IMPROVEMENT OF LAND COVER CLASSIFICATION WITH THE INTEGRATION OF TOPOGRAPHICAL DATA IN UNEVEN TERRAIN

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

BY

DENİZ GERÇEK

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

DECEMBER 2002

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Tayfur Öztürk
Director

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

Assoc. Prof. Dr. Oğuz Işık
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. Vedat Toprak
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Oğuz Işık

Assoc. Prof. Dr. Vedat Toprak

Assist. Prof. Dr. Mustafa Türker

Assist. Prof. Dr. Zuhal Akyürek

Dr. Lütfi Süzen

ABSTRACT

IMPROVEMENT OF LAND COVER CLASSIFICATION WITH THE INTEGRATION OF TOPOGRAPHICAL DATA IN UNEVEN TERRAIN

Gerçek, Deniz

M.Sc., Department of Geodetic and Geographic Information Technologies

Supervisor: Assoc. Prof. Dr. Vedat Toprak

December 2002, 112 pages

The aim of this study is to develop a framework for the integration of ancillary topographic information into supervised image classification to improve the accuracy of the classification product. Integration of topographic data into classification is basically through modification of training set in order to provide additional sensitivity to topographical characteristics associated with each land cover class in the study area.

Multi-spectral Landsat 7 ETM 30x30 meter bands are the remotely sensed data used in the study. Ancillary topographic data are elevation, slope and aspect derived from 1/25000 scaled topographic map contours.

A five-phase methodological framework was proposed for developing procedures for the integration of topographical data into a standard image classification task. Briefly; first phase is the selection of initial class spectral signatures, second phase is analyzing the information content of class spectral signatures and topographical data for a potential relationship, and quantification of the related topographical data. Third phase is the selection of class topographical signatures from the related topographical data. Fourth phase is redefinition of two training sets where one of which includes spectral information only and the other includes both spectral and topographical information. The last phase is classification. Two products were derived where, first product used bands as input and was trained by spectral information only and the second was the product for which bands and topographical data was used as input and it was trained with both spectral and topographical information.

Method was applied to image and associated ancillary topographical data covering rural lands mainly composed of agricultural practices and rangelands in Ankara.

Method provided an improvement of 10% in overall accuracy for the classification with the integration of topographical data compared to that depended only on spectral data from remotely sensed images.

Keywords: Image Classification, Integration of Ancillary Data, Topographical Data, Training set

ÖZ

TOPOGRAFİK VERİ ENTEGRASYONU İLE ARAZİ ÖRTÜSÜ SINIFLANDIRMA HASSASİYETİNİN ARTIRILMASI

Gerçek, Deniz

Yüksek Lisans, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü

Tez Yöneticisi: Doç. Dr. Vedat Toprak

Aralık 2002, 112 sayfa

Bu çalışmanın amacı, görüntü sınıflandırma sonucunu iyileştirmek üzere yardımcı topoğrafik verinin sınıflandırmaya entegre edilmesini sağlamak için bir yöntem geliştirmektir. Topoğrafik verinin sınıflandırmaya entegre edilmesi temel olarak, eğitim setlerinin arazi örtüsü sınıflarının topoğrafik karakteristiklerine daha fazla hassasiyet gösterecek biçimde yeniden düzenlenmesi yolu ile gerçekleştirilmektedir.

Çalışmada kullanılan uydu görüntüleri, Landsat Thematic Mapper 7 ETM bantlarından oluşmaktadır. Yardımcı topoğrafik veriler ise, 1/25000 ölçekli topoğrafik harita yükseklik konturlarından elde edilen yükseklik, eğim ve bakı verileridir.

Topoğrafik verilerin standart görüntü sınıflandırma işlemine entegre edilmesini sağlamak üzere beş aşamalı bir metodolojik yapı geliştirilmiştir. Özetle; birinci aşama sınıflara ait ilk spektral işaretlerin belirlenmesidir, ikinci aşama bu işaretlere topografik verilerin olası bir ilişkinin tespiti için analiz edilmesi ve ilişkili bulunan topografik verilerin seçilmesidir. Üçüncü aşama ilişkili bulunan topografik veriden sınıf topografik işaretlerinin belirlenmesidir. Dördüncü aşama iki eğitim setinden birinin sadece spektral bilgi içerecek, digerinin hem spektral hem topografik bilgi içerecek şekilde yeniden tanımlanmasıdır. Son aşama ise sınıflandırmadır. Sınıflandırma sonucunda, biri girdi olarak sadece spektral bantları kullanmış ve spektral bilgi ile eğitimlenmiş, diğeri ise girdi olarak bant ve topografik veri kullanmış ve hem spektral hem topografik veri ile eğitimlenmiş iki ürün elde edilmiştir.

Metodoloji, Ankara'nın kuzeyinde, çoğunlukla tarım, mera ve çalılık alanlardan oluşan kırsal bir araziyi kapsayan Landsat TM görüntüleri ve ilgili topoğrafik veriler üzerinde uygulanmıştır.

Metod topografik verinin entegrasyonu ile elde edilen üründe, sadece spektral bilginin kullanımı ile elde edilen ürüne göre doğrulukta %10 oranında iyileşme sağlamıştır.

Anahtar Kelimeler: Eğitim seti, Görüntü Sınıflandırma, Topoğrafik Veri, Yardımcı Veri Entegrasyonu.

To my Family

ACKNOWLEDGMENTS

I would like to express my special thanks to my supervisor Assoc. Prof. Dr. Vedat Toprak, for his guidance, patience and motivation. I am sincerely in debt to him for his continuous support throughout this study.

I would like to thank to Assist. Prof. Dr. Mustafa Türker for his guidance and valuable suggestions and to Assist. Prof. Dr. Şebnem Düzgün for her guidance in statistics.

I would like to thank to Dr. Lütfi Süzen for sharing his knowledge and valuable experiences and his guidance at every stage of this study.

I would like to thank to HAT A.Ş. for providing data.

I would like to thank to Ediz Kırman for his cooperation and acquaintance and Ömer Gözü for his consideration and motivation.

Finally, I would express my thanks to my family for their patience and compassion throughout the study.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1. An Overview of Remote Sensing for Land Cover and Land Use Detection

Information regarding the characteristics and spatial distribution of land cover and land use is critical in monitoring and management of environment. Information on actual land cover and land use serve as basis for various studies of geosciences and furthermore provide substantial background for determining and implementing strategies, policies and principles for planning in local, regional and global scale. However, available data on land cover and land use are often out-of-date, of poor quality or inappropriate for particular applications.

The earth resources data are collected using basically two methods including in situ (field) and/or remote methods (Jensen, 1996). Beginning with the early use of aerial photography, remote sensing has been recognized as a valuable tool for viewing, analyzing, characterizing environment, and making decisions about environment. Today, satellite remote sensing is defined as the use of satellite-borne sensors to observe, measure, and record the electromagnetic radiation reflected or emitted by the Earth and its environment. Rapid development of satellite techniques, the advances in geomatics technologies encouraged by the non-stop development of computer environment have created many advantages for monitoring and handling earth resources. Those advantages include; capability to capture a synoptic view, availability of multi-spectral data providing increased information, capability of repetitive coverage, being global in scale and being not limited by political or geographic boundaries, being contemporary, being fast and practical, giving easy access to the end user (Bruzzone et al., 1997, Campbell, 1996)

In situ methods for collecting data often include point based measurements with direct contact and produce discrete data. However, remote sensing offers a way to avoid the logistical and economic difficulties associated with obtaining continuous in situ measurements of surface features.

Aerial photographs also have long been primary data input for production of topographic maps and various types of ground truth information, but obviously after certain photogrammetric processing yet, they are subjected to considerable amount of geometric distortions. Considering the aforementioned advantages of satellite remote sensing with its relatively less amount of preprocessing compared to that required for aerial photographs, it is increasingly becoming more optimal method of data collection, even the only feasible approach to map land cover and land use, especially on regional to global scales if frequent updating is required.

Remotely sensed imagery provides enormous quantity of data of the earth, but those data are inherently raw and user interpretation for such data is often limited in quality. Thus, automated processing of those images is apparently required to extract particular thematic information.

Image classification is a widely used automated technique to extract information from remotely sensed images. Image classification is the process of converting image data into useful thematic information; it categorizes spectral data into classes with respect to statistical decision rules introduced by the classifier algorithm. The multi-spectral image classification techniques are various and performed using plenty of algorithms.

However, information gathered by the classification of remotely sensed data, based solely on spectral variability is often insufficient in accuracy (Janssen et al., 1990; Bruzzone et al., 1997). Attempt to improve the accuracy of image classification would be to extend the classification procedure with the integration of data and/or information (Westmoreland and Stow, 1992, Bolstad and Lillesand, 1992, Gahegan and Flack, 1996). This data and information, also known as “ancillary data” in the literature, are often composed of map-based thematic data, terrain data and non-spatial data. There have been numerous attempts to increase overall accuracy of classification during the period regarding the use of automated classification systems. Some of the representatives of those approaches are given in *Background Study* chapter.

1.2. Purpose and Scope

This study represents a framework and application to increase accuracy of image classification and yield more reliable land cover thematic information with the use of ancillary topographical data in a standard classification procedure. Method was intended to provide a simple and concise approach to the integration of ancillary topographical data into classification, with a series of straightforward procedures. The method was primarily based on integrating topographical data into supervised classification procedure as a component, in addition to the spectral bands of satellite imagery. The crucial point of integration is the modification of training set so as to take topographical signatures of classes into account, which efficiently yielded an improved training set sensitive to topographic characteristics of features as well.

In image classification applications, it’s important to select an appropriate set of multi-spectral imagery to satisfy the final expectations. Landsat Thematic Mapper (TM) which has far been the most widely used and effective type of earth observing satellite in multi-spectral image classification applications at local and regional scales was the primary source of data for the study.

Surface elevation above mean sea level and slope were the ancillary data being used in the study as they deployed significant relations with land cover data. Aspect, which is also a derivative of elevation data, was excluded from the analyses for being poorly related with land cover types in the study area.

The proposed methodology, in which the aim was to partially prevent misclassifications due to spectral confusion, is implemented in five stages. In the first stage, an initial training set which represent the spatial characteristics of four land cover classes was selected. In the second stage the degree of relation between the four classes and the elevation data (elevation, slope and aspect) were investigated, elevation and slope data after correlation analyses were realized to be related with the land cover classes and entitled to be used as ancillary data in the classification. The third stage was selection of topographic training set which is representing the class topographical signatures. Selection of the initial training set was finalized by this stage. The fourth stage was the modification of training set by taking ancillary data into consideration without changing the spectral training set definitions, which is almost impossible practically by manual editing of the supervisor. Thus, the initial training set is redefined by including all feature elements covered by the data ranges (minimum-maximum) of class signatures. After the redefinition, the fifth stage was classification. Two products are derived through classification. This was done to enable an objective comparison with the classification qualified with ancillary data to measure the effect of topography and classification of only images with maximum likelihood classifier by the standardized training set utilized for the images only. Then classification of images, elevation and slope together with maximum likelihood classifier by the standardized training set utilized for imagery, elevation and slope is performed. The fifth and the last stage is the comparison of the two results with the ground truth information. The two results were trained spectrally the same but in the second classification, training set included additive elevation and slope information yielding a result including effect of terrain.

1.3. Method of the Study

Study is primarily based on office work consisting of collection and evaluation of various data sets. Field work is also performed to understand the feature components of the study area and collect some test points to be used in production of ground truth data. The procedures to implement the study are mainly composed of geometric correction of the data, production of Digital Terrain Model (DTM) and derivatives, training set selection, automated training set selection where, all of the raster attributes were transferred to vector points, every one of which represented a standard raster grid of 30x30 meters to execute the selection procedure, followed by classification, and accuracy assessment.

The information about the nature and the evaluation of these data sets will be given in detail in following chapters.

For handling the processes required to implement the study, TNTmips; map and image processing system of MicroImages was used. The system provides powerful tools for digital image processing and offers a flexible environment for the integration of image and the ancillary data. MapInfo was also used in particular stages. Statistical computations were mainly carried out using Microsoft Excel and SPSS. Correlation analysis was performed by an online biserial correlation calculator, which is a utility provided on Vassar Collage web site.

1.4. Format of the Thesis

This thesis consists of five chapters. The first chapter is Introduction. In chapter 2, theoretical information on image classification and methods improving classification accuracy with the incorporation of ancillary data are represented and previous studies on image classification using ancillary data are introduced. In chapter 3, study area, data and preprocessing operations on data are described. In Chapter 4 the methodology and workflow of the study is represented. Chapter 5 involves the discussion of the results and the last chapter is conclusion and recommendations.

1.5. Study Area

This study represents a framework and application to increase accuracy of image classification. The proposed method was applied to the imagery and the associated data set covering rural areas in northern Ankara.

Study area is located in northern Ankara, in central Anatolia within the boundaries of Ankara municipality and including a part of Kızılcahamam city. It covers a rural area of approximately 66 km2 with dimensions 7.4 x 8.9 km (Figure 1.1).

