification was constituted using the data coming from the field and 1/25,000 forest management maps. Since the quality of training data affects the accuracy of classification, it is important to choose pixels from homogeneous, representative areas and away from the areas of mixed classes. Visual interpretation of TM 543 color composite with the use of forest management maps helped to find the best fitting training set. Table 4.1 gives the number of training pixels taken from each of the classes in this study.

The relationships of the classes could be understood from the spectral characteristics of the training pixels. In the training set dendogram in Figure 4.2, seperability of each class is calculated. The pair classes that join together near the left edge of the diagram are closely related with each other and the degree of relatedness decreases to the right. According to this figure; spectral reflectance values of CP and TC-BP pair and TC-O-OB-BP-CP and OB-O-C pair are very similar to each other. On the other hand; FS and OB classes could be easily discriminated from the rest of the classes.

Table 4.1. Number of training pixels.

Class Value	Number of Training Pixels
Oriental Beech-Oak-Carpinus (OB-O-C)	37
Oak (O)	24
Callabrian Pine (CP)	51
Forest Soil (FS)	64
Highly Degraded Oriental Beech-Oak-Carpinus Coppice (HD-OB-O-C-Copp)	45
Taurus Cedar-Oak-Oriental Beech-Black Pine-Callabrian Pine (TC-O-OB-BP-CP)	43
Oriental Beech (OB)	48
Highly Degraded Taurus Cedar-Callabrian Pine (HD-TC-CP)	43
Taurus Cedar-Black Pine (TC-BP)	37
Total Number of Pixels	**392**

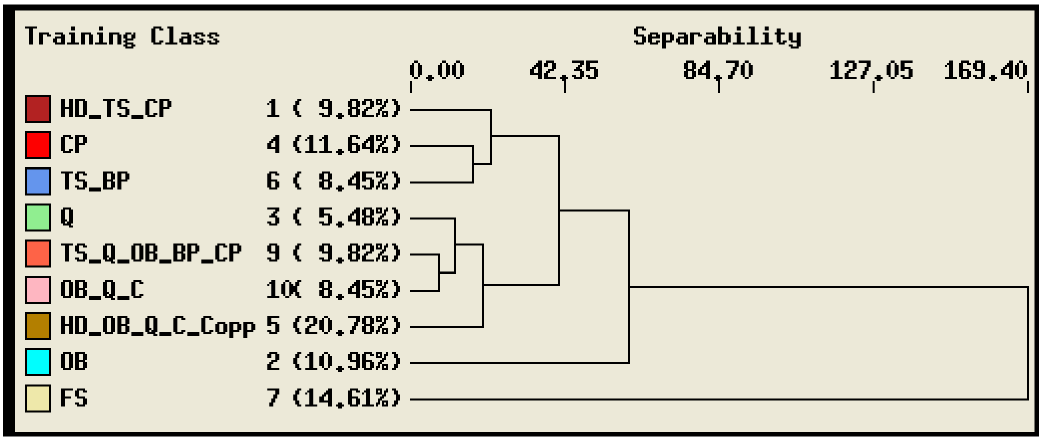

Figure 4.2. Training set dendogram

4.2.2. Maximum Likelihood Classification

After the selection of training pixels, classification was performed on the raw bands of the image. Nine vegetation classes were discriminated using maximum likelihood classification classifier. The resulting map of this process is given in Figure 4.3.

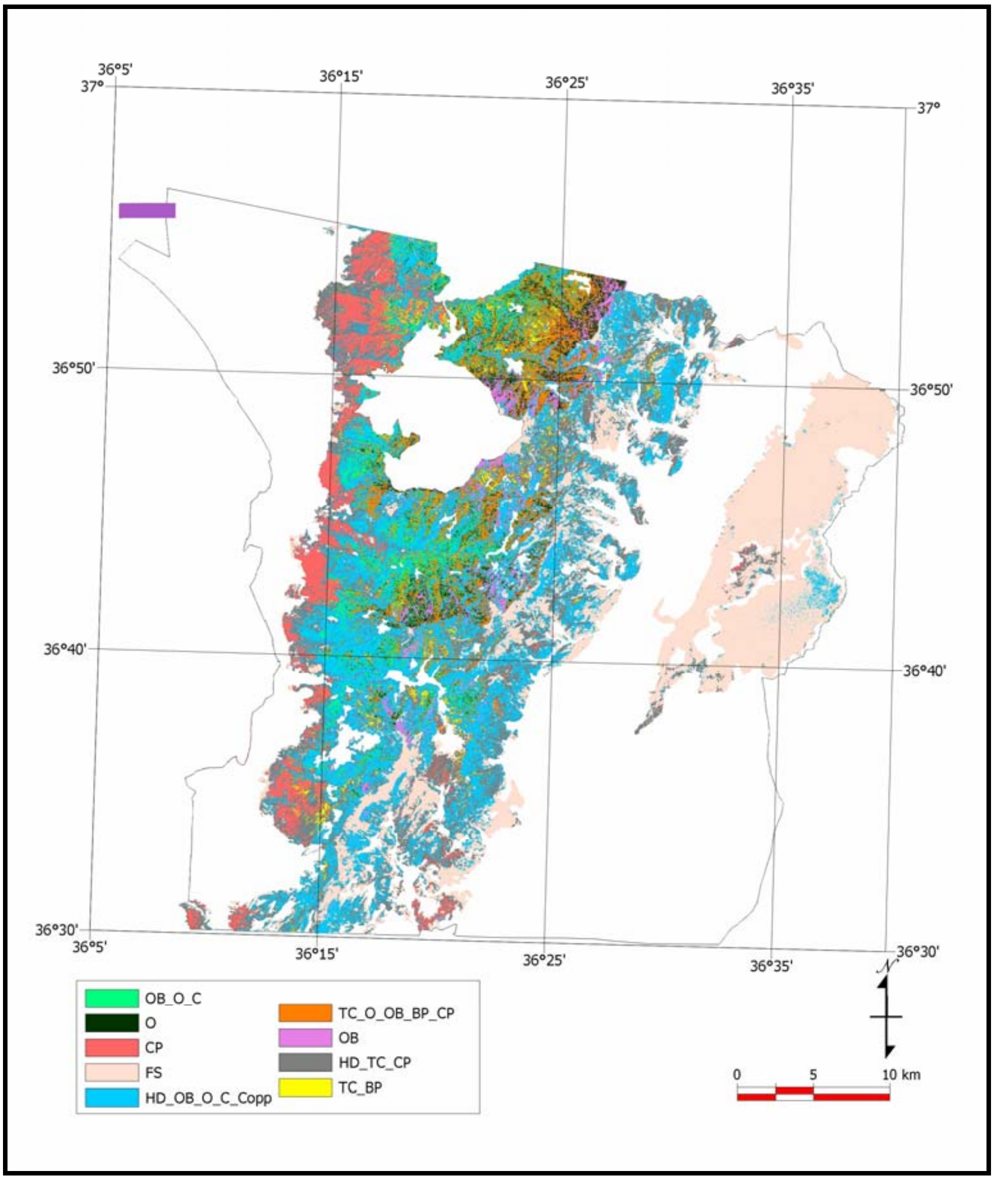

Figure 4.3. Coarse Classification result

4.3. Statistical Analyses

4.3.1. Sampling

In order to perform the Discriminant analysis, tree different sets of data were prepared to be used as statistical training sets in the analysis. In Set 1, the sample sizes were equal for each of the classes and stratified random sampling was used while generating the samples. In stratified random sampling, prior knowledge about the study area is used to divide the area into groups, and then each group is randomly sampled (Congalton and Green, 1999).

In Set 2, the numbers of points are proportional to the total area of those class polygons. The points were selected by using again by stratified random sampling. A minimum amount of 15 points were used as local minimum normalizer relative to the area, hence regardless of size any class is represented by at least 15 random points.

In Set 3, the points were taken in equal size from each of the classes, but in this case instead of stratified random sampling, the points were selected considering the representativeness of the point to which class it belongs to. The number of points for each set is given in Table 4.2.

After forming the data sets by sampling, a table covering the elevation, slope, aspect, forest management map classes, geology classes, TM1, TM2, TM3, TM4, TM5, and TM7 values of the points were prepared and saved for each of the set separately. Figure 4.4 is showing the point values of one of the sample point.

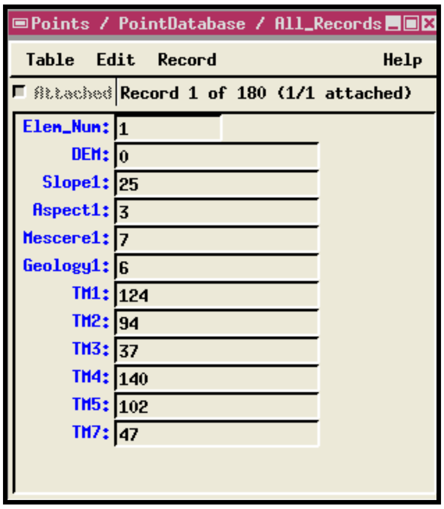

Figure 4.4. Table showing the point values of the Data Set3

Table 4.2. Number of points taken from the classes for each Data Set

	OB-O-C	O	CP	FS	HD-OB-O-C-Copp	TC-O-OB-BP-CP	OB	HD-TC-CPO	TS-BP	Total sample size
Data Set 1	40	40	40	40	40	40	40	40	40	360
Data Set 2	21	15	104	30	120	16	21	15	22	364
Data Set 3	20	20	20	20	20	20	20	20	20	180

4.3.2. Canonical Correspondence Analysis

In harmony with the research problem in order to improve the mapping action of the vegetation species classes, basic topographical information is needed to be combined with the vegetation data. Just before this combination the relationship between the vegetation cover and environmental variables were explored by using Canonical Correspondence

Analysis (CCA) (Ter Braak, 1986), to figure out if there exist any significant relationship. CCA is a multivariate direct gradient analysis which is widely used in ecological studies. In this analysis, ordination axes are generated using the linear combinations of environmental variables, and calculates the centroids of the species or communities within the ordination space. The resulting ordination diagram shows the patterns of variation and also the main relationships between the species and each environmental variable (Dirnböck et. al., 2003;).

In the diagram obtained from CCA (bi-plot) contains the environmental variables plotted as arrows, along with points for the samples or species. Each sample point lies at the centroid of the points for species that occur in those samples. Arrows representing the environmental variables indicate the direction of maximum change of that variable within the diagram. The length of the arrow is proportional to the rate of change and the position of the species points in relation to the arrows indicates the environmental preference of that species (MVSP User's Manual, 1986-1998).