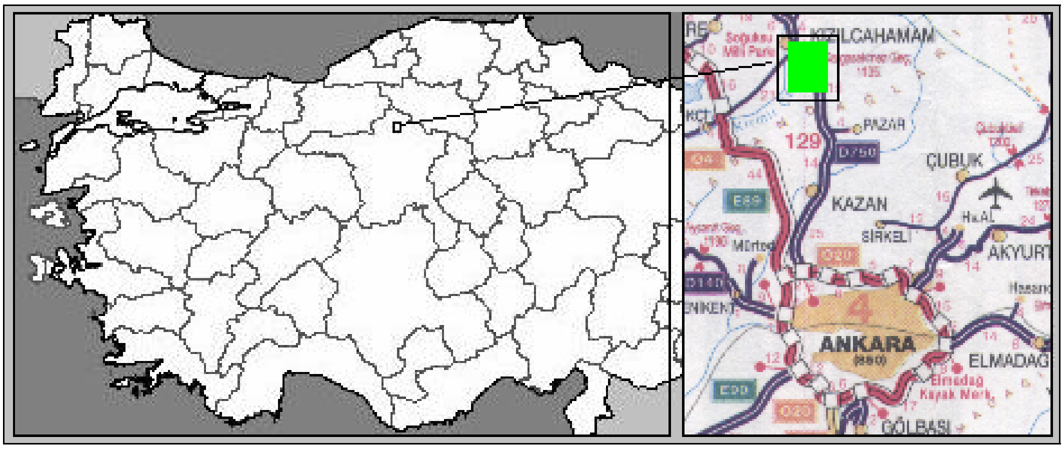


Figure 1.1 Location of the study area

Topography of the study area is uneven. It covers a typical volcanic mountainous terrain of middle Anatolia with dissected stream valleys, besides; there also exists some flat regions. Elevation in the study area ranges from 890 meters to 1440 meters, and slope varies from 0^{o} to 40^{o}. Middle and the southern parts are relatively low in altitude and have gentle slope. Northern and the north-western parts of the study area have higher elevation with moderately or steeply sloping terrain.

The native vegetation of the study area is typically composed of rangelands. Anderson et al. (1976) defined rangelands as natural vegetation types involving a variety of land from densely dominated shrub and brushes to sparsely vegetated herbaceous lands. Rangelands of the study area are composed of common steppe vegetation species in central Anatolian regions where typical continental climate is prevalent (DMİGM, 2002). The native shrubs and brushes of the study area are steppe species of maximum 2-meter height, distributed densely in the terrain. Herbaceous rangelands of the study area are poorly vegetated lands with herbaceous plants of maximum 20-30 cm height. In some areas, herbaceous rangelands are mixed with the native shrubs and brushes of the study area. Moving through the north, particular areas are dominated with trees composed mainly of coniferous and partially of deciduous tree species.

Apart from the natural land cover, particular land use classes are present in the study area. Land use classes in the study area are primarily agricultural, residential, industrial and transportational. Agricultural lands in the study area are mainly located in the mid and the southern parts. Predominant agricultural activity is cereal farming, also limited amount of vegetable farming is practiced nearby the rural settlements and the lower wetlands along Meraçayı River and branching intermittent streams. Residential development consists of several rural settlements and a part of Kızılcahamam city located at the upper edge of the study area. Industrial uses are expanding especially on the Kızılcahamam fringe and along Ankara–Kızılcahamam highway (D750).

The main reason for choosing this area was that the data required for implementing the analyses within the methodological frame were available and somewhat current. The other reason was the variability of the classes as all of the superior classes were typically of continental environment of central Anatolia. Observations of the available data set and the site visits convincingly affirmed the presence of these classes. Other reason was unevenness of the terrain, where topography is significantly varying within the study area. And the final issue to mention is that the study area is located near Ankara and site visit is rather feasible and easy.

CHAPTER 2

BACKGROUND STUDY

In this chapter brief theoretical information about image classification, and approaches on improvement of classification accuracy with the integration of ancillary data or information are provided. Majority of the researches in the literature about the concept of accuracy improvement concentrate on increasing the measure of the accuracy of image classification procedures with the integration of ancillary data.

In the first section of this chapter, image classification, factors influencing classification results and concept of classification accuracy are discussed. In the subsequent section focus is shifted a little deep into the technical issues concerning the basic, conventional image classification. And in the last section, particular approaches of integrating ancillary data into the classification procedure are introduced and previous studies on the subject are provided.

2.1. Image Classification

Spectral signatures of the features are recorded as reflectance values by remote sensing systems at different wavelengths. This introduces the multi-spectral concept to a remote sensing system. Those bands within these different spectral settings include enormous amount of raw data regarding earth surface. A widely used method to convert these data into useful information or into thematic data is image classification. Image classification is an information extraction process that analyses the spectral signatures for classes and assigns each pixel in the image to its appropriate class according to the class signatures (Sabins, 1977).

Image classification methods depending on the spectral reflectance values yield satisfactory classification accuracies when land features are spectrally separable. But on the contrary case, where features of environment are spectrally complex, great difficulties are encountered. Spectral confusion affecting the class signatures is the main reason of misclassifications (Bruzzone et al., 1997).

Factors which lead to spectral confusion basically have two different sources which are nature of environment and irregularities of terrain (Lunetta et al., 1991). Some different land cover types naturally can give similar reflectance values, that a single spectral class may represent more than one information class, is a source of confusion depending on nature of environment and hard to surmount with classification of spectral data only. A single information class may not show the same spectral characteristics due to topographical effect which is defined as variation in radiance from inclined surfaces as a function of the orientation of the surface relative to the sun and sensor position (Burgess et al., 1995).This is an error changing the original reflectance values of features but may be removed partially by radiometric correction.

Under these circumstances, it is evident that image classification may have some uncertainties. Therefore; thematic information extracted from classified imagery can be served as an end product, but in case, provided together with a known degree of uncertainty. Considering the issue, a measure indicating the degree of uncertainty is useful. Uncertainty of a classification product is measured by a procedure widely known as accuracy assessment. Accuracy assessment is simply the comparison of two sources of information; remote sensing derived classification product and reference test information or ground truth data (Jensen, 1996).

2.2. Methods of Classification

Image classification often includes the categorization of spatial features on earth into land cover and land use classes like forest, settlement, water, etc. by means of selected classifier algorithm. Various classification algorithms which basically depend on statistical decision rules are performed for classifying images. Major difference between these classification methods is their emphasis and ability to incorporate information obtained by remote sensing.

The two generalized methods for image classification are supervised and unsupervised classification. Being supervised or unsupervised is the basic point of distinction for image classification where, supervised classification involves user interference, and unsupervised not. Nevertheless, image classification methods can be further categorized by means of being per-pixel or contextual, hard or fuzzy and parametric or non-parametric.

Most commonly used classification algorithms depend on per-pixel decision rules, which evaluates the membership of pixels individually. In contextual classification, considering the spatial context, a pixel with weak observational evidence for being a member of a specific class is classified according to its neighboring pixels' membership tendencies (Deusen, 1995).

Soft or fuzzy methods of classification provide a partial membership for each pixel, in other words, a measure of degree of similarity for every class is assigned to every single pixel where, hard classifiers assign every single one of pixel to a single class according to the highest class membership information (Foody, 1992; Mather, 1999).

Most of the classification methods are parametric which means, they are primarily based on Bayesian decision models and work under the theoretical assumption that the pixels in the image have a probability density function with normal distribution. Non-parametric classifiers do not depend on a particular probabilistic models (Favela and Torres, 1998), thus they are free of restrictions introduced by the assumptions of the probabilistic models.

2.2.1. Unsupervised Classification

Unsupervised classification automatically categorizes the image data into spectral classes by means of an unsupervised classifier such as K means and ISODATA. The classifier algorithm estimates the mean values of the classes and other essential descriptive statistics required for clustering operation, after several iterations. In unsupervised classification algorithm-defined test pixels are utilized instead of user-defined training samples to construct class signatures. This method is inevitable where insufficient or no information of the area covered by the image is available.

2.2.2. Supervised Classification

In supervised classification a priori information of the area covered by the image becomes mandatory. The identity and location of feature classes or cover types are collected beforehand through field study and/or interpretation of aerial photographs or using up-to-date and reliable maps. Existing surface features on the ground are observed and corresponding pixel/pixels representing those features on the image are selected by the analyst and accepted without doubt that they successfully represent the characteristics of the associated class. These samples are called training samples or training sets. The initial procedure is calculation of univariate and multivariate statistics where, univariate statistics are; minimum, maximum, mean, variance, standard deviation and multivariate statistics are; variance-covariance matrix and correlation matrix for each training set. Then the spectral characteristics can be used to train a classification procedure to assign each pixel in the image to one of these classes. Each pixel is evaluated for its degree of being likely to be a member of a specific class and then, assigned to that class.

Supervised classification is more effective in terms of accuracy in mapping substantial classes whose validity depends to a degree on the cognition skills of the image specialist (Bolstad and Lillesand, 1992)

The most widely used supervised methods are minimum distance to mean, parallel piped and maximum likelihood algorithms.

Parallelepiped requires the least information from the supervisor compared to other three methods. For each information class, user identifies the minimum and maximum pixel values for each band. This ranges form parallelepipeds. Then, the algorithm assigns pixels to a class if value of the pixel is within the minimum and maximum ranges of a pre defined class. The method has drawbacks due to two extreme cases that may inherently occur. First one is point in the spectral space representing the pixel may not lie inside any of the regions defined by parallelepipeds. Then the pixel is not assigned to any class. An the second, point may lie inside two or more overlapping parallelepipeds, then decision becomes more complicated that it can not be solved within the capabilities of the classifier (Mather, 1989). Method may be successful in classification of data which is showing high seperability with no overlapped or unidentified regions but is not appropriate for data of natural phenomena which is not often so.

Minimum Distance mean classifier analyzes the training set provided and calculates a mean vector for each information class, described by the class center coordinates in feature space (Jensen, 1996). The Minimum Distance to Mean algorithm then determines the Euclidean distance from each unclassified cell to the mean vector for each prototype class and assigns the pixel to the closest class.

In Maximum Likelihood classification, an unknown pixel is assigned to a class which, the pixel is most likely to be member of. A set of observations based on mean and variance-covariance matrix corresponding to a class is generated, an individual pixel’s probability of being a member of each class is computed, then the pixel is assigned to the class for which the probability value is greatest

(Strahler, 1980). The probability that an unknown point belongs to a particular class depends on the distance from the point to the class center, and also on the variance and covariance of the class which define the size, shape and orientation of the distribution of points in the class (Microimages, 2002).

The probability P(x) that a pixel vector x of p elements (a pattern defined in terms of p bands) is a member of class i is given by the equation of multivariate normal density:

$$P(x) = 2\pi^{-0.5p} \ | S_i |^{-0.5} \ \exp (y' S_i^{-1} y)$$

Where | . | denotes the determinant of the specified matrix,

S_i is the sample variance-covariance matrix for class i, and

y is (x-x_m) where x_m is the multivariate mean of class i

(Mather, 1989)

Maximum likelihood is a more complicated method compared to other two methods since it uses mean and variance-covariance matrix to compute class membership, where parallelepiped classifier use only minimum maximum value, and minimum distance to mean algorithm use only mean of the training set. The superiority of Maximum Likelihood classifier over the other supervised methods such as minimum distance to mean, parallelepiped and etc. is because it takes into account the shape, size and orientation of a cluster (Shrestra and Zinck, 2001).

However, Maximum Likelihood is a parametric model based on the assumption that the data has normal distribution. However this is usually not the case for remotely sensed data.

2.3. Image Classification with the Integration of Ancillary Data

Image classification depending on the spectral reflectance values only, is often limited in content and accuracy. Results derived from the classification of only image data may not satisfy expected accuracy for particular applications. For instance United States Geological Survey (USGS) affirmed that thematic maps extracted from data set including remotely sensed imagery should satisfy minimum level of 85% accuracy (USGS, 2002). Under these circumstances it's fair to state that image classification based on spectral data only may be unfeasible to satisfy desired level of accuracy required for particular applications. At this point solution can be the incorporation of non-image information into image classification. External non-image data so-called ancillary data are any type of spatial or non-spatial data such as elevation, slope, aspect, geology, vegetation, crop yield statistics etc. that may contribute to image classification procedure (Jensen, 1996). These data are usually utilized as additional information for assigning pixels, which are subject to spectral confusions or have some uncertainty of being a member of a particular class.

Various methods were developed to improve accuracy of image classification with the use of ancillary data. These methods were grouped into three by Hutchinson (1982) as (1) use of ancillary data before classification; preclassification scene stratification, (2) use of ancillary data after classification; post classification sorting and (3) use of ancillary data during classification; classifier modification.

Stratification involves segmentation of the image into smaller scenes before classification takes place in order to provide spectrally similar classes to be classified independently. Post-classification sorting is the use of ancillary data after classification based on the problem that a single class of objects may be assigned to more than one classes due to the fact that a class can show different spectral characteristics. These problematic cases are treated based on decision rules to assign problem pixels into appropriate class using ancillary data.