The sum of canonical eigenvalues indicates the amount of variation explained by the site variables supplied (Ter Braak, 1986). According to the table 4.3, 62.63% of the variation in the species distribution is explained by using 5 environmental variables, in the first two Canonical Correspondence Axes (CCA) where 76.19 % is explained in the first three axes.

Table 4.3. Eigenvalues of CCA axis

	Axis 1	**Axis 2**	**Axis 3**
Eigenvalues	0,54	0,37	0,18
Percentage	6,75	3,84	2,29
Cum.Constr.Percentage	39,94	62,63	76,19
Spec.-env. correlations	0,75	0,55	0,43

Although the CCA literally explains the relation of vegetation species classes with environmental variables, the quantification of this relation in terms of crisp decision rules are hardly possible. The graph of multivariate analysis of environmental variables with the species distribution didn't give clear relations between species and variables (Figure 4.5). This is probably due to the fact that the interaction within the variables are quite complex and could not be modeled by original feature components in three axes dimensions. Further factorials of features and interactions between variables might be attributed to this non-selective nature of this modeling space.

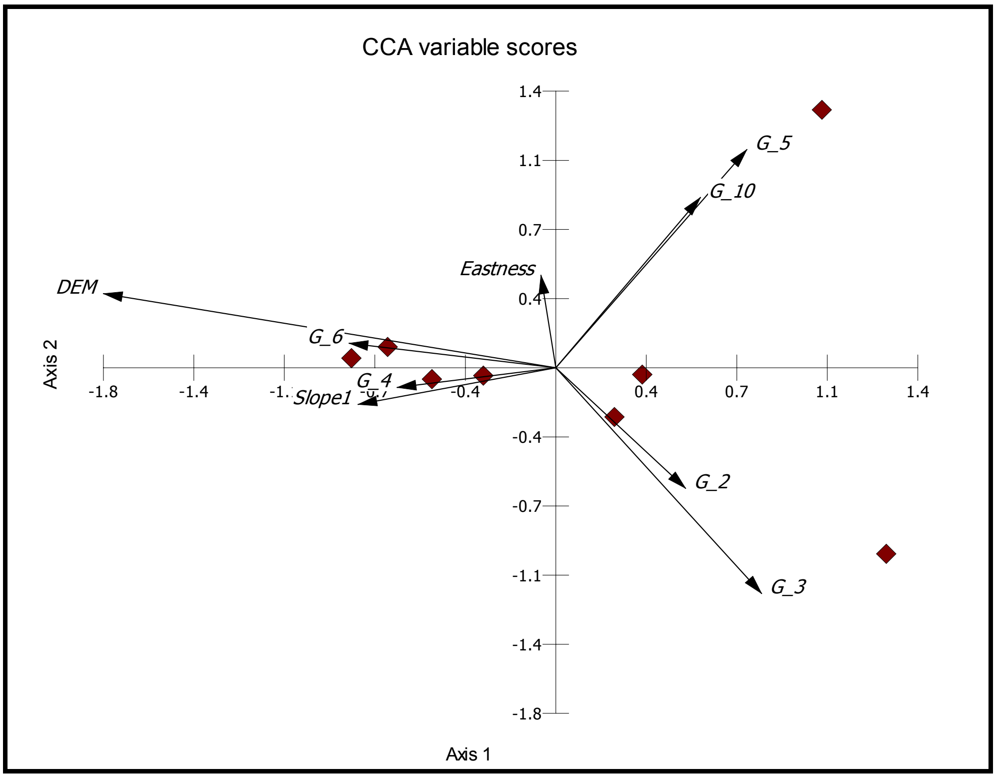

Figure 4.5. Biplot of species vs. environmental variables as a result of multivariate analysis

So each environmental variable univariately analyzed with the species distribution, in order to see if there exist any significant relation that could be utilized as an embryonic decision rule. The results of these analysis, express the occurrence of the species according to DEM, Slope and Aspect individually. After analyzing these graphs with an expert, some decision rules were extracted (Oral communication, Uğur Zeydanlı, 2004). The results of CCA can be explained by:

- *Forest Soil:*

Areas with lower elevation are more easily accessible thus suspect to human impact. Because of that these areas usually lost their forest cover and covered with the herbecous vegetation and usually occur below 500 m elevation (Figure 4.6).

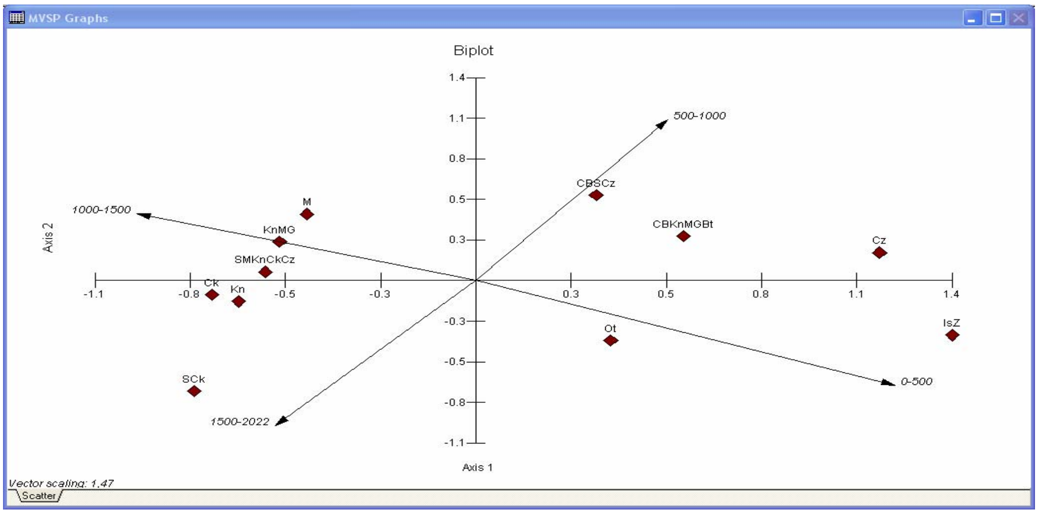

Figure 4.6. Graph of DEM vs. Species distribution

- *Callabrian Pine:*

Its distribution rarely exceeds 1000 m. According to Figure 4.6, CCA shows that elevation of 0-500 and 500-1000 m gives information about the distribution of the Callabrian Pine but it is not affected by the slope.

On the other hand Callabrian Pine prefers the areas under the effect of sea because of humidity. According to Figure 4.7 distribution of the Callabrian Pine is explained by the W-NW and W-SW aspects because of the sea effect.

- *Oak:*

For the distribution of Oak DEM gives good information (Figure 4.6). According to this figure, 1000-1500 m explains the Oak's distribution sufficiently.

Although Oak's distribution does not show much relationship with the slope (Figure 4.8), it can be characterized in between 20-40 degrees, of which this is a wide range that many species falls in.

According to aspect diagram (Figure 4.7) Oak's distribution related with the W-NW and W-SW. However from an ecological perspective this does not yield any meaningful criteria that can be successfully generalized

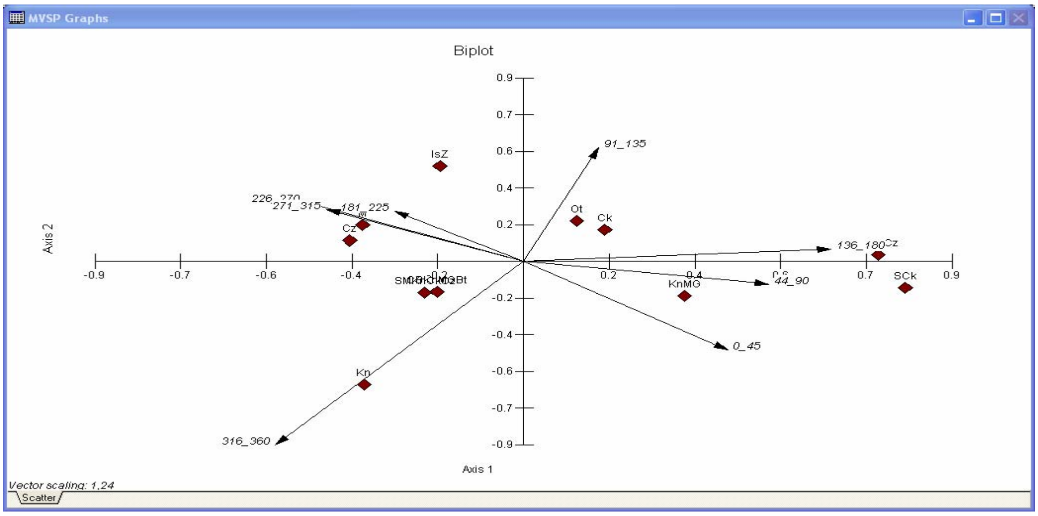

Figure 4.7. Graph of aspect vs. species distribution

- *Oriental Beech:*

According to DEM CCA diagram 1000-1500 m and 1500-2022 m range gives information about the distribution of the Oriental Beech. The figure 4.6 shows that it is not found below 1000 m.

The Slope diagram of CCA result (Figure 4.8) does not give any explanatory rule about the distribution of Oriental Beech.

Aspect CCA diagram gives very good information about the distribution of this species. Since Oriental Beech prefers humid environments thus humid aspects, they usually occur in the N-NW directions in the study region (Figure 4.7).

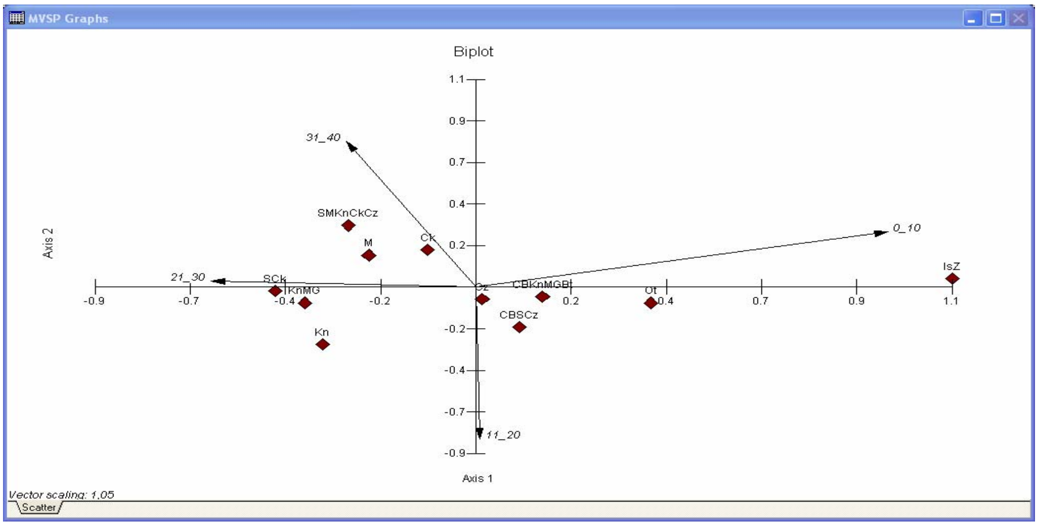

Figure 4.8. Graph of slope vs. species distribution

- *Taurus Cedar-Black Pine (TC-BP):*

CCA DEM diagram explains that TC-BP distribution occurs between 1500-2022 m elevations (Figure 4.6). The results of aspect and slope diagram are not giving clear rules for the distribution of TC-BP species.