Integration of ancillary data during classification has followed two approaches; inserting ancillary data as an additional channel or modifying prior probabilities derived from data statistics. (Hutchinson, 1982; Harris and Ventura, 1995; Mesev, 1998)

The first and the most obvious method is aimed to increase the number of attributes or channels of information used in the classification. For instance n bands plus one or more ancillary data layers can together be involved as input into classification. This technique is also called Logical Channel. Strahler et al. (1978) practiced simple additional channel technique for assessing urban change. He involved a land use map derived from a topographical map with SPOT panchromatic image into multispectral image classification procedure. However, simple addition of non-spectral data into classification without making any modifications on the training statistics may add little to classification.

The second is classifier modification, which involves changing a priori probabilities according to areal composition of the expected product based on image statistics, ancillary data or a known relationship between classes and ancillary data. In conventional classification, prior probabilities are assumed to be equal for all classes. Classifier can be modified by changing the prior probabilities before classification. Prior probability is the probability of occurrence of classes which are based on separate, independent information concerning the area to be classified (Strahler, 1980). When used in classification procedure, the probabilities weight the classes according to their expected distribution in the data set by shifting decision space boundaries to produce larger volumes in measurement space for classes that are expected to be large or smaller volumes for classes expected to be small.

The effect of ancillary information in improvement of image classification accuracy have long been a concept of research in remote sensing literature and many researchers made use of ancillary data to improve accuracy of image classification.

Harris and Ventura (1995) developed a post-classification method incorporating ancillary spatial data to improve the accuracy of a land use classification derived from multi spectral imagery for pollution modeling in an urban area. Landsat TM images were classified with maximum likelihood method and the product was modified with zoning and housing data. This provided an increase both in accuracy and content or namely, number of information classes of the classification.

Mesev (1998) demonstrated a supervised classification method integrating population census data with urban land cover from remotely sensed data. Census data which is composed of points is converted into a continuous surface raster is then used to modify maximum likelihood classification through class a priori probabilities and in terms of post-classification sorting to resolve misclassified pixels.

Vogelmann et al. (1998) aimed at generating a large region land cover ancillary data. Procedure involved unsupervised classification of Landsat TM, labeling of classes and further development of decision rules for splitting confused classes into the appropriate land cover category using one to several data sets some of which were elevation, slope, aspect, population, density and city lights.

Maselli et al. (1995) have extended the methodology based on the inclusion of prior probabilities derived from the frequency histograms of the training sets in to maximum likelihood (Maselli et al., 1992) and proposed a new method to integrate ancillary data layers a priori probabilities into classification process instead. This led to significant increase in accuracy compared to the method previously proposed by the authors Maselli et al. (1992). Since they integrated additional information derived from the images once more into the classification procedure.

A study by Eiumnoh and Shrestha (1997) attempted to explore the effect of DTM in accuracy of image classification by combining it as a component band with Landsat TM band combinations for land cover classification. The maximum likelihood supervised classifier was applied to digital terrain model and different band combinations of Landsat TM imagery. It was concluded that DTM as one type of ancillary data can improve the classification result.

Irvin et al. (1997) performed an automated classification on Digital Terrain Model and its derivatives; slope, aspect, profile curvature and plan curvature to automate the production of landform information. The aim was to produce a Soil map and landform information was one of the attributes used in soil mapping. First they have investigated whether units created from DTM and derivatives correlate with observed soil properties, they have hypothetically assumed that there is correlation between them and classified DTM and derivatives with standard ISODATA and fuzzy classification methods to be used in soil mapping. The classification product of land forms was quite similar to the ones which were manually delineated.

CHAPTER 3

DATA AND PREPROCESSING

3.1. Data

A subscene of Landsat 7 ETM imagery of May 2000 extracted from the scene labeled 177/32 is the primary data source for the analyses (Figure 3.1). The subscene covering the study area was composed of 295 rows and 248 columns. Bands 1, 2, 3, 4, 5 and 7 with 30x30 meters spatial resolution were selected for the analyses. Band 6 and Band 8, which are also present in the Landsat TM 7 data set were excluded from the analyses since, their information content and spatial resolution was quite different than the selected data set. Band 6 is presenting the thermal radiation emitted from the Earth surface with a spatial resolution of 60x60 meters. Band 8 is a panchromatic imagery with 15x15 meters resolution. Although Band 8 is excluded from the analyses, it was used as a complementary data in the geometrical correction of 30x30 meter Landsat bands.

The ancillary data source for the study is the topographical data. Topographical data were derived from elevation contours of 1/25000-scaled (Figure 3.2) topographical map obtained from General Command of Mapping (GCM).

These two sets of data are described as the major data and were used in the analyses. Other data used in the study, so called minor data were set apart from the analyses and were only used for gathering ground truth information. Those data are two IRS panchromatic images with 5x5 meters resolution from 1c sensor (Appendix A.1) and from 1d sensor (Appendix A.2), 1/25000 forest map

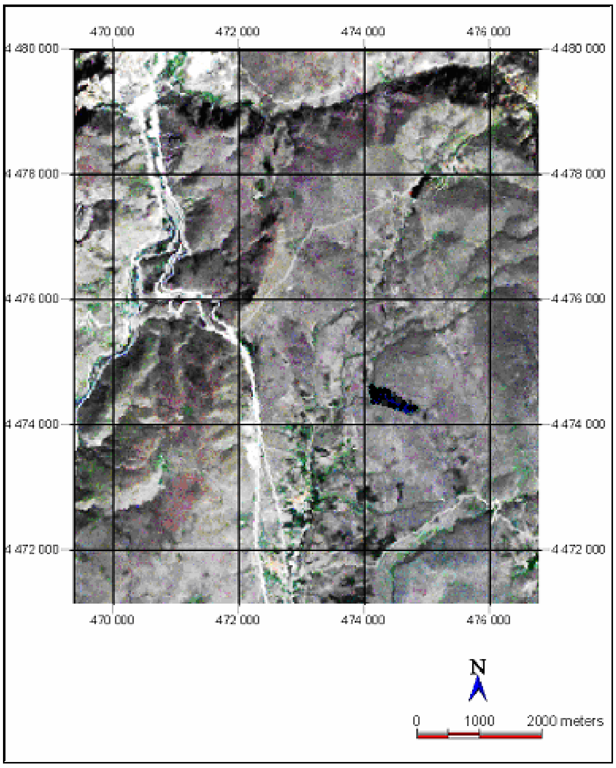


Figure 3.1: Landsat True Color Composite (321) of the study area

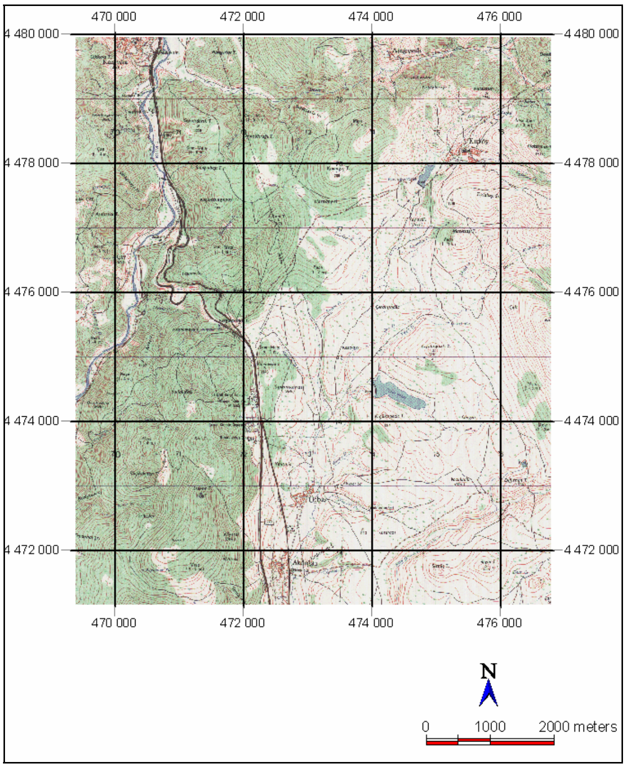


Figure 3.2: 1/25000-scaled topographical map of the study area

including two regions Yıldırım (Appendix A.3) and Kızılcahamam (Appendix A4), obtained from General Directorate of Forest (GDF) a digital land use and land cover map which was obtained from a pilot project of General Directorate of Rural Affairs (GDRA) (Appendix A5), an aerial photograph stereo pair and field observation data. Particular thematic information within the 1/25000-scaled topographical map may also be mentioned since it was used to extract some thematic information for the ground truth data. It was accomplished as a major data source actually.

Detailed information on data used in analyses and ground truth information are given in Table 3.1 and Table 3.2.

Table 3.1 Major Data; Data used in the analyses

Primary - Remotely Sensed Data			Ancillary – Topographical Data		
Data	Acquision Date	Source	Data	Date	Source
Landsat TM 7	10.05.2000	GDRA	1/25000 scaled topographical map	1985	GCM

Table 3.2 Minor Data; Data used in gathering ground truth information

Data	Date	Source
IRS panchromatic (1C)	16.08.2000	GDRA
IRS panchromatic (1D)	05.08.2000	GDRA
1/25000 forest map	1994	GDF
1/25000 topographical map	1985	GCM
Digital Land use – land cover map	2001	GDRA
1/35000 scaled aerial photograph	1983	GCM

3.2. Preprocessing

Pre-processing operations, sometimes referred to as image restoration or correction are intended to correct radiometric and geometric distortions that degrade the quality of the remotely sensed data. Image correction produces an image, geometrically close to ground truth and radiometrically close to the radiance characteristics of the features (Jensen, 1996, Campbell, 1996). Preprocessing is a valuable and necessary process prior to particular analyses or processing on the data set. However, it must be taken into consideration that preprocessing operations change the original pixel values and this may introduce other errors as well.

3.2.1. Radiometric Correction

Landsat TM subscene covering the study area is free of internal distortions created by the sensor itself. These errors are predictable and constant for entire scene and generally corrected before offered. Remote sensor data may be acquired properly with a system functioning perfectly and may be free of internal distortions, but error may creep into the data acquision process due to platform perturbations and environmental factors, these errors are called external errors and require radiometric corrections (Jensen, 1996). The two most important sources of environmental error are atmospheric defects caused by atmospheric scattering and topographic defects due to slope and aspect especially in the mountainous terrain affecting the original pixel brightness values. (Franklin and Giles, 1995)

In particular, applications involving image classification, radiometric correction due to environmental error is considered to be unnecessary, since it is regarded to have little effect on the product’s accuracy (Song et al., 2001; Fraser et al., 1977). There is no doubt that the image covering the area of interest involve some environmental influences, nevertheless for this study removal of probable atmospheric and topographic defects were regarded unnecessary after the following evaluations.

Atmospheric effects depending on the existence of aerosols and particles are assumed to make homogenous affects for small regions of interest on the Earth. Atmospheric effects influence all of the bands, and make an identical change in the pixel values of each individual band. Considering the issue for classification, this actually makes a relative change in class spectral signatures and whatever the magnitude of change, it is almost the same for all class spectral signatures. As a consequence of this classification product from the corrected bands would not cause a reasonable amount of improvement compared to uncorrected bands.

Topographic error is a kind of error affecting all of the bands, but making partial changes in the pixel values of each individual band especially corresponding to places with steeply slope due to orientation of the surface relative to the sun (Lunetta et al., 1991). Considering the issue for methodological frame of the study, the improvement that may be obtained by topographical correction is negligible. What if the input bands are corrected topographically, they will be input for both of the products derived to be compared after classification, and improvement will affect both of them in the same magnitude. That would add no to the comparison of the two products.

3.2.2. Geometric Correction

Remotely sensed data contain both systematic and unsystematic geometrical errors. Systematic or sensor specific errors due to scan skew, mirror-scan velocity, panoramic distortion, platform velocity, earth rotation and perspective are predictable and corrected by using some mathematical formulations based on the knowledge of sensor distortion (Jensen, 1996). These errors are often corrected in the ground station before offered to the end user. The unsystematic errors due to alterations in altitude and attitude however, cannot be corrected without ground control. Ground control is provided by assigning geographic coordinates via Ground Control Points (GCPs) to the remotely sensed data (Kardoulas et al., 1999)

This procedure is followed by image rectification. Image rectification is described as geometrically rectifying the image to its correct position and it involves basically two operations that are; spatial interpolation (geometric transformation) and intensity interpolation (resampling) (Jensen, 1996). Once several well-distributed GCP pairs have been identified, the coordinate information is evaluated by the image processing system to determine the proper transformation equations to apply to the original (row and column) image coordinates to map them into their new ground coordinates (CCRS, 2002).

Spatial interpolation involves the geometrical transformation of raw image into geometrically correct image or in other words relocating every pixel in the original image to its proper location in the rectified image (Lillesand and Kiefer, 1994). Mathematical transformation equations are applied to rectify images, those transformation equations use coefficients determined by the relation between image coordinates and geometric coordinates; so-called Root Mean Square Error (RMSE).

RMSE is the root of the mean of the square of the errors and defined as;

$$\mathrm{RMSE} = \sqrt{\frac{\sum r_i^2}{n}}$$

where;

ri = residual for each point

n = number of test points

Intensity interpolation involves extraction of a pixels value from the image and relocation of this value to its appropriate location in the rectified output image, several methods of image interpolation or resampling are; Nearest Neighbor, Bilinear Interpolation and Cubic Convolution (Jensen, 1996).

Geometric correction is essential in the point that it makes the data usable in conjunction with other spatial data, such as other images, maps, or GIS data. For this study, where remotely sensed imagery was proposed to be integrated with other sources of data, the whole data set must be in precise spatial correspondence.

3.2.2.1. Geometric Correction of Major Data

For the geometric correction of the image data, 1/25000-scaled topographic map was available to be used as reference for performing image-to-map registration. But, prior to this procedure the topographic map should be registered to serve as a reference map.

1/25000 scaled topographical map was scanned from an original and unfolded colored sheet. It was imported to TNTmips environment and was registered by means of 8 Ground Control Points read from the map sheet. The geometric correction procedure involved identifying the image coordinates (row, column) of UTM grid's intersection points physically existing on the map sheet, and matching them to their true positions in ground coordinates. UTM projection, accepted widely as a national projection system was used where, zone is 36, which is the appropriate zone covering the study area, and datum is European 1950 Mean.

The total RMSE calculated was 1.00 meters with maximum residual of 1.14 meters (Appendix B.1), which was quite satisfactory since the accuracy of the original 1/25000 topographical map sheets were affirmed as 5 meters in the horizontal plane and 2.5 meters in vertical plane by General Command of Mapping (HGK, 2002). The geometric correction of the data was implemented by 1st order transformation and nearest neighbor resampling.

In consequence, the topographical map became ready for use as reference for image-to-map geometric correction of the Landsat TM 7 imagery. Image-to-map correction involves matching clearly identified features such as road junctions in both map and the image so as to get the coordinate information for the image pixel from the map that is accepted geometrically correct. However finding corresponding features on the images to match with the map is a quite difficult task if the image spatial resolution is coarse. Providentially, Landsat TM 7 unlike the previous Landsat series includes a panchromatic image with 15x15 meters spatial resolution. And it is reasonable to register this image first and then convert the GCP information into a form that the other bands with 30 meters resolution can use.

Landsat Panchromatic image with 15 meters resolution was registered image-to-map. Total number of GCPs used was 22. The total RMS error was 12.45 meters with the maximum error of 18.88 meters (Appendix B.2), slightly exceeding one pixel size. Since Images with total RMSE less than a pixel size are accepted of reasonable accuracy (Fonseca and Manjunath, 1996), these GCPs were admissible.

The GCPs collected for Landsat panchromatic were saved into a text file. This file was then imported to Microsoft Excel. The raster coordinate values in columns with the titles line and column of Landsat Panchromatic GCPs were divided by 2 in order to adjust raster coordinates of GCPs to TM bands with 30 meters resolution. This data is then transferred to 30-meters resolution Landsat images with “read GCPs from text” option. See column and line numbers in B2 and B3 in Appendix B. By this way, 22 GCPs were practically located to their appropriate positions in Landsat TM1 band with 30x30 meters resolution. Figure 3.3 representatively illustrates transfer of a GCP from 15x15 meters Landsat Pan to Landsat TM1 with 30x30 meters resolution

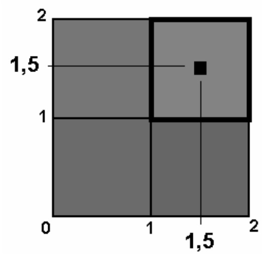
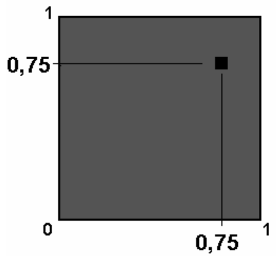
GCP for TM panchromatic 15x15 m	**GCP for TM1 30x30 m**
	

Figure 3.3: Raster coordinates of the same GCP for Landsat Panchromatic and Landsat TM1 band

The residual values are again almost the same for Landsat TM1 band where; maximum RMS error was computed 18.86 meters. Total RMS error is 12.47 meters (Appendix B.3). Ground Control information is then saved and transferred to remaining 30-meter resolution bands by simple copy-paste procedure offered by the system. Geometric correction procedure was followed by 2nd order transformation. Higher polynomial transformations are appropriate when the image has considerable distortions however; it is recommended that very high degree polynomials should be carefully used since they introduce local aberrations (CORINE, 1993). 2nd order transformation reduces the RMS error compared to the affine transformation (Novak, 1992; Fonseca and Manjunath, 1996) and is the most widely used in image geometric transformation. Nearest neighbor resampling method was used to implement geometric correction procedure applied to six bands (1, 2, 3, 4, 5 and 7).

Twelve check points first four of which were lying on GCP locations and remaining eight of which were randomly selected were used for checking the accuracy of the image after geometric correction (Table 3.3). Test was simply a comparison of the coordinates of the checkpoints on the image and on the topographical map. The residual values for the first four points lying on or very near to GCP locations were naturally smaller, since those locations were the inputs of the geometric transformation. Other eight check points deployed relatively higher residuals. Anyway, errors measured for twelve check points were of acceptable quantity since, they were under a pixel size.

Table 3.3: Landsat 1, 2, 3, 4, 5, 7 bands geometric error check

#		Measured Coordinates	Actual Coordinates	x-y residual	Residual (meter)
1	N	4475996.39	4475988.29	8.10	8.33
	E	471634.94	471636.92	1.98	
2	N	4474604.57	4474610.72	6.15	10.95
	E	474041.71	474032.65	9.06	
3	N	4472464.9	4472456.93	7.97	9.51
	E	476019.96	476014.77	5.19	
4	N	44779180.35	44779184.35	-4.00	6.40
	E	472568.69	472563.69	5.00	
5	N	4476198.05	4476184.76	13.29	15.28
	E	472799.04	472806.6	7.56	
6	N	4473078.79	4473065.77	13.02	15.59
	E	472859.28	472867.87	8.59	
7	N	4472895.02	4472885.78	9.24	14.41
	E	475494.28	475505.35	11.07	
8	N	4474351.12	4474357.22	-6.10	7.18
	E	472182.7	472186.5	-3.80	
9	N	4473872.25	4473868.1	4.15	7.21
	E	472337.36	472331.46	5.90	
10	N	4475441.54	4475431.11	10.43	11.60
	E	471841.83	471836.74	5.09	
11	N	4475858.69	4475849.65	9.04	10.33
	E	470925.46	470930.46	-5.00	

12	N	4476758.87	4476744.87	14.00	14.14
	E	471196.29	471198.29	-2.00	

3.2.2.2. Geometric Correction of the Minor Data

Minor data are used in the accuracy assessment of the final products. Those data are; IRS panchromatic image, forest map, aerial photograph pair, digital land use and land cover map and field observation respectively. Accuracy assessment, which will be mentioned in section 4.4, is a procedure of testing the accuracy of the product with ground truth information. Ground truth information for the study was extracted from these four data sets and field observations data. Therefore, matching those data geometrically with respect to each other and to the primary data was of great importance for this study and it is possible with an accurate registration operation.

Map-to Image geometric correction was applied to IRS panchromatic image with 5 meters resolution. The study area coincides with two IRS scenes one of which is B1E15A6D from 1c sensor and the other is, D1E15A6D from 1d sensor. Both of those images were geometrically registered via image-to-map method, taking 1/25000 topographical map as reference. The clearly identified matching points found both in the image and the topographical map were matched. Compared to registration of Landsat TM imagery, procedure was easier since the resolution of IRS panchromatic image was of higher quality. Total number of GCPs used for registering was 33 for B1E15A6D, and 40 for D1E15A6D.

Total RMSE was calculated 7.11 meters for B1E15A6D, which exceeded a pixel size. Hence; four GCPs with the greatest amount of individual error were removed from the GCPs set. The recalculated value of total RMSE was reduced to 6.89 meters but is still over a pixel size Therefore, to lower the RMSE, 2nd order polynomial transformation method was used to transform the image. This reduced the total RMSE for remaining 29 GCPs to 5.55 meters (Appendix B.4). The image is then resampled by nearest neighbor method.

Total RMSE calculated for D1E15A6D; 5.72 meters also exceeded a pixel size.

Six GCPs with the greatest amount of individual error were removed from the GCPs set. The recalculated value of total RMSE with 34 GCPs for 2nd order polynomial transformation was reduced to 4.85 (Appendix B.5). The image is transformed using 2nd order polynomial interpolation and nearest neighbor resampling method.

Another data used as ground truth was 1/25000-scaled Forest Map produced by General Directorate of Forest. The study area coincided with two maps one of which is labeled “Kızılcahamam 1” and the other; “Yıldırım 2” regions. Those maps had several tick points indicating the associated coordinates, however the coordinate values were hardly recognized since the maps were worn and wrinkled. Therefore, forest maps were registered with the use of matching points such as river joints present both in 1/25000-scaled topographical maps and the forest maps.

16 GCPs were used for registering Kızılcahamam 1 with a total RMSE of 13,24 meters and 11 GCPs were used for registering Yıldırım 2 with a total RMSE of 9,94 meters. These errors appear to be high, however there is no available information about the geometric accuracy of these data, and taking the physical distortion of these old sheets into account, the high RMS error was disregarded and GCPs were accepted. The geometric correction was implemented by 2nd order transformation and nearest neighbor resampling of the data.

1/35000-scaled Aerial Photograph pair consisted of consequent left and right scenes covering more than a half of the area. Those two aerial photographs were not scanned but were utilized as hard copy. They were examined with the help of a Stereoscope to obtain the 3D visualization of the study area. This provided better visualization of the area and gave valuable information about features existed within the study area.

The remaining digital land use and land cover data has already been processed and was ready to use as a reference layer.

Field observation data consisted of two data sets one of which was the point observations gathered for the pilot project of General Directorate of Rural Affairs "Digital land use and land cover map" in August 2001 and other of which was the field observation for this study held in August 2002.

3.2.3. Production of Ancillary Topographical Data

Ancillary data produced to be used in the analyses involve Digital Terrain Model (DTM), and DTM driven slope and aspect data.

3.2.3.1. Production of Digital Terrain Model (DTM)

A digital terrain model (DTM) is a digital representation of a portion of the earth's surface in a grid form where each grid point represents a ground elevation value (Ardiansyah and Yokoyama, 2002). Elevation can be presented as contour lines, elevation points or Triangulated Irregular Network (TIN) in a GIS environment. DTM differentiates in being a continuous surface raster modeled from discrete elevation data. Modeling enables presentation of the spatial variability of elevation in every part of the data. This makes DTM a valuable data entity especially when processed together with the image data. A digital terrain model also provides additional derivatives. Attributes derived from a DTM can be grouped into two as primary and secondary where primary attributes are directly derived slope, aspect, and secondary or compound attributes are derivatives from slope, aspect or elevation itself. (Gallant and Wilson, 1996). Within the study in addition to DTM, slope and aspect data derived from the DTM were also practiced.