- *Highly Degraded Taurus Cedar-Callabrian Pine (HD-TC-CP)*

Callabrian Pine usually occurs up to 1000 m elevation and Taurus Cedar prefers elevations higher than 1000 m. The diagram gives the distribution of these species between 500-1000 m ranges which is quite meaningful. But since these altitudes are low, these areas could easily affected by the human factor, which could be the reason of degradation.

According to the diagram of slope (Figure 4.8), it has no effect on the distribution of the HD-TC-CP. The occurrences of HD-TC-CP species are in the aspects facing S-SW direction according to diagram (Figure 4.7).

- *Over Degraded Oriental Beech-Oak-Carpinus Coppice (HD-OB-O-C-Copp):*

According to DEM CCA diagram (Figure 4.6), 0-1000 m is the elevation in which the HD-OB-O-C-Copp distribution usually occurs. Meanwhile, this range is also explaining the reason of degradation.

0-10° is the slope where the distribution of HD-OB-O-C-Copp species occurs according to CCA diagram (Figure 4.7).

The aspects facing in the west direction are generally humid and important for the deciduous forests. The result of CCA Aspect diagram shows that HD-OB-O-C-Copp is occurring in west aspects (Figure 4.8).

- *Taurus Cedar-Oak-Oriental Beech-Black Pine-Callabrian Pine (TC-O-OB-BP-CP):*

According to DEM CCA diagram 1000-1500 m is a suitable environment for the distribution of TC-O-OB-BP-CP which is consistent with the elevation range of the species composition (Figure 4.6)

Although TC-O-OB-BP-CP distribution does not show much relationship with the slope, it can be seen between 20-40 degrees. However this is a wide range that many species falls in (Figure 4.7).

Such as HD-OB-O-C-Copp case, the result of CCA Aspect diagram shows that these species occurring in west aspects (Figure 4.8).

Although the results of the Canonical Correspondence Analysis show a distinct separation between certain species qualitatively, it is not giving exact rules when the mixed forests are considered. For example it could successfully discriminate Callabrian pine from the species occurring in higher altitudes. But when considering the species occurring in high elevation this analysis become inadequate in discrimination of Oak from Taurus Cedar-Black Pine class. These results prevent the use of CCA in the post classification sorting stage of the vegetation classification.

4.3.3. Evaluation of Discriminant Analysis

Discriminant analysis is used to determine which variables discriminate between two or more naturally occurring groups. Multiple Discriminant Analysis (MDA) is used to classify a categorical dependent which has more than two categories, using as predictors a number of interval or dummy independent variables. MDA is sometimes called discriminant factor analysis or canonical discriminant analysis.

Discriminant analysis uses two types of variables; first one is a dependent "grouping variable" score in categorical form which is the Forest Management Classes in this study. Scores on a grouping variable are used to identify the groups by predicting the membership of each variable (Huberty, 1994). Second type of variable is a "discriminating variable", which is the independent data that gives the characteristics of the cases to be used to distinguish the groups.

Ultimately Discriminant Analysis (DA) is used to classify the given cases, and try to find out a linear equation or a series of linear equations that would be discriminating the groups. The backbone of DA is to find out the peculiar functions that would minimize within group variance and maximize the differences between groups.

The major assumptions that should be fulfilled in order to apply this analysis is that the cases should be independent, discriminating variables should have a multivariate normal distribution, and within-group variance-covariance matrices should be equal across groups.

Discriminant analysis explores the data in two steps; in the first step, the significance of the Discriminant functions are tested by comparing the matrix of total variances and covariances and the matrix of pooled within-group variances and covariances using a multivariate F-test. Once seeing that group means are statistically significant, functions had been determined by using the combination of variables to provide discrimination between groups.

It is known that Fisher's Linear Discriminant functions are the best for discrimination among all linear functions of the original variables (Kanal and Krishnaiah, 1982). Besides, its ease in practical applications, these functions is derived without an assumption of normality which makes it attractive to users.

Fisher (1936) (in Klecka, 1990), proposed to use linear combinations that maximize group differences while minimizing the variation within the groups while quantization of the discriminating functions. This idea leads the derivation of separate linear combinations which are classification functions for each group. Assigning each case into one of the group means assigning it to the group which it has the highest probability of belonging (Klecka, 1990). After assigning all of it into the groups, the discriminant analysis perform a classification process either to predict the group membership for the unknown cases or to test the accuracy of the classification procedure.

In the previous stages, the sample sets were prepared to be used in the statistical analysis with SPSS 11.0 software. The analysis was performed for each of the three data sets individually by using the environmental variables as discriminating variables. After applying steps of statistical analysis and checking the validity of the tests, DA performs a classification process with given cases and calculate the correctly classified points. The results of this DA classification shows that only 52.2% of the cases were classified correctly which was not successful in discrimination of the data. It is obvious that other discriminating variables are in need to increase the success of the DA, which was considered as the raw Landsat ETM bands. The degree of success of the DA classification results for each Data Set is given in Table 4.4. According to the table, addition of ETM Bands increased the performance of the analysis. Furthermore, among the three data sets, third one with ETM Bands classified the 90% of the cases correctly which is quite successful. Resulting functions of this data set 3 were considered to be used in the next steps and outputs of statistical analysis are going to be explained in detail in proceeding sections.

Table 4.4. DA classification results for each data set

	Without TM's	With TM's	% improvement
Data Set 1	49.7%	57.1%	7.4 %
Data Set 2	44.0%	63.1%	19.1 %
Data Set 3	52.2%	90.0%	37.8 %

Box's M test (Table 4.5), tests the assumption of homogeneity of covariance matrices. This test is very sensitive to meeting the assumption of multivariate normality. For the data below, the test is significant so it could be concluded that the groups do differ in their covariance matrices, which is an assumption of DA.

Table 4.5. Box's M test results

Test Results

Box's M		1226,386
F	Approx.	1,764
	df1	528
	df2	29607,39
	Sig.	,000

Eigenvalue shows the discrimination power of the function. The larger the eigenvalue, the greater the discrimination is. The square root of each eigenvalue is an indication of the length of the corresponding eigenvector. The % of variance column allows evaluating which canonical variable accounts for most of the spread, and % of variance column allows evaluating which canonical variable accounts for most of the spread. In this case the first eigenvector gives the 95.6% of the total variance explained by the variables. The eigenvalues for the Discriminant functions are given in Table 4.6.

Table 4.6. Eigenvalues of Discriminant functions.

Eigenvalues

		Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
Function	1	*17,540[a]*	*68,7*	*68,7*	*,973*
	2	*6,851[a]*	*26,8*	*95,6*	*,934*
	3	*,568[a]*	*2,2*	*97,8*	*,602*
	4	*,259[a]*	*1,0*	*98,8*	*,454*
	5	*,170[a]*	*,7*	*99,5*	*,381*
	6	*,095[a]*	*,4*	*99,8*	*,295*
	7	*,031[a]*	*,1*	*100,0*	*,173*
	8	*,011[a]*	*,0*	*100,0*	*,105*

a. First 8 canonical discriminant functions were used in the analysis.

Wilks's lambda is a multivariate measure of group differences by calculating the ratio of the within-groups sum of squares to the total sum of squares. The value of lambda is an inverse measure, which means while the values near zero denote the higher discrimination, the values close the one shows less discrimination. By looking at the Wilks's lambda values in Table 4.7, it could be understood that the discrimination power of band 3 is highest and northness is lowest among the variables. The significance of lambda could be tested by converting it into F distribution

In this table also the significances of the variables are given. The low significance value indicates the significant group differences. By looking at the significance values of discriminating variables in Table 4.7, the variable differences could be considered as significant.

Table 4.7. Wilks' lambda values

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
DEM	,525	19,330	8	171	,000
SLOPE1	,838	4,145	8	171	,000
GEOLOGY1	,884	2,808	8	171	,006
Northness	,919	1,895	8	171	,064
EASTNESS	,772	6,313	8	171	,000
TM1	,154	117,365	8	171	,000
TM2	,146	124,970	8	171	,000
TM3	,099	193,970	8	171	,000
TM4	,129	143,763	8	171	,000
TM5	,132	140,881	8	171	,000
TM7	,101	190,366	8	171	,000

In Table 4.8 Fisher Linear function coefficients are given. Columns of the table contain the coefficients of each group for a classification function. According to these coefficients, functions for each case are calculated and then case is assigned to the class whose value is higher.

Table 4.8. Fisher's Linear function Coefficients

Classification Function Coefficients

	Forest Management Map Classes								
	1	2	3	4	5	6	7	8	9
DEM	,0584	,0580	,0583	,0622	,0585	,0605	,0612	,0577	,0648
SLOPE	-1,0557	-1,0410	-1,1671	-1,0295	-1,0595	-1,0787	-1,1287	-1,1051	-1,0672
GEOLOGY	-2,5095	-2,5000	-2,7422	-3,3341	-2,8544	-2,6408	-2,5863	-2,4037	-2,6624
NORTHNESS	-5,0782	-6,0241	-5,3602	-1,2040	-5,6528	-5,1469	-4,6645	-5,6423	-5,7250
EASTNESS	-28,9139	-29,1696	-27,6386	-32,7034	-28,8484	-28,7784	-28,2694	-27,0282	-26,7224
TM1	22,4861	22,5667	22,7887	22,2715	22,7467	22,5377	22,9139	22,2503	22,4413
TM2	,8779	,8307	,9277	-,2989	,7608	,6428	,5998	1,1209	,6385
TM3	-11,3349	-11,2899	-10,9433	-9,4949	-11,1730	-11,0441	-11,6246	-10,8777	-10,6611
TM4	2,9218	2,9433	2,3456	1,9147	2,5977	2,7557	3,4394	2,3955	2,4434
TM5	-1,7643	-1,8289	-1,9257	-1,3092	-1,8427	-1,9104	-1,8952	-1,9103	-2,0542
TM7	2,3995	2,4121	2,2661	2,6287	2,6006	2,4396	2,6526	2,3517	2,3509
(Constant)	-1429,05	-1434,07	-1420,98	-1391,81	-1428,1	-1399,16	-1512,02	-1379,66	-1368,82

Fisher's linear discriminant functions

By using the coefficients of the Forest Management Classes in the Table 4.8, functions for each group could be written. The equation of first class is given as an example:

$$Group1 = DEM \times 0{,}584 - SLOPE \times 1{,}0557 - 2{,}5095 \times GEOLOGY - 5{,}0782 \times NORTHNESS - 28{,}9139 \times EASTNESS + 22{,}4861 \times TM1 + 0{,}8779 \times TM2 - 11{,}3349 \times TM3 + 2{,}9218 \times TM4 - 1{,}7643 \times TM5 + 2{,}3995 \times TM7 - 1429{,}05$$

To evaluate how well the Discriminant function works, analysis performs a classification by using the known cases. According to the results of the table 4.9 about 90% of the cases correctly classified when Data set 3 is used in the analysis. The result of the Discriminant Analysis of the other sets is given in Table 4.4.