There are several methods for generating a DTM. These are direct photogrammetric measurements, interpolation from elevation points or contours and more recently, interferrometry from SAR satellite and laser scanner. (Mizukoshi and Ania, 2002). However, digital terrain model generation from contour lines continue to be used widely, because in most countries, this data covers the whole area in different scales, thus presenting cheap data source. (Ardiansyah and Yokoyama, 2002)

Elevation contours from 1/25000 topographical data available for the whole country were used to generate digital terrain model for the study area. Surface elevation in a 1/25000 topographical map is presented by contour lines in red, which are at ten meters interval in general but can go down to 0.5 meter interval in flat areas and with altitude points at the hilltops. The contour lines of 1/25000 topographical maps can be either supplied from the General Command of Mapping in return of some price or digitized by oneself. Digitization procedure is performed in two ways one of which is manual digitization where the user traces each line and other one is automated where image processing system automatically performs the digitization. For this study manual digitization, which is time consuming and tedious was not preferred, instead automated digitization was performed, obviously followed by some automated and manual editing. Automated digitization is offered by the software together with raster-to-vector conversion utilities. Automated digitization process generated a new vector product composed of contour lines together with unwanted dangling lines, bubbles, sliver polygons etc. so called topological errors. Majority of these errors were removed by topological error filters and the remaining of the errors were removed by manual editing. Finally elevation values read from the underlying topographical map were entered into database table and each line is assigned with the appropriate elevation record. And the digital contour map generation was finalized (Figure 3.4)

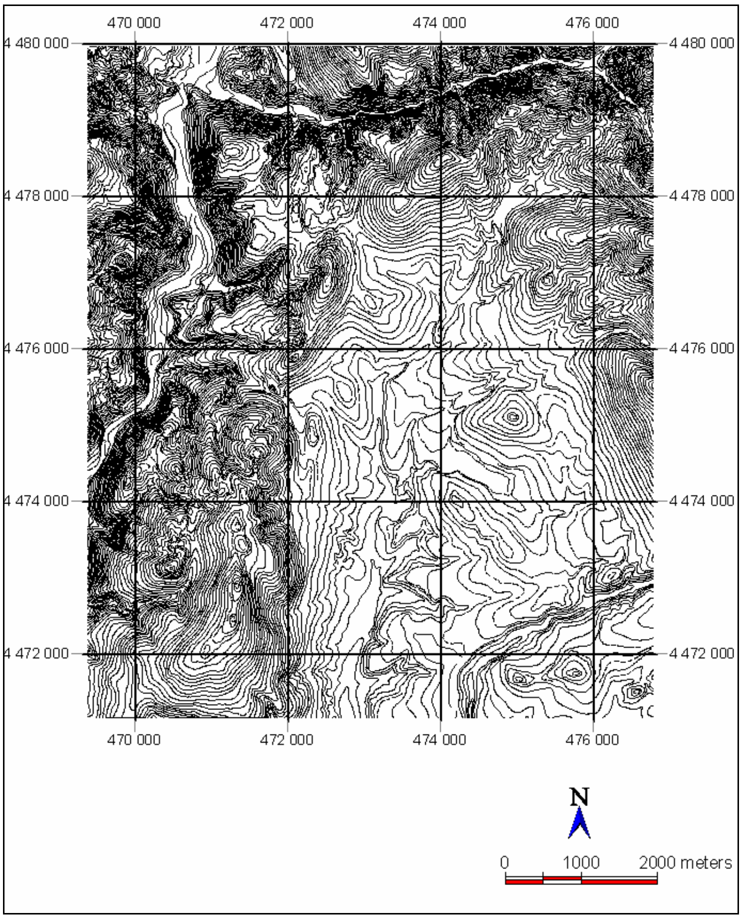


Figure 3.4: Digital contour map of the study area

DTM production depends on interpolation of the known elevation data. Interpolation is one of the basic estimation methods in order to describe the spatial variability of the data, for the data lacking locations, from locations of known data. DTM interpolation is a kind of function for determining elevation of unknown points using a set of proper known data. Selecting a set of appropriate neighboring reference data points is one of the key steps for DTM interpolation. The selected reference points are used for estimating the value of elevation at any location in the given area.

A 30 meter DTM was produced using Kriging method via surface modeling utility provided by the software. Kriging is a sophisticated method of determining the best estimate for each point in a target matrix, based on statistical principles. Kriging works best with datasets that have both regions of densely scattered data and regions of lightly scattered data (RSI, 2002). Method requires point data as input (Microimages, 2002). In order to convert the line data at hand into points, vertices and nodes from contour lines were converted into points by "poly2pnt" utility of MapInfo software and imported back to the TNTmips environment.

Kriging method is appropriate to model elevation data, since the elevation points converted from contours are densely distributed along the contour lines and almost no elevation points exist except for the hill tops and some special points in the data. The parameters for Kriging were selected through the recommendations in the reference manual of TNTmips and a sequence of trial and error procedure. A variogram model must be used for running the Kriging method, because the variogram plot of the discrete control points serves as a basis for determining the appropriate model to select. The Linear model is adequate for general use. For drifting order linear method is again used, because it is assumed that there is a degree of drift in the input and linear model uses a first-order polynomial equation to model the drift. Initial sill and nugget values were not adjusted since adjusting those values was observed to yield unsatisfying results with the elevation data. Simple search type for points were used, search parameters were set as 80 points per sector and 16 minimum total points after a procedure of trial and error. The parameters used in the Kriging interpolation are given in the (Appendix C).

Digital Terrain Model generated from elevation contours using Kriging method is presented with Figure 3.5.

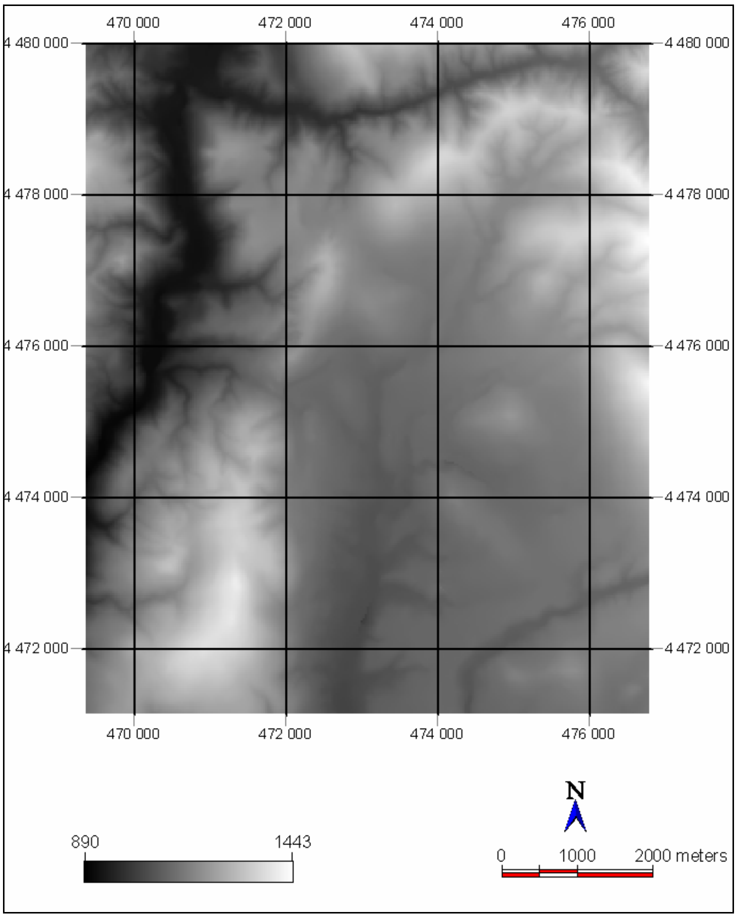


Figure 3.5: Digital Terrain Model of the study area

The elevation ranges from 896 to 1443 in the study area with a maximum concentration at 1115 meters (Figure 3.6).

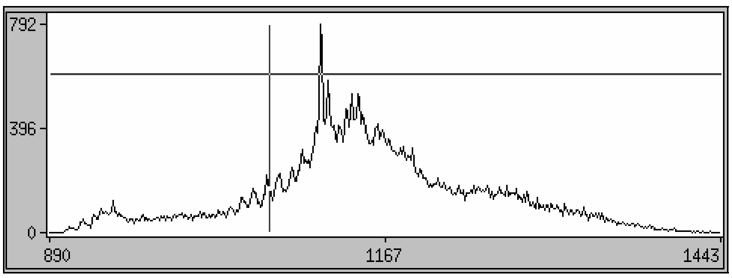


Figure 3.6: Elevation histogram of the study area

Quality of the digital representation of the terrain is important for the study in the way that the topographical parameters will numerically be compared and analyzed with those of images. Once a high-quality DTM has been generated to meet the requirements of the study, the use of these data in analysis will result in valid conclusions.

In order to test the accuracy of the produced DTM, some principles adopted by USGS were used. According to USGS National Mapping Program Technical Instructions, a representative sampling of test points is used to verify the accuracy of any category of DTM. A minimum number of 28 test points per DTM, 20 of which are interior and 8 of which are from the edges is required (USGS, 2002a).

The vertical root-mean-square error (RMSE) statistic is used to describe the vertical accuracy of a DTM, since the only measurable or perceivable errors in the DTM exist as vertical errors that may be partially attributable to horizontal error inherent in the source data. (USGS, 2002b)
Vertical RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum (z_i - z_t)^2}{n}}$$

where;

Zi = interpolated DTM elevation of a test point

Zt = true elevation of a test point

n = number of test points

Accuracy is computed by a comparison of interpolated elevations in the DTM with corresponding measured elevations from the topographical map. A measured elevation mentioned is the average of elevation values corresponding to a 30 meter-DTM pixel in the topographical map.

The test points collected from 28 points on the DTM were then compared with the elevations precisely measured from the topographical maps. The total vertical RMSE was found 1.72 meters, which is quite admissible within the accuracy requirements of the study. The elevation values both from the model and the original data and the vertical errors are listed in Table 3.4.

Table 3.4 Vertical RMSE in Digital Terrain Model (DTM) of the study area

# check point	elevation from DTM	elevation measured from topographic map	residual
1	1044	1045	1.00
2	1130	1130	0.00
3	1311	1314	3.00
4	1166	1165	1.00
5	1205	1203	2.00
6	1070	1069	1.00
7	1152	1151	1.00
8	1183	1183	0.00
9	1390	1390	0.00
10	1180	1179	1.00
11	1219	1220	1.00
12	1036	1032	4.00
13	1222	1225	3.00
14	1140	1141	1.00
15	1409	1407	2.00
16	1239	1239	0.00
17	1214	1214	0.00
18	1266	1266	0.00
19	1044	1048	4.00
20	1138	1138	0.00
21	1115	1115	0.00
22	1099	1099	0.00
RMSE			**1.72**

3.2.3.2. Production of Slope and Aspect

Slope is the measure of rate of change in elevation or in other words; measure of the steepness of an area on the Earth's surface. Slope is the first derivative of a digital terrain model and is often calculated by means of a neighborhood operation applied on a DTM. The slope value assigned to each cell reflects the overall slope based on the relationship between that cell and its neighbors. Slope for the study area was computed by 3x3 window, which uses all eight cells that surround each raster cell to determine the slope value for that cell. Slope value was calculated in degrees where the minimum value can be zero and the maximum possible value can be 90 (Figure 3.7). The slope value ranges between 0 and 48 in the area however, slope greater than 40 compose only a slight portion (Figure 3.8)

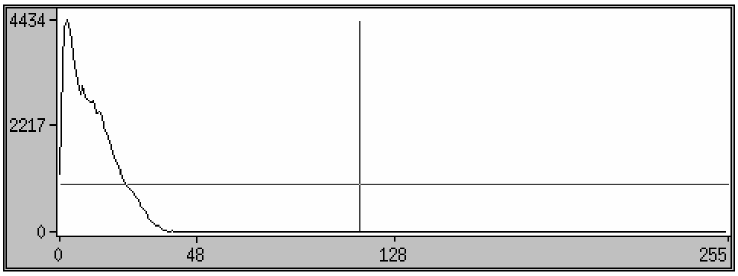


Figure 3.8: Slope histogram of the study area

An Aspect map is also derived from elevation values in a DTM. The Aspect value assigned to each cell in an Aspect map tells us the direction of slope (north, south, etc.) to which that cell is oriented. Just as slope is a measure of the rate of change of a DTM, aspect is the direction of change. While in simple terms this means we can use an aspect map to determine the direction faced by any part of the Earth's surface, aspect maps can be used to determine the direction of change of any phenomena. Aspect output cell values for the study area describe the direction that the slope inclines relative to north. The output was created in degrees and in the range zero to 360 (Figure 3.9). Zero indicates that the surface slopes north, 90 = east, 180 = south, 270 = west, and -1 indicates no slope (flat). Distribution of the aspect values for the study area is given in Figure 3.10.