Table 4.9. DA classification result of Data set 3

Classification Results[a]

				Predicted Group Membership									Total
				1	2	3	4	5	6	7	8	9	
Original	Count	MESCERE1	1	14	5	0	0	0	1	0	0	0	20
			2	3	16	0	0	0	1	0	0	0	20
			3	0	0	17	0	0	0	0	2	1	20
			4	0	0	0	20	0	0	0	0	0	20
			5	0	0	0	0	18	2	0	0	0	20
			6	0	0	0	0	0	19	0	1	0	20
			7	0	0	0	0	0	0	20	0	0	20
			8	0	0	0	0	0	0	0	19	1	20
			9	0	0	1	0	0	0	0	0	19	20
	%	MESCERE1	1	70	25	0	0	0	5	0	0	0	100
			2	15	80	0	0	0	5	0	0	0	100
			3	0	0	85	0	0	0	0	10	5	100
			4	0	0	0	100	0	0	0	0	0	100
			5	0	0	0	0	90	10	0	0	0	100
			6	0	0	0	0	0	95	0	5	0	100
			7	0	0	0	0	0	0	100	0	0	100
			8	0	0	0	0	0	0	0	95	5	100
			9	0	0	5	0	0	0	0	0	95	100

a. 90,0% of original grouped cases correctly classified.

4.4. The Use of DA Results in the Classification of Satellite Images

After gathering the Fisher linear discrimination functions from the results of Discriminant Analysis, the next step was to use these functions in the classification of the satellite images. The incorporation of these results with the satellite images was used in two different stages. In the stage 2 the satellite images directly classified only with the classification functions gathered from the DA. A geoformula (Appendix) is prepared to calculate the class value of each pixel and assign it into one of the nine classes which has a higher Fisher's Linear Function value. The output of this classification is given in Figure 4.9.

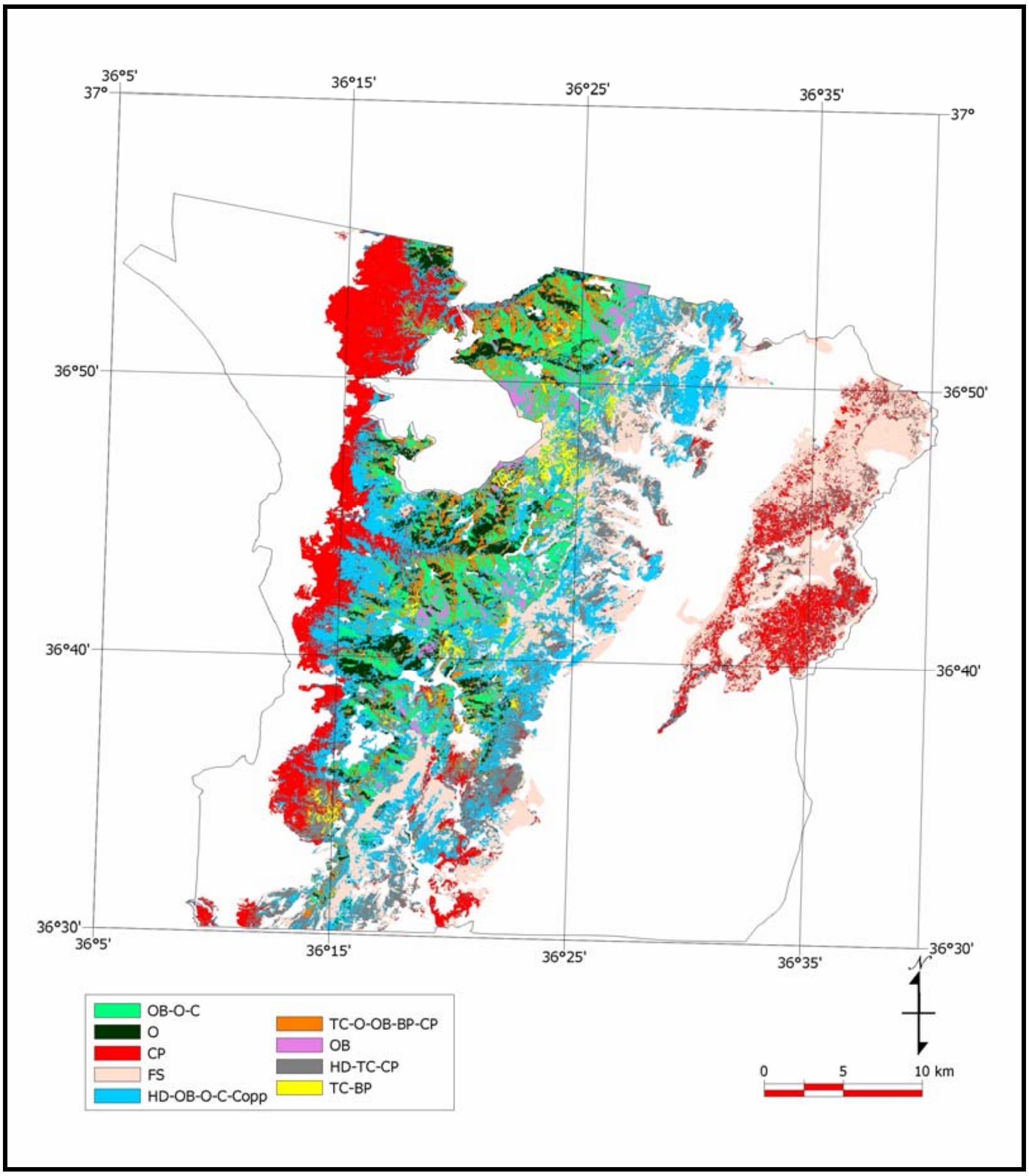

Figure 4.9. Figure of DA classification result (Stage 2)

In stage 3, both the outputs of Maximum likelihood and Discriminant analyses was used. The basic outputs of the maximum likelihood classification are class raster and distance raster. Class raster was constituted by assigning each cell into one of the predefined classes, whereas the distance raster gives the probability of each cell of belonging to the assigned class. If distance raster value of a pixel is one, this means that the pixel is 100 % correctly classified according to the Maximum Likelihood classifier. In Figure 4.10, areas with a green color shows the pixels having a probability of one, and the black areas are showing the pixels having a distance raster value less than 1. According to the histogram of the distance raster values; 47.38 % of the classified pixel values are one. Thus, these pixels could be considered as correctly classified with the Maximum Likelihood classifier.

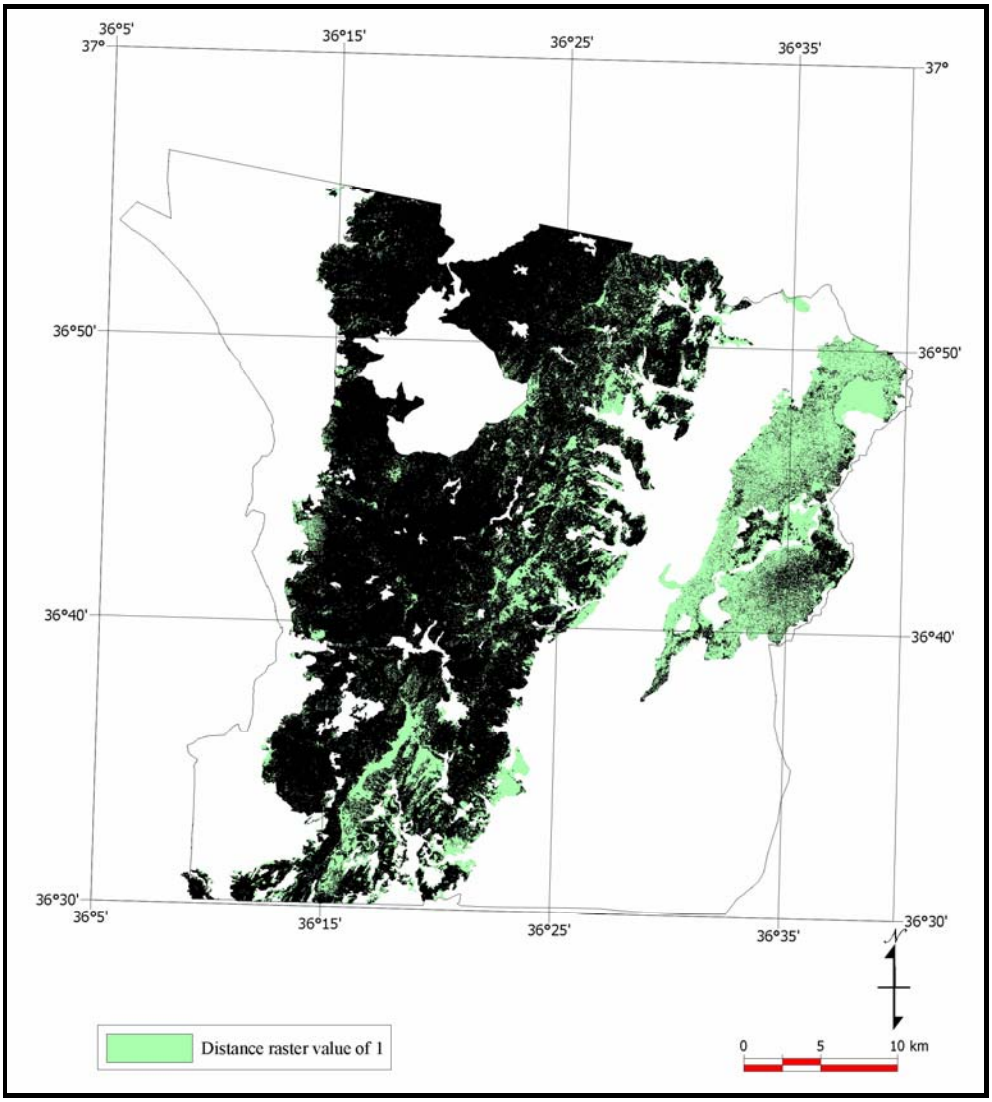

Figure 4.10. Distance raster showing the pixel values of 1

In this stage of this study, the pixels having a distance raster value less than one was reclassified according to the Fisher's linear functions which could be called as a Post

Classification Sorting. In the prepared geoformula, it is asked to take original classification values where the distance raster value is one, and take the "same data set" classification value of Stage 2 where the distance raster value is less than one. The result of the second stage classification with data set 3 is given in Figure 4.11.

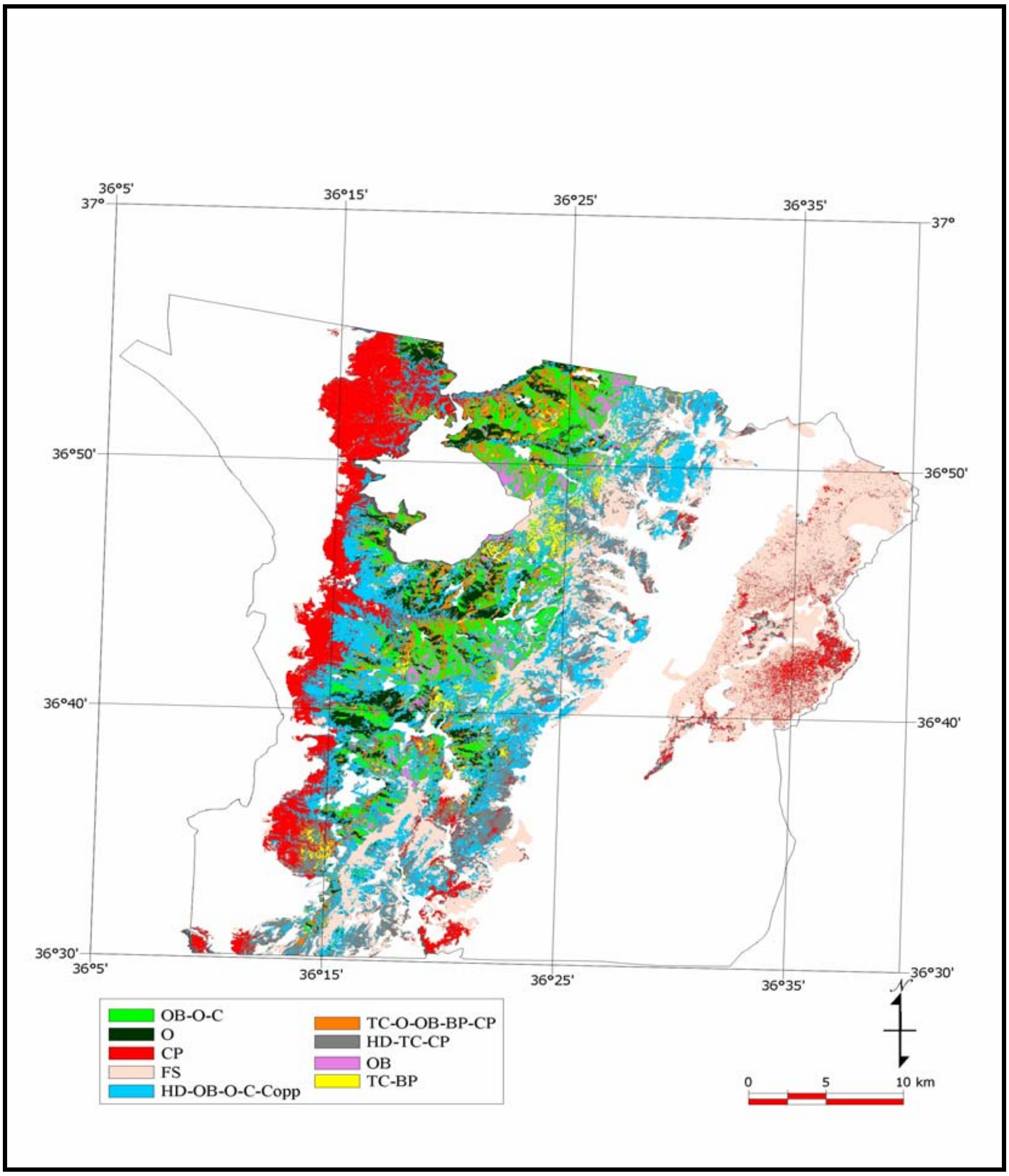

Figure 4.11. Classification map of Data Set 3 Stage 1

4.5. Accuracy Assessment

During the classification process many errors could be happened due to the quality of the training pixels, presence of mixed pixels and quality of ancillary data. Image classification is not considered as complete until an accuracy assessment has been performed as in Tso and Mather (2001), the accuracy term is explained as the level of agreement between the labels assigned by the classifier and the class allocations based on the ground data collected by the user.

The most common way to represent the accuracy of a Landsat classification is in the form of an error matrix or contingency table (Congalton et al, 1983). This is an array of numbers in a form of rows and columns providing a comparison between the results of classification and known reference data. The main diagonal of the matrix gives the correctly classified pixels. Dividing the number of correctly classified pixel into the total number of sample points, gives the overall accuracy of the classification. For calculating the accuracy of each class separately, producer's accuracy, and user's accuracy can be used. The producer's accuracy is calculated by dividing the entry (i,i) by the sum of column i, while the user's accuracy is obtained by dividing the entry (i,i) by the sum of row i (Tso and Mather, 2001). In addition to overall accuracy, Kappa statistics were also calculated which is a discrete multivariate technique used in accuracy assessment for determining if one error matrix is significantly different than another (Congalton and Green, 1999).

In order to assess the accuracy of classification, accurate ground truth or reference data is needed. The data set which used to measure the accuracy of this classification could be gathered from the field studies with the help of GPS's, from the large scaled maps or from the aerial photograph. In this study the accuracies of the classification results were calculated using the Data Set 4. This data set is prepared to check the accuracies of the classification results from the 1/25.000 scaled forest management maps with stratified random sampling and field studies carried out in the summer of year 2002 and 2003 by WWF-TR ecologists (Figure 4.11). In total, 225 points were used in the measurement of the accuracy.

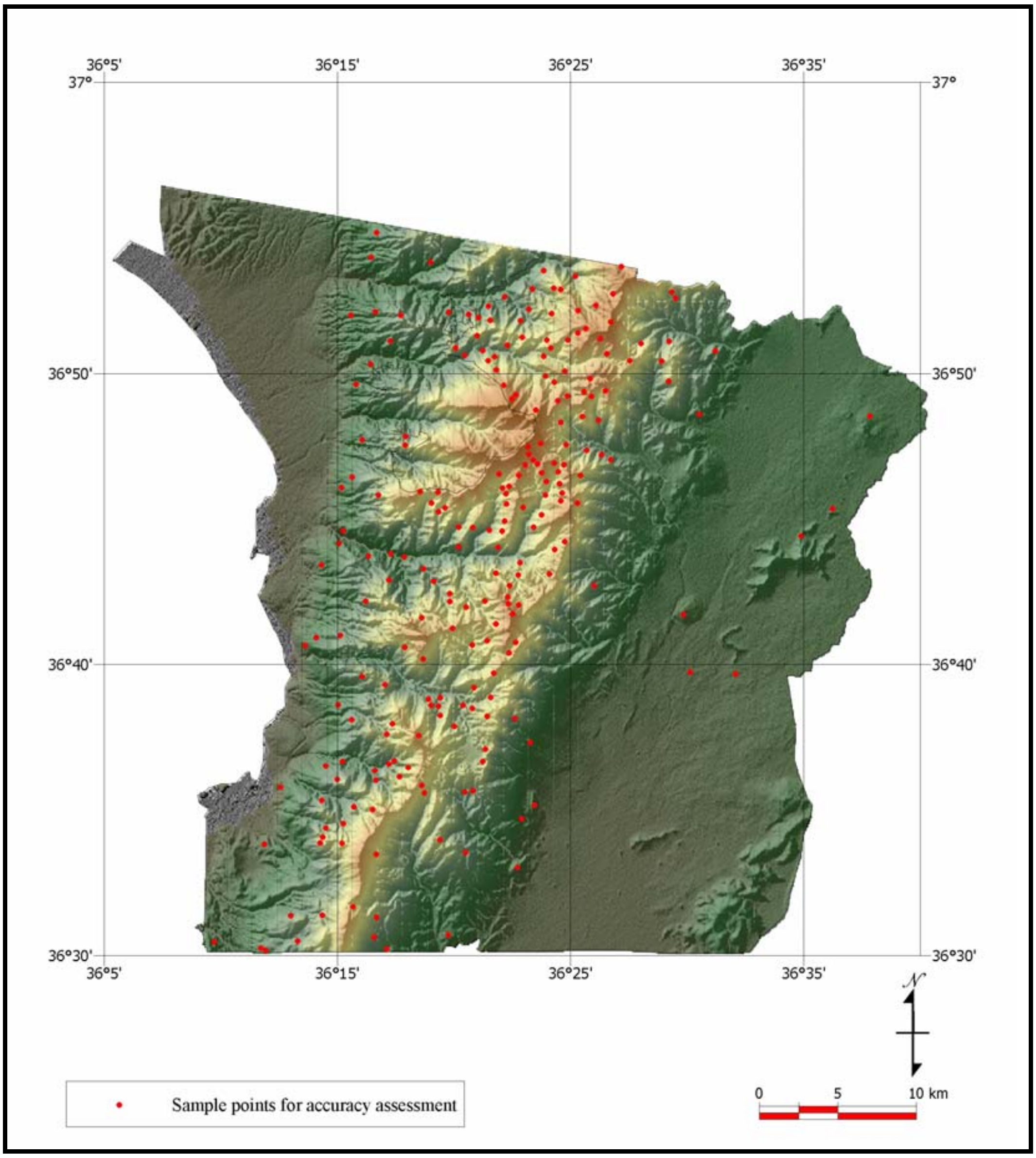

Figure 4.12. Sample points for accuracy assessment

4.5.1. Accuracy of Coarse Classification (Stage 1)

According to data set 4, the accuracy of the coarse classification result is 44.4 % and is presented in Table 4.10, with a 0.37 KHAT value.