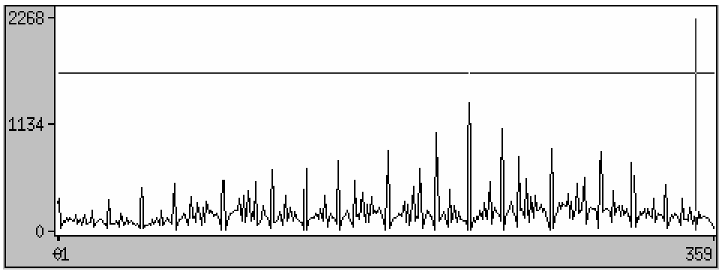


Figure 3.10: Aspect histogram of the study area

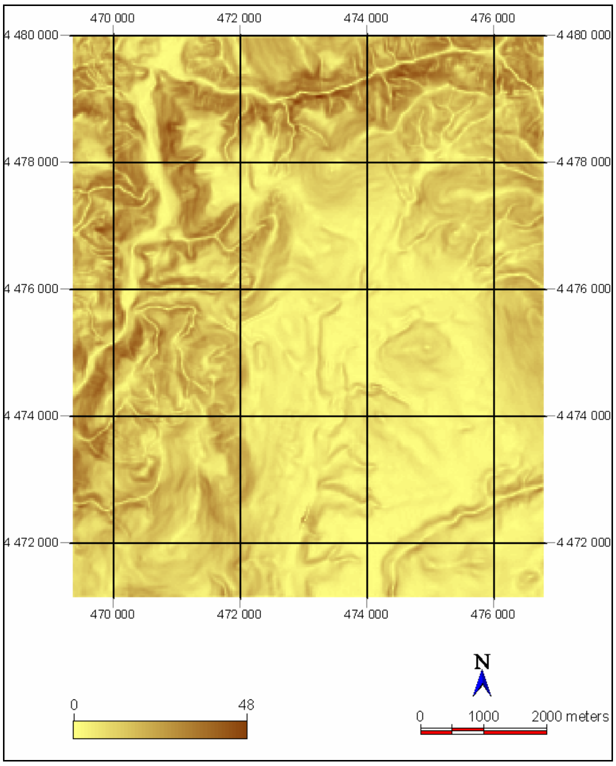


Figure 3.7: Slope map of the study area

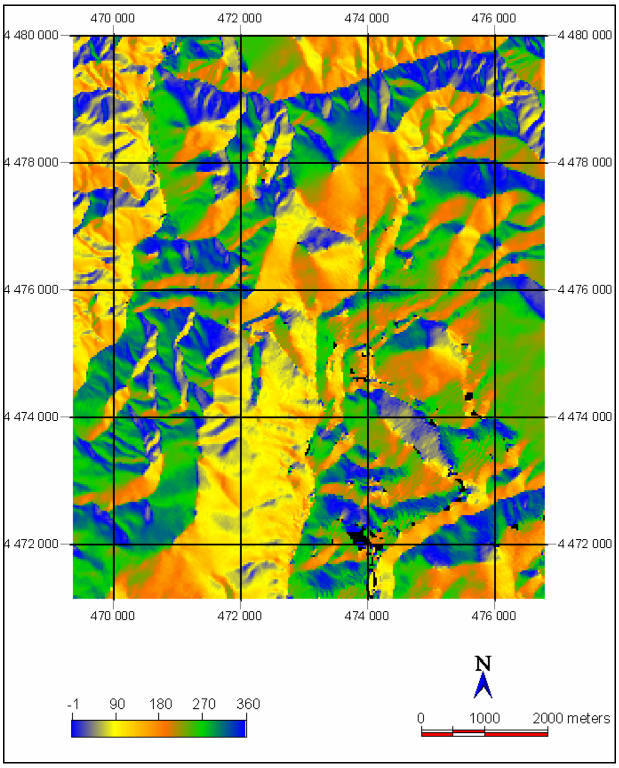


Figure 3.9: Aspect map of the atudy area

3.2.4. Production of Ground Truth Data

Production of the ground truth data is aimed at generating a reliable reference to test classification outputs in the accuracy assessment phase of the study. Accuracy assessment is the indispensable part of the classification task subsequent to the implementation of analyses associated with classification. In order to test the accuracy of the classification product, a reference data which is accepted to be a perfect representation of the actual phenomena is required and that is called ground truth data. For this study, ground truth information is composed of minor data and are IRS panchromatic image, aerial photo, Forest map which includes accurate delineations of land cover from 1/15000 aerial photographs, a digital land cover-land use map previously produced by General Directorate of Rural Affairs, and field observations from the study area.

Ground truth for the study is gathered by interpreting all of this information in other words; a kind of synthesis task is carried on for combining information from all of these data. Production of this data basically depends on visual interpretation of fine resolution imagery and additional thematic data. Some point observations from two field studies one of which was the study of General Directorate of Rural Affairs held in 2002 for digital land cover and land use mapping and other for this study specifically, were also utilized in producing the ground truth information. The product is a thematic map including principle classes in the study area (Figure 3.11) and after a sampling operation to gather test points it was used for testing the accuracy of the end products in Chapter 4, section 4.4.

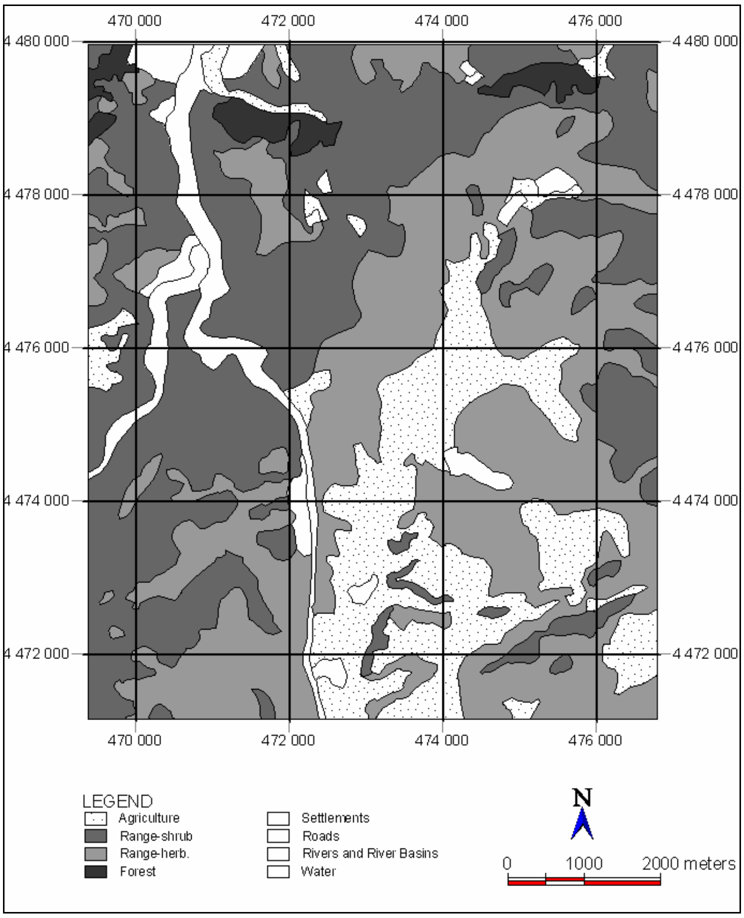


Figure 3.11: Ground truth information for the study area

CHAPTER 4

METHOD AND THE ANALYSES

4.1. Methodology and Framework

This chapter explains the methodology of the study and gives the results of analysis obtained at different phases. A simplified flowchart of the methodology is given in Figure 4.1. The methodology involves four major steps: 1) Preprocessing, 2) Basic Definitions, 3) Analysis and, 4) Accuracy Assessment.

Details of “Preprocessing” are explained in Chapter 3, which is related to the preparation of major and minor data. The data were initially processed and following input layers were produced:

1. Topographic map (1/25.000 scale) was geometrically corrected
2. Landsat TM 7 bands (1, 2, 3, 4, 5 and 7) were geometrically corrected
3. Elevation contours from topographical map were digitized.
4. Digital Terrain Model (DTM) was generated, from DTM, Aspect and Slope was derived.
5. Some of Minor data (IRS Panchromatic image, 1/25.000 scaled forest map) were geometrically corrected.

In the rest of the following sections, other three steps of the methodology will be explained.

4.2. Basic Definitions

Before starting the classification procedure, defining some basic criteria regarding classification with respect to the scope of the study is fundamental.

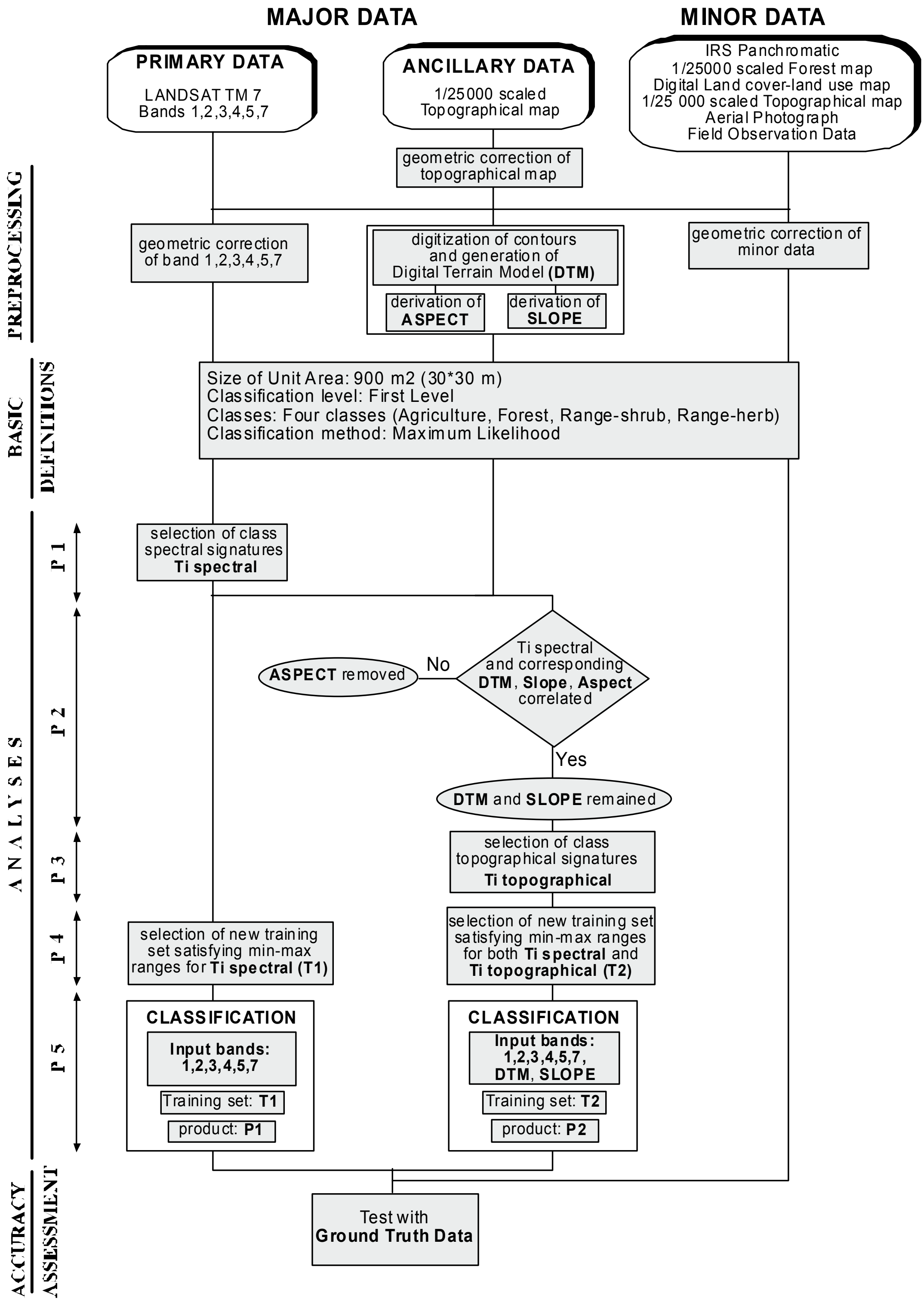


Figure 4.1: General Framework of the Study

Clearly identifying the principles also prevents the probable confusion and ambiguities that would occur later during classification.

Principle criteria for this study were defined in a manner that the restrictions introduced by principles should comply the scope of the study. In CORINE (Coordination of Information of the Environment) Program the principle criteria prior to image classification were affirmed as (1) mapping scale, (2) size of the unit area and (3) land cover nomenclature respectively (CORINE, 1993). These headings were adapted to the study as (1) Size of the unit area, and (2) land cover nomenclature, mapping scale was excluded from the list of definitions since the product derived from this study is not bound to be a part of any standard or national geographical map. “In what scale the product should be?” is not the question of this study.

4.2.1 Size of the Unit Area

Size of the unit area is the surface of the smallest unit within a map. Once the unit area is defined, a feature with dimensions smaller than the unit area cannot be delineated or clustered throughout classification.

Unit area, so-called minimum mapping unit is defined for a particular study after the thematic objectives, printing legibility and budgetary constraints are evaluated (CORINE, 1993). For this study, the only priority is given to the thematic information and other two objectives mentioned were negligible. Thus the unit size was defined as one pixel size, which corresponds to $900m^2$ (30 x 30 meters) for the study area.

Defining one pixel size as the unit area restricts any generalization, provided that the minimum area that was allowed to be mapped is already the minimum area that can be mapped. As a result of this definition; the classification product will not be modified using any generalization procedures such as hole-filling or modal filter.

4.2.2. Classification Level and Classes

Describing the nature of the Earth's surface has been the problematic of geoscientists interested in the definition and the distribution of the phenomenon. When describing the land cover, the issue is similar to that of taxonomists who, for instance, classify an animal into particular place in the hierarchy of animal kingdom. Likewise, a phenomenon on the Earth is classified into a certain type of unit in the hierarchy. The procedure of identification and labeling into classes is called classification (Mather, 1999).