Table 4.10. Error matrix of coarse classification

Classification \ Ground Truth	OB-O-C	O	CP	FS	HD-OB-O-C-Copp	TC-O-OB-BP-CP	OB	HD-TC-CP	TC-BP	Column Total	User's Accuracy (%)
OB-O-C	**4**	8	0	0	6	0	3	2	0	23	17,4
O	3	**1**	0	0	0	2	3	0	0	9	11,1
CP	0	0	**16**	0	1	0	0	2	2	21	76,2
FS	0	0	1	**23**	0	0	1	1	0	26	88,5
HD-OB-O-C-Copp	7	10	1	0	**10**	2	1	4	2	37	27,0
TC-O-OB-BP-CP	8	5	0	0	5	**13**	0	1	1	33	39,4
OB	2	0	1	0	0	0	**16**	0	0	19	84,2
HD-TC-CP	1	0	4	1	4	4	0	**12**	14	40	30,0
TC-BP	0	0	2	0	3	4	0	3	**5**	17	29,4
Row Total	25	24	25	24	29	25	24	25	24	**225**	
Producer's Accuracy (%)	16,0	4,2	64,0	95,8	34,5	52,0	66,7	48,0	20,8		44,4

4.5.2. Accuracy of Statistical analyses (stage 2)

The accuracies of statistical analyses classification are calculated with the same data set, and the error matrix for each classification is given in separate tables. Overall accuracy of the statistical analysis classification Data Set 1 Stage2 is 39.1 % and KHAT statistics is 0.32 (Table 4.11), where overall accuracy of the statistical analysis classification Data Set2 Stage2 is 37.3 % and KHAT statistics is 0.30 (Table 4.12), and overall accuracy of the statistical analysis classification Data Set3 Stage2 is 62.2 % and KHAT statistics is 0.58 (Table 4.13).

Table 4.11. Error matrix of statistical analysis classification Data Set 1, Stage 2 classification

	Ground Truth										
Classification		**OB-O-C**	**O**	**CP**	**FS**	**HD-OB-O-C-Copp**	**TC-O-OB-BP-CP**	**OB**	**HD-TC-CP**	**TC-BP**	**Column Total**	**User's Accuracy (%)**
	OB-O-C	**12**	13	1	0	4	4	10	0	0	44	27,3
	O	0	**0**	0	0	1	0	0	0	0	1	0,0
	CP	0	1	**11**	4	5	0	0	6	0	27	40,7
	FS	0	0	1	**10**	0	0	1	0	0	12	83,3
	HD-OB-O-C-Copp	1	1	0	0	**4**	1	1	1	0	9	44,4
	TC-O-OB-BP-CP	5	3	0	0	5	**12**	1	2	2	30	40,0
	OB	6	5	0	0	2	3	**11**	0	0	27	40,7
	HD-TC-CP	1	0	11	4	5	1	0	**14**	8	44	31,8
	TC-BP	0	1	1	6	3	4	0	2	**14**	31	45,2
	Row Total	25	24	25	24	29	25	24	25	24	**225**	
	Producer's Accuracy (%)	48,0	0,0	44,0	41,7	13,8	48,0	45,8	56,0	58,3		39,1

Table 4.12. Error matrix of statistical analysis classification Data Set2, Stage 2 classification

	Ground Truth										
Classification		**OB-O-C**	**O**	**CP**	**FS**	**HD-OB-O-C-Copp**	**TC-O-OB-BP-CP**	**OB**	**HD-TC-CP**	**TC-BP**	**Column Total**	**User's Accuracy (%)**
	OB-O-C	**7**	11	1	0	0	1	8	1	0	29	24,1
	O	8	**5**	0	0	8	8	0	0	0	29	17,2
	CP	0	0	**17**	2	8	0	0	6	4	37	45,9
	FS	0	0	1	**20**	0	0	1	1	0	23	87,0
	HD-OB-O-C-Copp	1	1	0	0	**4**	1	0	2	1	10	40,0
	TC-O-OB-BP-CP	2	3	1	0	3	**1**	0	0	6	16	6,3
	OB	6	4	0	0	3	7	**15**	0	0	35	42,9
	HD-TC-CP	1	0	5	2	3	7	0	**15**	13	46	32,6
	TC-BP	0	0	0	0	0	0	0	0	**0**	0	0,0
	Row Total	25	24	25	24	29	25	24	25	24	**225**	
	Producer's Accuracy (%)	28,0	20,8	68,0	83,3	13,8	4,0	62,5	60,0	0,0		37,3

Table 4.13. Error matrix of statistical analysis classification Data Set3, Stage 2 classification

Classification \ Ground Truth	OB-O-C	O	CP	FS	HD-OB-O-C-Copp	TC-O-OB-BP-CP	OB	HD-TC-CP	TC-BP	Column Total	User's Accuracy (%)
OB-O-C	**17**	5	0	0	4	5	2	0	2	35	48,6
O	4	**16**	1	0	3	4	0	0	0	28	57,1
CP	0	0	**20**	1	1	0	0	3	4	29	69,0
FS	0	0	0	**23**	0	0	1	2	0	26	88,5
HD-OB-O-C-Copp	0	1	1	0	**11**	3	0	5	2	23	47,8
TC-O-OB-BP-CP	2	1	0	0	3	**8**	0	1	2	17	47,1
OB	1	1	0	0	1	0	**19**	0	0	22	86,4
HD-TC-CP	0	0	2	0	4	0	0	**13**	1	20	65,0
TC-BP	1	0	1	0	2	5	2	1	**13**	25	52,0
Row Total	25	24	25	24	29	25	24	25	24	**225**	
Producer's Accuracy (%)	68,0	66,7	80,0	95,8	37,9	32,0	79,2	52,0	54,2		62,2

4.5.3. Accuracy of Integration (stage 3)

Overall accuracy of the Integration stage for Data Set1 is 32.9 % with a KHAT statistics of 0.25 (Table 4.14), where the overall accuracy of the Integration stage for Data Set 2 is increased to 39.1 % (KHAT statistics is 0.32) (Table 4.15) and the overall accuracy of the Integration stage for Data Set3 is 64,00 % with a KHAT statistics of 0.60 (Table 4.16).

Table 4.14. Error matrix of Data Set1, integration Stage classification

Classification	OB-O-C	O	CP	FS	HD-OB-O-C-Copp	TC-O-OB-BP-CP	OB	HD-TC-CP	TC-BP	Column Total	User's Accuracy (%)
OB-O-C	**12**	14	1	0	5	4	9	0	0	45	26,7
O	0	**0**	0	0	1	0	3	0	0	4	0,0
CP	0	1	**9**	1	5	0	0	6	0	22	40,9
FS	0	0	2	**0**	0	0	0	0	0	2	0,0
HD-OB-O-C-Copp	1	1	0	1	**4**	1	1	1	0	10	40,0
TC-O-OB-BP-CP	5	3	0	0	5	**12**	1	2	2	30	40,0
OB	6	5	1	21	2	3	**10**	1	0	49	20,4
HD-TC-CP	1	0	11	0	5	1	0	**13**	8	39	33,3
TC-BP	0	0	1	1	2	4	0	2	**14**	24	58,3
Row Total	25	24	25	24	29	25	24	25	24	**225**	
Producer's Accuracy (%)	48,0	0,0	36,0	0,0	13,8	48,0	41,7	52,0	58,3		32,9

Columns: Ground Truth

Table 4.15. Error matrix of Data Set2, Integration Stage classification

Classification	OB-O-C	O	CP	FS	HD-OB-O-C-Copp	TC-O-OB-BP-CP	OB	HD-TC-CP	TC-BP	Column Total	User's Accuracy (%)
OB-O-C	7	11	1	0	0	1	7	1	0	28	25,0
O	8	5	0	0	8	8	0	0	0	29	17,2
CP	0	0	17	0	8	0	0	6	4	35	48,6
FS	0	0	1	23	0	0	1	1	0	26	88,5
HD-OB-O-C-Copp	1	2	0	0	5	2	0	2	1	13	38,5
TC-O-OB-BP-CP	2	2	1	0	2	0	0	0	6	13	0,0
OB	6	4	0	0	3	7	16	0	0	36	44,4
HD-TC-CP	1	0	5	1	3	7	0	15	13	45	33,3
TC-BP	0	0	0	0	0	0	0	0	0	0	0,0
Row Total	25	24	25	24	29	25	24	25	24	225	
Producer's Accuracy (%)	28,0	20,8	68,0	95,8	17,2	0,0	66,7	60,0	0,0		39,1

Columns: Ground Truth

Table 4.16. Error matrix of Data Set3, Integration Stage classification

		Ground Truth										
		OB-O-C	**O**	**CP**	**FS**	**HD-OB-O-C-Copp**	**TC-O-OB-BP-CP**	**OB**	**HD-TC-CP**	**TC-BP**	**Column Total**	**User's Accuracy (%)**
Classification	**OB-O-C**	**16**	4	0	0	3	5	4	2	1	35	40,00
	O	4	**16**	0	0	4	3	1	0	0	28	46,43
	CP	0	0	**21**	0	1	0	0	2	3	27	66,67
	FS	0	0	1	**23**	0	1	0	2	0	27	85,19
	HD-OB-O-C-Copp	0	1	1	0	**12**	3	0	4	2	23	47,83
	TC-O-OB-BP-CP	4	2	0	0	2	**9**	0	0	0	17	41,18
	OB	1	1	0	0	1	0	**19**	0	0	22	86,36
	HD-TC-CP	0	0	1	1	4	1	0	**12**	2	21	52,38
	TC-BP	0	0	1	0	2	3	0	3	**16**	25	44,00
	Row Total	25	24	25	24	29	25	24	25	24	**225**	
	Producer's Accuracy (%)	64,0	66,7	84,0	95,8	41,4	36,0	79,2	48,0	66,7		64,00

CHAPTER 5

DISCUSSION

The purpose of this research was to discriminate the classes of vegetation in formation or dominant species level. Even as distinguishing the broad-leaved forest from needle-leaved forest is maintained with higher accuracies, (Dymond et. al., 2002; Jensen, 2000), classifying them within themselves is not a very simple task as all of them are chlorophyll bearing spectral objects having similar spectral reflectance curves. The aim of this chapter is to endure step by step through this rigorous vegetation classification procedures used in this study while lightening the possible errors with their reasons, and evaluating the significance and success of all the sequential process.

5.1. Data

In this study a new approach is experimented on Amanos Mountains to map the vegetation types which are discriminated using Landsat ETM images with a spatial resolution of 30 meters. Globally in vegetation classification suitability of both the spectral and spatial resolution of the image to the desired map content and scale are of the key parameters of high accuracy. In recent operational satellites spectral choice is not too much as either Landsat like satellites operating in visible region or hyperspectral sensors should be chosen. On the other hand in spatial domain theoretically the spatial resolution of the image must be smaller than the size of the feature being classified to fully cover the classified objects. However in practice area that could be obtained by classifying the one Landsat image pixel is 900 m^2 in which the gradations from one species to another or even complete different association of 3 other species is possible. Hence these naturally mixed pixels would always create problems and they will reduce to some extent the accuracy of classifications due to spectral confusion they create.