A figure is offered (Figure 4.2) that attempts to represent the hierarchical structure of the feature domain on the Earth's surface. The figure is actually a synthesis of the two schema previously adopted by Jain (1989) (Appendix D1) and CORINE program (1993) (Appendix D2)

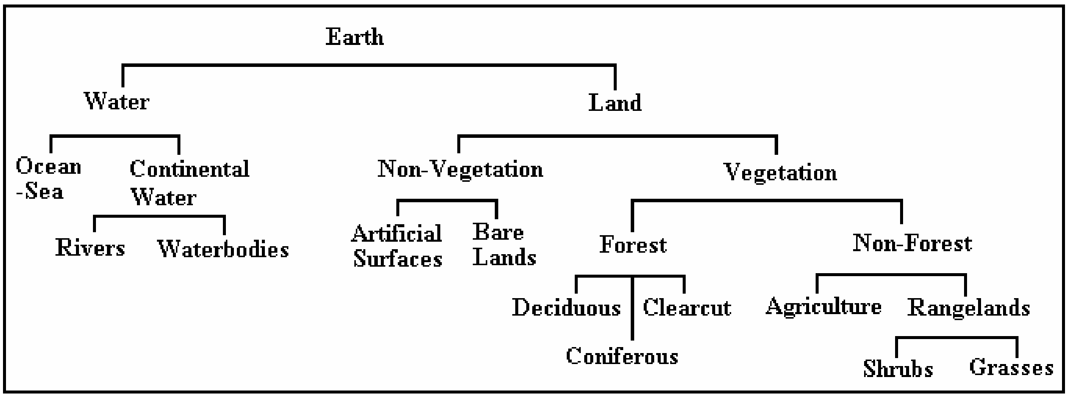


Figure 4.2: Hierarchical Categorization of Feature Domain on Earth

The hierarchical categorization of the earth surface can be extended with subdivisions until an individual feature unit is reached. However, classifying an image is rather different than classifying land cover of a specific area on Earth. Identification in individual feature level is dependent on the scale of observation and quantity of attributes available for the features that are used to determine class membership (Mather,1999), and is not possible yet through today's remote sensing technologies.

For a successful image classification aimed to detect land cover, it is critical to carefully select and define the classes of interest. This requires the use of a classification scheme containing taxonomically correct definitions of information classes (Jensen, 1996). There are several classification schemes attempted to categorize land use and land cover for the use with remote sensing data. The two of the most widely accepted and practiced are USGS Land Use and Land Cover Classification System by Anderson et al. (1976) (Appendix E.1) and CORINE (1993) Land Cover Classification System (Appendix E.2). Both of the schemes are attempted to cover whole feature variety on the Earth surface but the categorization slightly differs. Classification scheme; involving two information; (1) level of classification, (2) classes and their definitions for this study was defined under the requirements and restrictions of the study.

4.2.2.1. Classification Level

Classification level denotes the level of thematic detail for classification. Most of the classification schemes are designed to use multi-level information. A multi-level system is devised because different degrees of detail can be obtained from different types of product with different spatial and spectral resolution. Since the level of classification is dependent on the sensor system and image spatial resolution (Table 4.1), the level of classification for the study was set taking the image's information capability into account, since the primary data source for the study is Landsat 7 ETM+, which has multispectral bands with 30x30 meters resolution; it was quite reasonable to perform a first level classification.

Table 4.1: Relation between classification level and data characteristics (Jensen, 1996)

Classification Level	Typical Data Characteristics
I	Landsat MSS (79x79 m), Thematic Mapper (30x30 m), and SPOT XS (20x20 m)
II	SPOT Panchromatic (10x10 m) data or high-altitude aerial photography acquired at 12,400 m or above
III	Medium-altitude data acquired between 3,100 and 12,400 m
IV	Low-altitude data acquired below 3.100 m

4.2.2.2. Classes

For the study it was proposed to classify all the land cover and land use categories within the scene at the beginning. The land cover and land use categories in the study area were composed of five Level I classes which were;

(1) Urban and Built-up Land
(2) Agricultural Land
(3) Range Land
(4) Forest
(5) Water Bodies

Urban and Built-up Land in general involve areas of intensive use with majority of land covered by artificial structures. This category includes cities, villages, transportation, power and communication facilities, commercial centers, and industrial units. Urban Built up land in the study area was composed of a part of Kızılcahamam urban development, several rural settlements, industrial developments, sites under construction and transportation.

Agricultural Land can be broadly defined as land where crops and other yields for particular uses are cultivated. The category includes cropland, pasture, orchards, grooves, vineyards, nurseries and confined feeding. The agricultural lands in the study area were composed mainly of cereal farming also some vegetable farming was practiced; however it was a small proportion in the total area of agricultural vegetation. Agricultural land in the study area is composed of vegetated and non-vegetated cultivation lands.

Rangeland is defined as the land with natural vegetation species such as grasses, grasslike plants, brushes and shrubs. Rangeland definition covers a broad variety of natural plants from i.e., grasslike plants to shrubs with distribution from sparse to dense in the nature. Rangelands of the study area are composed of common Anatolian steppe cover, which includes a wide range from particular land with little or no vegetation to native shrubs and brushes of maximum 2-meter height distributed densely in the terrain. The term range is hence, a very broad expression in defining the natural cover of the study area.

Forest is defined as the land where trees with height of five meters or more are densely distributed (OGM, 2002). The measure of density of distribution is the percentage of the closure of canopy, where areas with closure of 10% or more are accepted as forest. However, the lands where the closure is less than 10% but there is no other use or activity dominating the area are also accepted to be forest (Lillesand and Kiefer, 1994). Forestlands in the study area are composed mainly of black pine (Pinus nigra Arnold.); a coniferous species and a less amount of deciduous species majority of which are oaks (Quercus sp. L) especially in the northern parts of the study area and at the fringes of Kızılcahamam.

Water bodies include streams, canals, lakes, reservoirs and bays in general. Water bodies were composed of a dam reservoir and two streams within the extent of the study area.

Photographs taken at the field study are available for some of the classes including agriculture, rangelands and settlements (Appendix F).

From the five Level I classes in the study area mentioned above, two of them were excluded from the procedures associated with image classification. These classes are Urban or built-up land and water bodies.

For this study, the main reason for excluding the urban built-up land from the classification procedure was that it is a land use class specifically. Although ancillary data is widely used to improve classification in land cover applications, their use in land use applications is less common (Westmoreland and Stow, 1992).

The term land cover relates the type of feature present on the surface of the earth, however the term land use relates the human activity; a social or economic function practiced on a particular area. Whereas land cover information can be directly interpreted by means of spectral characteristics of an image, additional information sources are needed to reinforce the image data in order to identify whether the area mentioned is an area associated with human activities (Lillesand and Kiefer, 1996). This supplementary data is a usually a thematic map or information regarding the type of use of a specific area or construction and often becomes more critical than the spectral data. Since the remotely sensed imagery is the primary data source for this study, superiority of an ancillary data, its becoming more important than, and even supersede spectral data is unacceptable. Other reason for excluding the urban built-up land was regarding the scope of this thesis where the aim is to test the affirmative effect of topography on improving the accuracy of classification. However, human factor when exceeded a trade-off between required development area and present suitable area, is often challenging, means that land use associated with human activities can be practiced anywhere even unusual, regardless of the topographical restrictions, but dependent on other parameters instead. Development of new residential areas at the fringe of the city on the steeply sloping terrain is an example of independency on topography. Construction of industrial units nearby the transportation is an example of dependency on the parameters other than topography. Another reason for excluding the class from classification relates the spectral properties of the urban or built-up land class. Multispectral image classification tends to be most successful for scenes characterized by homogenous cover. Urban built-up land however shows high spectral variability, which may cause high amount of misclassifications yielding unreliable classification results. Use of an unreliable result in testing effect of topography would not make sense.

The main reason for excluding the water bodies from the classification was that clear water bodies with distinct and unambiguous spectral signatures are the most easily classified information class within a multispectral image, and there is no need to support classification of such water bodies with additional information, also considering that adding topographical information especially the elevation data may reduce the accuracy of the water bodies class rather than improving it, since elevation is constant for corresponding water pixels, and this may not yield meaningful relationship between elevation and water class signatures.

As a consequence of these statements, land cover classes remained to be classified were; (1) Agriculture, (2) Range Land and (3) Forest. At this point, rangelands in the study area were reevaluated, because; rangelands of the area apparently consisted of two contextually different categories, which are herbaceous rangeland and shrub rangeland. Thus it was reasonable to make a subdivision for rangeland category although it may violate Level I of some well-known classification schemes (Anderson, 1976; CORINE, 1993). As a consequence of this subdivision; ultimate list of land cover ended up with four classes;

Class1: Agriculture

Class2: Rangeland-shrub (Range-shrub)

Class 3: Rangeland-herbaceous (Range-herb)

Class4: Forest

When an information class defined by the analyst and spectral class defined for a particular category coincides or is very close, classification yields successful results. This statement is provided for the three classes but not for Agriculture.

Agriculture is a special class because; it is regarded as a land use class. Related to this, any particular land dominated with agricultural activities is considered to be agricultural land, regardless of its being vegetated or non-vegetated. Vegetated and non-vegetated lands inherently have different spectral reflectance values and patterns and this raises intolerable results in classification.

Under these circumstances, non-vegetated agricultural lands were set apart from the classification procedure; hereafter agriculture class represented the vegetated agricultural land only. In other words it was treated as a specific land cover rather than a land use class.

4.2.3. Classification Method

Each feature or type of land cover may have their own spectral characteristics in different bands of electromagnetic spectrum. Image classification for this study is aimed to convert spectral data into four land cover classes.

An information class implies the class defined by the analyst. However spectral classes are those that are inherent in the remotely sensed data (Jensen,1996). Thus, a spectral class may not necessarily correspond to a specific information class defined by the analyst.

In order to extract afore mentioned land cover classes from the image data supervised classification method was quite appropriate. The reason for selecting supervised classification was that, a supervised classification ideally yields information classes since the training samples were defined for each specific information class by the supervisor. However, an unsupervised classification predictably yields spectral classes, because the training set is automatically calculated from the image spectral data.

A conventional supervised classification clusters the pixels into information classes by means of training data based on probability distribution models for the cases of interest. (Favela and Torres, 1998). Maximum Likelihood classifier is the most commonly used supervised method and is supposed to provide better results compared to the other supervised methods (Strahler, 1980; Bolstad and Lillesand, 1992; Foody et al.,1992; Deusen,1995; Maselli et al., 1995). Superiority of Maximum Likelihood classifier over the other supervised methods such as minimum distance to mean, parallelepiped and etc. is due to its sensitivity on the shape, size and orientation of a cluster (Shrestra and Zinck, 2001). However, Maximum Likelihood classifier is not particularly a robust technique, in handling natural geographical data since it assumes that the data has normal distribution.

4.3. Analyses

A five-phased methodological framework was proposed for developing a procedure for the integration of ancillary topographical information into image classification.

First phase basically involves understanding class spectral characteristics. In this phase, representative sets of pixel values that spectrally characterize classes to be extracted from the image data were defined. Class spectral signatures compose the initial training set for the multispectral image data. However, this set is not used for training a classification procedure but, it serves as prior information for the later redefinition of training data.

Second phase involves analysis of the information content of data sources and quantification of the relationship between specific classes and particular topographical data. Dependent on the significant relationships, ancillary topographical data that may contribute to improvement of classification accuracy was determined. The topographical data, which was considered to add no to classification, was excluded from the remainder of the analyses.

In the third phase the ancillary topographical data, which was tested and qualified to be used in the classification, was examined for the selection of class topographical signatures. A procedure similar to that performed in the first phase was carried on. However, this time the aim was to define the representative sets of values that topographically characterize classes of interest.

The fourth phase; redefinition of training sets is very critical. The aim is to define two training sets where Training Set 1; involves class spectral signatures only and Training Set 2; involves both class spectral and class topographical signatures. The signatures were selected with respect to class spectral and topographical signatures previously defined. That's why the task is described with the term redefinition. Training set 1 was then used to train classification of spectral data only and Training set 2 was used to train classification of both spectral and topographical data

The fifth phase is classification. Classification for the study was aimed to prove the affirmative effect of topographical data on image classification accuracy by generating two products to be compared, where first was a product of classification based solely on spectral data (Figure 4.10) and second was a product of classification based on both spectral and topographical data (Figure 4.11).

The five phases were followed by accuracy assessment, which is essential task for implementation of classification. Classification products were assessed to be verified for their correspondence to the ground truth information. In the accuracy assessment step; the effects of the method were achieved and the extent of contribution of topographical data to image classification with the integration of ancillary topographical data was quantified.

4.3.1. Phase 1: Definition of Class Spectral Signatures

Supervised classification as the term supervised denotes, is the classification supervised by the analyst. Supervising procedure is often performed by manually defining training samples to guide classification. Whereas image classification is a highly automated procedure, selection of training set is nothing but manual.

The objective of training set selection is to assemble a set of statistics that describe the spectral signature/response pattern for each category to be classified within the image. Supervised classification methods define classes by analyzing the spectral signatures in each training area and determining their statistical properties. The defined class properties are then used as the basis for classifying the input raster set. The success of supervised classification to some extent depends on the quality of the training set. The goal is to define training areas of sufficient quality and quantity, so that every type of feature in the input can be assigned to a meaningful class.

Training samples were selected for all classes overall the image, ensuring that they are good representatives of each information class. When selecting training samples there is a trade-off between having a sufficiently large sample size to generate accurate statistical parameters used by classifiers, and a restrictive sample size to ensure class seperability. It is essential not to exclude any samples that would contribute to the representation of the class's signature, but it is also equally essential not to include some of the samples redundantly not to computationally yield a training set of inferior quality. Generating training sets for spectral classes that are distinct and mutually exclusive do not pose problems (Mesev, 1998). However, it should always be kept in mind that this is the ideal case; samples representing a particular phenomenon may also cover a part of another. Thus, training set selection for natural phenomenon is often problematic.