Bearing the above constraints in mind the possible optimum scale that can be obtained from result of these images is around 1/100.000. Also the scale of the ancillary and ground truth data should be coherent to explain the variance within the data set and to get consistent results.

Forest management maps with a 1/25.000 scale was the main input of this study, which is a result of detailed field studies discriminating the forest stands according to species, age class and canopy of the vegetation type. After digitization of these maps, 26 vegetation classes were attained but since some of the classes were not ecologically meaningful and besides gaining this detail is impossible with the spatial and spectral resolution of the Landsat images, this layer is generalized into 10 variables of interest, including settlement and agriculture. Despite this reclassification is based on the ecological criteria and removes the unnecessary details of the map, this process is completely expert dependent causing heterogeneous vegetation groups, which makes the discrimination difficult with the satellite images. Another problem about the forest management map is that some of the classes could be named as another class intentionally to protect some species by the forest management department staff, which in turn completely hampers the efforts to collect correct training signatures, statistically analyze the data and to collect objective accuracy assessment data sets. Both deciduous shrublands and maquis with the coracious leaves are called "Coppice" although ecologically they are two separate classes.

Similar scale and reclassification problems are also effectual for the lithology layer. The only complete coverage of the region was available at a scale of 1/100.000 and consists of 17 classes some of which are distinct in geological aspect but of no interest or relation to ecology. The best effort to use this lithostratigraphical map as an environmental predictor was to reclassify into some ecologically meaningful legend, which could be the rock material itself. Consequently this reclassification was completely based on expert field knowledge, which could easily be addressed as a deficiency about the universal appliance of this variable. The resulting map consists of 11 classes that may lead to the loss of geological information but seems to be informative behind the scope of ecological interest.

Elevation is a valuable data that could increases the accuracy of classification in a significant amount, since the elevation differences cause climatic variations influencing the vegetation type, aspect differences creates differential illumination regarding reflectance amount and sun capture, slope differences yield in different density of vegetation and different reflectance again. In this study Shuttle Radar Topography

Mission DEM is used which is having the highest resolution amongst all publicly available sources. As stated by the product description in web site of Jet Propulsion Laboratory at NASA the vertical accuracy of this product is approximately 10 meters in rough terrain ranging down to 5 meters in flat areas. These values of vertical accuracy sufficiently fulfills the needs of such an ecological characterization of the cover units as the resulting data would have a scale of 1:100.000. However the derivatives of this elevation model namely the Slope and Aspect, were also produced from the SRTM data. As expected the vertical accuracy constraints have migrated also to these derivatives, but the final mapping scale compensates these small errors.

5.2 Classification

The gathered accuracy result of vegetation classification with Maximum likelihood classifier is calculated as 44.4 %, which could be considered as a lower value. This result could be affected by various reasons.

Although the Landsat ETM bands of the region registered to the UTM coordinates, classification was performed on the unregistered images, to preserve the original digital number values. Other layers which are used as an ancillary data were all registered to the 1/25.000 scaled topographical maps. While selecting the spectral signatures from the image and in the forming the Discriminant Analysis samples, this could result a minute amount of shifting from the original coordinate which is ignored in the study, but controlled individually at every single observation point. However, although minor, there exists a chance of misalignment, which could be attributed as operator error that would migrate down to final accuracy.

Another main factor of this low accuracy is probably the rigidity of the legend used. The classification legend is based on the aggregated classes of forest inventory map, hence the classes are quite different in definition and on map, but the 44.4 % accuracy shows that they are not so much different in real world. The different pine species are the least distinguishable classes whereas forest soil is the best discriminated class among the others. The heterogeneity of the forest management classes also affects the classification accuracy. An additive factor related to the effect of the forest management maps on classification accuracy is in forest management maps the canopy closure of the forest units are also used. If the forest has open canopy with less than 40 % crown closure it is called degraded. While merging these degraded areas class type are generally disregarded. The degree of subjectivity in forest management classes such as “highly degraded” does not reveal any quantitative difference relative to “less degraded” class. Different stages of degradation lead some problems in the classification since according to the canopy closure of the vegetation, soil and lithology type becomes a more important factor helping the discrimination of vegetation.

The maximum likelihood classifier is one of the most popular methods of classification in remote sensing but one of the drawbacks of the method is normal distribution assumed for the each class in the method but in nature matching with normal distribution is a little probability, where the same assumption is valid for Discriminant analysis also.

Both maximum likelihood classification and DA classification only consider the pixel itself regardless of the neighboring pixels. When classifying the contents of imagery, there are only a few attributes accessible to human interpreters. For any single set of imagery these are shape, size, color, texture, pattern, shadow, and association. Traditional image processing techniques incorporate only color (spectral signature) and perhaps texture or pattern into a process. Using ancillary data could be accepted as a way of adding contextual data into classification. As an example some of the neighboring pixels having same land cover type could be assigned as different classes due to some additional factors. Elevation, slope and aspect are the data types gradually varying rather than nominal nature. When these data types added to classification as ancillary data, similar pixels are classified in the same class regardless of the additional factors. Even though the spectral bands shows correlation with each other, their correlation with elevation, slope and aspect is low, means that they are covering different information channels, which is in fact quite valuable to discriminate the vegetation species type.

5.3 Analysis and Accuracy

Low accuracy value of maximum likelihood classification forced researchers to search alternative methods to be used in the classification. Addition of ancillary data in the classification by increasing the number of channels is a widely used method and it is believed that increases the accuracy. But addition of non-spectral information into classification could create some problems and require additional field study since the results are unpredictable. Instead of direct addition of the ancillary data, it is preferred to examine the statistical properties of the sample points and relationships of the environmental variables and species distribution.

Another recommended way to increase the accuracy is to use the multi-seasonal satellite images. Species show different spectral responses in different seasons according to their phenological stages which could be used to separate the species with similar spectral response (Nagendra, 2001). The selection of species specific phenological seasons could increase the accuracy obviously but also the cost of the study.

Canonical Correspondence Analysis is used before applying the Discriminant Analysis to see if there exists any relationship between environmental variables and the species distribution. Bi-plots of the CCA analysis showed that the distribution of vegetation is highly effected by DEM, slope, and aspect variables. Since this analysis doesn't give any quantitative result Discriminant analysis is applied on the data sets.

Statistical analysis classification was performed in two different stages. In the first stage (Stage 2) Discriminant analysis functions are used to classify the image, and in the second stage (Stage 3) only the misclassified pixels of the Maximum likelihood method is replaced with the first stage results. It was expected that the second stage results will be higher than the first stage. Unfortunately according to the accuracy results of these two stages there is not any significant increase is observed (Table 5.1). This means that the pixels having a distance raster value of one is classified as the same class according to the Discriminant Analysis results.

Table 5.1. Accuracy changes through different classification methods

	Stage 1 (MLC Classification)	**Stage 2** (Stat. Analy.Class)	**Stage 3** (Integration of st1&st2**)**
MLC	44,4 %	-	-
Data Set 1 (random stratified)	-	39,1 %	32,9 %
Data Set 2 (random stratified area normalized)	-	37,3 %	39,1 %
Data Set 3 (random stratified area normalized, expert oriented**)**	-	62,2 %	64 %

First and second data set which are used in the Discriminant Analysis are prepared by using stratified random sampling method. It could generate problems especially in the inter-grading regions, mixed pixels and in the areas where land cover class has changed since the date of the formation of the forest management maps. Vegetation map is a choropletic map which means that each polygon has a discrete value (Hutchinson, 1982). So each polygon has one attribute but actually these polygons could include different vegetation or land cover classes with a little difference in their nature. In random sampling, points could have an attribute of the polygon but the actual case in the nature could be different. It is seen that doing analysis with this data sets decreases the accuracy. When the sample points selected using an expert knowledge from the most representative areas of that groups the increase of accuracy is significant.

In this study only 10 % of the ground truth data were collected from the field and the remaining was dependent on forest management maps. Selection of ground truth data was performed by stratified random sampling thus again it is difficult to be sure about the attributes of those points because of the nominal nature of the reference maps. While calculating the accuracy of classification these mislabeled samples would decrease the accuracy of the classification.

When the accuracies of the individual classes are examined, maximum increase is observed in OB-O-C class with a 62.5 % and then 48 % in O class. On the other hand in heterogeneous classes the individual accuracies do not change much which means that discrimination of these classes are troublesome with all classification methods. As a result, overall accuracy increase between coarse classification and fine classification are around 20 %, but increase in producers and users accuracies of individual classes are in significant amount, which helps mapping the vegetation except the problematic classes.

It could be thought that the amount of increase in vegetation classification is not that much high when considering the effort spend on this issue. By using only the Landsat bands accuracy could be increased up to 44.4 % in a species level vegetation classification. But if the discrimination is in floristic level, classification issue gets more complicated and addition of additional information and further improvements such as use of statistical analysis is highly required.

Even though this application increased the overall accuracy of the classification in Amanos Mountains with these environmental variables; different study areas could require different variables affecting distribution of vegetation. For the regional scale biodiversity studies the minimum mapping unit of the study could be increased up to

100 ha, which require filtering of the resultant map. As the extent of the study increased, the level of detail must be deceased (Nagendra, 2001). Filtering process removes the unnecessary detail of the map, removing or assigning the outlier pixels to majorities and consequently it directly increases the accuracy while decreasing the level of detail.

Generalizations of the legend units stands as a way to increase the general accuracy of the produced maps, however they are standing far away from the real purpose of the classification efforts. Although this study demonstrates that with a moderate effort level and with the help of few environmental variables, the overall accuracy increases nearly 50 % from its initial stage. Where as it is obvious that addition of more environmental variables would yield in a more accurate mapping campaign of the species, bearing in mind that the statistical models will be saturated in further stages, yielding in such a fact that no more effort or variable inclusion would improve the accuracy.

CHAPTER 6

CONCLUSION

Improvement in vegetation classification accuracy is achieved with the use of ancillary data. Although the spectral resolution of Landsat images is considered as low in species level classification, use of additional information is increasing the discrimination power of Landsat bands. Ancillary data could be used to reduce the effects of correlation between spectral bands and as a way of adding contextual information in the analysis stage.

The improvement in classification could be performed in various ways, but in this study environmental variables are incorporated into classification process by modeling the effect of these variables via Discriminant analysis. The result of Maximum Likelihood Classification, which is named as Stage 1 classification is 44.4 %. After the use of Fisher's Discriminating functions the accuracy is increased up to 62.2 % in the Stage 2. Integrating these two classification map give the higher accuracy with the usage of distance raster which is a bi-product of Maximum Likelihood Classifier and named as Stage 3. The resulting classification map both includes statistical outcome of the environmental variables and information of satellite bands with a value of 64 %.

The composition and the structure of vegetation could be influenced by various environmental variables hence there is no perfect correlation between environment and species, since some species are tolerant to varying extreme conditions. The correlation between species and environmental variables helped in mapping the vegetation at unobserved locations; ultimately with the use of elevation, slope, aspect, and lithological units as an ancillary data, 19.6 % increase in the vegetation classification was obtained in this study.

Although the overall increase in the accuracy is about 20 %, in classes such as Oriental Beech-Oak-Carpinus (OB-O-C), Oak (O), Callabrian Pine (CP), Tauruc Cedar-Black Pine

(TC-BP) this value increases up to 62.5 %, whereas in Highly Degraded Oriental Beech-Oak-Carpinus-Coppice (HD-OB-O-C-Copp), Taurus Cedar-Oak-Oriental Beech-Black Pine-Callabrian Pine (TC-O-OB-BP-CP) and Highly Degraded Taurus Cedar-Callabrian Pine (HD-TC-CP) classes the increase is not noticeable. From these values it could be concluded that the approach gives good result in relatively homogeneous classes which includes less than 3 different species.

This method could be recommended for areas where the traditional approaches are unsuccessful and where laborious field work is necessary.

REFERENCES

Aksay, A., Tekeli, O., Ürgün, B.M., Işık, A. 1988. Amanoslar'ın Paleozoik Birimleri ve Mezozoyik Platform Karbonat İstifleri. MTA Genel Müd. Jeoloji Etüdleri Dairesi.

Carpenter, G. A., Gjaja, M.N., Gopal, S., Woodcock, C.E. 1997. Art Neural Networks for Remote Sensing: Vegetation Classification from Landsat TM and Terrain Data. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 2.

Congalton, R. G., Oderwald, R.G., Mead, R. 1983. Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques. *Photogrammetric Engineering and Remote Sensing.* Vol. 49, No.12 1671-1678.

Congalton, R. G., Green, K., 1999, Assessing the Accuracy of Remotely Sensed Data. Lewis Publisher, London

Cingolani, A.M., Renison, D., Zac, M.R., Cabido, M.R., 2004. Mapping vegetation in heterogeneous mountain rangeland using landsat data: an alternative method to define end classify land-cover units. *Remote Sensing of Environment*. 92, 84-97.

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

De Bruin, S., Gorte, B.G. H., 2002. Probabilistic image classification using geological map units applied to land-cover change detection, *International Journal of Remote Sensing.* Vol. 21, No. 12, 2389-2402.

Dirnböck, T., Dullinger, S., Gottfried, M., Ginzler, C., Grabherr, G., 2003. Mapping alpine vegetation based on image analysis, topographic variables and Canonical Correspondence Analysis. *Applied Vegetation Science* 6: 85-96.

Domac, A., Zeydanli, U., Yeşilnacar, E., Süzen, M.L., 2004, Integration and Usage of Indices, Feature Components and Topography in Vegetation Classification for Regional Biodiversity Assessment. ISPRS 2004, 12-23 July 2004, Istanbul (published in CD form as PDF)

Dorren, L.K.A., Maier, B., Seijmonsbergen, A.C., 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. *Forest Ecology and Management.* 183, 31-46

Dymond, C. C., Mladenoff, D.J., Radeloff, V. C., 2002. Phenological differences in Tasseled Cap indices improve deciduous forest classification. *Remote Sensing of Environment*. 80, 460-472

Fahsi, A., Tsegaye, T., Tadesse, W.,Coleman, T., 2000. Incorporation of digital elevation models with Landsat-TM data to improve land cover classification accuracy. *Forest Ecology and Management*. 128, 57-64.

Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*. 19: 474-499.

Hietel, E., Waldhardt, R., Otte, A., 2004, Analysing land-cover changes in relation to environmental variables in Hesse, Germany, *Landscape Ecology* 19:473-489.

Hoersch, B., Braun, G., Schmidt, U., 2002. Relation between landform and vegetation in alpine regions of Wallis, Switzerland. A multiscale remote sensing and GIS approach. *Computers, Environment and Urban Systems*. 26, 113-139.

http://iss.sfo.jaxa.jp/shuttle/flight/sts99/mis_srtm_e.html (visited on 18. 08.2004)

http://www.wwf.org.tr/en/ormanlar_dunya_sn_ad.asp (visited on 23. 10.2004)

Huberty, C.J., 1994. Applied Discriminant Analysis. John Wiley & Sons, INC. New York, Chichester, Brisbane, Toronto, Singapore.

Hutchinson, C.F., 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering and Remote Sensing*. 48, 123-139.

Janssen, L.L.F., Jaarsma, M.N., van der Linden, E.T.M., 1990. Integrating Topographic Data with Remote Sensing for Land-Cover Classification. *Photogrametric Engineering & Remote Sensing*. Vol.56, No.11

Jensen, J.R., 2000. Remote sensing of environment: an earth resource perspective. Prentice Hall series in geographic information science. pp.333-373. New Jersey:Prentice Hall.

Kääb, A., 2004. Combination of SRTM3 and repeat ASTER data for deriving alpine glacier flow velocities in the Bhutan Himalaya. *Remote Sensing of Environment*. Article in press.

Kanal, L.N., Krishnaiah, P.R., 1982. HandBook of Statistics 2. Classification, Pattern recognition and Reduction of Dimensionality. North-Holland Publishing Company. Amsterdam, New York, Oxford

Klecka, W.R., 1990. Discriminant Analysis. SAGE Publications. Newbury Park, London, New Delhi

Kimmins, J.P., 1997. Forest Ecology: A Foundation for Sustainable Management. Second edition. Prentice Hall, Upper Saddle River, New Jersey, USA.

Lewis, M. M., 1998, Numerical classification as an aid to spectral mapping of vegetation communities*. Plant Ecology*, 136: 133-149.

Liu, X, H., Skidmore, A. K., Van Oosten, H., 2002. Integration of classification methods for improvement of land-cover map accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing.* 56 257-268

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

Maselli, F., Rodolfi, A., Romanelli, B.S., Conese, C., 2000. Classification of Mediterranean vegetation by TM and ancillary data for the evaluation of fire risk. *International Journal of Remote Sensing.* 21: 3303-3313.

Moore, I.D., R.B. Grayson., A.R.Ladson, 1991. Digital Terrain Modelling: a review of hydrological and biological applications, Hydrological Processes, 5:3-30

MVSP User's Manuel., 1986-1998. Kovach Computing Services, Pentraeth, Wales, UK

Nagendra, H. 2001. Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*. 22: 2377-2400.

Pfeffer, K., Pebesma, E.J., Burrough, P.A., 2003. Mapping alpine vegetation using vegetation observations and topographic attributes*. Landscape Ecology*. 18, 759-776

Pinder, J.E., Kroh, G. C., White, J.D., Basham May, A. M., 1997. The relationship between vegetation type and topography in Lassen Volcanic National Park. *Plant Ecology*. 131, 17-29

Roberts, D. W. 1986. Ordination on the basis of fuzzy set theory. *Vegetatio* 66:123-31.

Strahler, A., 1980, The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment*, 10, 1135-1163.

Shrestha, D.P., Zinc, A.J., 2001. Land use classification in mountainous areas: integration of image processing, digital elevation data and field knowledge (application to Nepal). *JAG.* Volume 3. Issue1.

Tayku, M., Abia, S, I., Kitayama, K., 2002. Effects of topography on tropical lover montane forest under different geological conditions on Mount Kinabalu, Borneo*. Plant Ecology*. 159: 35-49

Ter Braak, C.J.F., 1986. Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. *Ecology*. 67(5). Pp.1167-1179

Tso, B., Mather, P.M., 2001. Classification Methods for Remotely Sensed Data. Taylor and Francis Inc, New York, London.

Yılmaz, Y., Gürpınar, O., Yiğitbaş, E. 1988. Amanos Dağları ve Maraş Dolaylarında Miyosen Havzalarının Evrimi ; Türkiye Petrol Jeologları Derneği Bülteni., c.1/1, s.52-72.

Zeydanlı, U., Domac., A. 2004. Landscape. In GAP Biodiversity Research Project 2001/2003 Final Report. (ed Welch, H.J.)

Wang, G., Gertner, G., Xiao, X., Wente, S., Anderson, A, B., 2001. Appropriate Plot Size and Spatial Resolution for Mapping Multiple Vegetation Types. *Photogrametric Engineering & Remote Sensing*. Vol.57, No.5.

Whittaker, R.H., 1970. Communities and ecosystems. Current concepts in Biology Series. The Macmillan Company, London. p.p162

APPENDIX

```
[GeoFormula]

Script=if ( s1_class_1_mref_Value > s1_class_2_mref_Value )
Script__2=Value = s1_class_1_mref_Value
Script__3=else
Script__4=Value = s1_class_2_mref_Value
Script__5=if ( s1_class_3_mref_Value > Value )
Script__6=Value = s1_class_3_mref_Value
Script__7=else
Script__8=Value = Value
Script__9=if ( step1_class_5_Value > Value )
Script__10=Value = step1_class_5_Value
Script__11=else
Script__12=Value = Value
Script__13=if ( step1_class_6_Value > Value )
Script__14=Value = step1_class_6_Value
Script__15=else
Script__16=Value = Value
Script__17=if ( step1_class_8_Value > Value )
Script__18=Value = step1_class_8_Value
Script__19=else
Script__20=Value = Value
Script__21=if ( step1_class_9_Value > Value )
Script__22=Value = step1_class_9_Value
Script__23=else
Script__24=Value = Value
Script__25=if ( step1_class_10_Value > Value )
Script__26=Value = step1_class_10_Value
Script__27=else
Script__28=Value = Value
Script__29=if ( step1_class_11_Value > Value )
Script__30=Value = step1_class_11_Value
Script__31=else
Script__32=Value = Value
Script__33=
Script__34=if
```

```
Script__35=( Value == s1_class_1_mref_Value )
Script__36=Value = 1
Script__37=else if
Script__38=( Value == s1_class_2_mref_Value )
Script__39=Value = 2
Script__40=else if
Script__41=( Value == s1_class_3_mref_Value )
Script__42=Value = 3
Script__43=else if
Script__44=( Value == step1_class_10_Value )
Script__45=Value = 4
Script__46=else if
Script__47=( Value == step1_class_11_Value )
Script__48=Value = 5
Script__49=else if
Script__50=( Value == step1_class_5_Value )
Script__51=Value = 6
Script__52=else if
Script__53=( Value == step1_class_6_Value )
Script__54=Value = 7
Script__55=else if
Script__56=( Value == step1_class_8_Value )
Script__57=Value = 8
Script__58=else if
Script__59=( Value == step1_class_9_Value )
Script__60=Value = 9
Script__61=else
Script__62=Value = 100
```