Selection of training set was given great importance for the study and samples were selected precisely as possible to represent land cover classes' signatures well. Because; the generation of representative training statistics is sometimes more important for obtaining accurate classifications than is the selection of classifier algorithm itself.

There are several ways to collect training samples. For this study on-screen polygonal selection was preferred. Polygonal selection tool is an effective way of training samples selection compared to those which are point based, because it enables the selection of multiple pixels located nearby, belonging to the same class instead of pointing them one by one. Training samples were collected by polygonal selection tool randomly from all over the scene for the four land cover classes to be extracted from the multispectral Landsat TM bands.

Selection of the training set for the study basically depended on visual interpretation. A certain time was devoted to understanding visual components of land cover classes in the study area within particular band combinations and other reference data. Visual interpretation of a feature involves examining its tone, color, texture, pattern, shape, etc. Landsat 453 color composite was the primary reference to select training samples, since this band combination was regarded to represent distinctions of land cover classes better in the study area compared to other particular combinations. Training samples were collected from 453 composite making use of color and tone information. IRS panchromatic image, an aerial photograph, topographical map and forest map was used to obtain texture, pattern, shape information especially when 453 composite was insufficient in spatial detail. Visual interpretation of a feature involves examining its tone color, texture, shape, etc. (Figure 4.3).

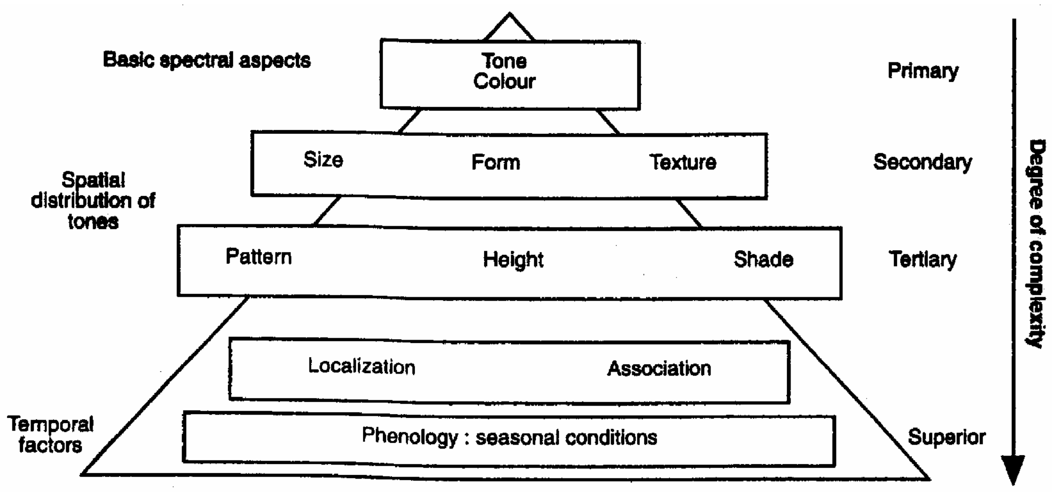


Figure 4.3: Relationship of Image Elements to Visual Interpretation (CORINE, 1993)

Training set selection for the study was a multiphase procedure; it was not completed all at once. Selected training set was tested both for seperability and representativity, if not satisfied with the results; the training set was modified and tested again. This procedure continued since a balance between sample size and sample error was supplied. In other words; it was concluded that the training set is satisfactory, when samples were at that critical balance where attempts to make modification on a class's training samples so as to generate more representative class signatures violates general seperability or visa versa; attempts to increase seperability violates a class's representativity.

The seperability of the training samples was tested with the help of a dendrogram. A Training Set Dendogram is used to obtain the results of a hierarchical analysis of the class signatures in graphic form. This analysis performs a successive grouping of pairs of classes on the basis of distance between class centers in feature space. The closest two classes are merged, a new joint class center computed, and class-center distances recalculated; this process repeats until all classes are merged into a single class. Classes that join together near the left side of the diagram are closely related in their spectral properties, and the degree of relation decreases to the right as the size of the class groups increases. The spectral seperability of signatures were tested by "Transformed Divergence" since it provides more appropriate seperability definition for Maximum Likelihood method (Swain and Davis, 1978). Transformed divergence values range between 0.0-2.0; where 0 indicates the class pairs that cannot be distinguished and 2 indicates full seperability between two classes Seperabilty between 1.9-2.0 is accepted as best seperability, 1.7-1.9 as fairy good seperability, 0-1.7 as poor seperability (Swain and Davis, 1978).

Figure 4.4 shows the seperability of the classes making use of a dendrogram.

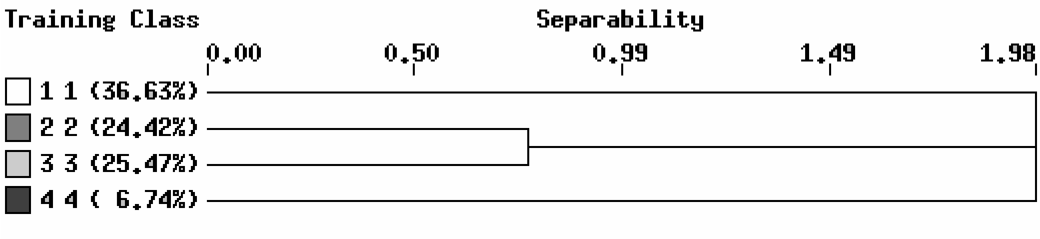


Figure 4.4: Seperability of Initial Training set by means of Transverse Divergence measurement

Transformed divergence of approximately 1.5 is often accepted to represent spectral region of merging (Wiseman, 2002). However class2 and class 3; range-shrub and range-herb which have low seperability were previously defined as classes necessarily to be classified. And furthermore this set was not directly used for classification. The seperability of the training set may change after redefinition procedure mentioned in section 4.3.4.

Although there is no certain upper limit for the number of samples to be collected for each class, there are several suggestions about minimum number of samples to be collected for each class. Jensen (1996) stated that the general rule for size of training samples was 10n where; n is the number of bands to be classified. However Mather (1989) recommended that sufficient number of samples to be collected is 30n. For the study, 32-174 pixels were selected by means of 8-28 polygons to define the training set for each class, and number of pixels used were 443 totally (Table 4.2). This number corresponds to approximately 70n where number of bands used was six and is quite sufficient for generating class signature statistics. Number of training samples collected for forests were noticeably small; however this was a consequence of the small proportion of forestland within the image, rather than an undersampling problem.

Table 4.2: Number of training samples and selecting polygons for each class

Class	Number of polygons	Number of Training Samples
Agriculture (Class1)	25	174
Range-shrub (Class2)	21	116
Range-herb. (Class3)	20	121
Forest (Class4)	6	34
TOTAL	72	443

Training set is then used for generating class spectral statistics that will guide further refinements on training data. Fundamental characteristics that define class's spectral signatures are univariate and multivariate statistics where, univariate statistics are; minimum, maximum, mean, variance, standard deviation) and multivariate statistics are; variance-covariance matrix and correlation matrix. Minimum, maximum, mean and the variance values for the Training set is given in Table 4.3.

Table 4 3: Minimum, maximum, mean and variance values for the Initial Training set

CLASS1 Agriculture	minimum	maximum	mean	variance
TM1	64	84	72,81	12,99
TM2	55	75	63,28	17,01
TM3	41	73	53,96	45,18
TM4	82	152	100,85	173,92
TM5	67	112	84,01	87,09
TM7	34	70	48,66	55,23

CLASS 2 Range-shrub	minimum	maximum	mean	variance
TM1	70	86	78,18	13,03
TM2	57	77	67,03	14,66
TM3	54	82	67,68	44,70
TM4	51	85	67,97	56,02
TM5	82	135	109,61	139,61
TM7	52	98	76,04	108,45

Table 4 3: Minimum, maximum, mean and variance values for the Initial Training set (continued)

CLASS 3 Range-herb.	minimum	maximum	mean	variance
TM1	75	95	83,61	19,63
TM2	63	87	72,71	28,32
TM3	60	98	76,14	68,82
TM4	64	91	77,80	45,53
TM5	107	165	129,55	135,46
TM7	71	124	90,97	124,76

CLASS 4 Forest	minimum	maximum	mean	variance
TM1	61	71	65,90	6,66
TM2	46	58	50,03	8,80
TM3	35	52	42,25	18,76
TM4	50	68	58,31	18,60
TM5	43	89	57,37	112,99
TM7	24	49	34,93	45,27

Actually, this training set was an initial study. It was aimed to serve as a basis for generating the redefined training sets by making use of minimum and maximum ranges. Hence, this training set was remarked as Initial Training set and was not used directly to train any classification procedure.

4.3.2. Phase 2: Determination of Class–Topographical Data Relationship

The effective use of ancillary data requires a consistent and known relationship between ancillary data and the subject of interest (Westmoreland and Stow, 1992). For a particular application regarding the integration of ancillary data into classification; understanding ancillary data in the context and knowing how this data will contribute to image classification is of great importance.

In this study, each of the ancillary data was evaluated for its degree of correspondence with land cover categories. The analyses were aimed to find out whether there is a relation between a specific class and a specific ancillary data. But the case involves comparison of two different types of variables, where class is dichotomous and topographical data are continuous. Biserial Analysis is quite adequate for making this kind of comparison since it pertains to the case where one variable is dichotomous and the other is non-dichotomous. By convention, the dichotomous variable is treated as the X variable, its two possible values; being class A or not being class A, coded as X=1 or X=0; and the non-dichotomous variable is treated as the Y variable (Vassar, 2002). The task is simply correspondence analysis of Y variables associated with class A, and Y variables not associated with Class A. If the variation between these two sets is wide, than this means X variables being A or not being A extensively effects Y variables and is the confirmation for a high correlation between class A and Y variable. If there is little or no significant difference between the two data sets than there is no relation between Class and Y variable.

Data to be tested for correspondence were the four land cover classes and the ancillary topographical data consisted of elevation, slope and aspect.

Land cover data involved training samples assigned to land cover classes; agricultural land, range-shrub, range-herb., and forest. The training samples were collected randomly from all over the study area and were spectrally good representatives of their associated classes, so, they formed an adequate test set.

Topographical data involved the pixel values corresponding to the training samples.

Test was aimed to discover the effect of topography on the land cover classes. Thus, comparison of topographical variables corresponding to a specific class's training samples and topographical variables corresponding to other remaining classes' training samples may offer meaningful information about the effect of topography on that specific class. To be more specific, for example, elevation values corresponding to agriculture's training samples are gathered and labeled as set1. On the other hand, elevation values corresponding to samples of other three classes are gathered and labeled as set2. If we can compare set1 and set2, we can achieve valuable correlation pertaining to information about the effect of elevation as an entity on class: Agriculture.

Biserial analysis was performed with 95% confidence interval for each class and each topographical attribute to measure a probable correlation. An online biserial coefficient calculator was used (Vassar, 2002).

The correlation coefficients for 24 tests ranged between the minimum of 0.02 to maximum positive 0.65, and maximum negative 0.41 (Table 4.4); where 0 denotes is there is no relationship (no correlation), 1 is perfect relationship and -1 is perfect negative relationship.

Table 4.4: Biserial correlation coefficients for four land cover classes and topographical data

Topographic Parameter	Test Set		Correlation Coefficient
Elevation	agriculture	non- agriculture	**+0,62**
Slope	agriculture	non- agriculture	**+0,65**
Aspect	agriculture	non- agriculture	**+0,02**
Elevation	range-shrub	non- range-shrub	**-0,34**
Slope	range-shrub	non- range-shrub	**+0,48**
Aspect	range-shrub	non- range-shrub	**-0,11**
Elevation	range-herb.	non- range-herb.	**-0,41**
Slope	range-herb.	non- range-herb.	**+0,08**
Aspect	range-herb.	non- range-herb.	**+0,06**
Elevation	forest	non- forest	**+0,1**
Slope	forest	non- forest	**-0,5**
Aspect	forest	non- forest	**+0.24**

The result of the biserial correlation analysis was evidence for the relation between specific land cover classes in the area and the terrain attributes. The significance test of the correlation coefficient verified that all of the correlations which have correlation coefficients greater than approximately 0.30 are significant. Significance level, often called the p value is the probability that a statistical result as extreme as the one observed would occur if the null hypothesis were true. If the observed significance level is small enough, usually less than 0.05 or 0.01, the null hypothesis is rejected (SPSS help).

$P < 0.0001$ for all of the high correlations is a very low p value under 0.05 (Appendix G), which is generally accepted as border line limit for significance (Statsoft, 2002).

In general it was comprehended that land cover classes in the study area are highly related or in other words highly dependent on the elevation and slope of the terrain. However, aspect has little effect on most of the classes' subsistence and distribution within the study area.

Examining individually;
Agriculture (C1) is highly correlated with both elevation and slope, but with coefficient value of 0.02 the least correlation is also between agriculture and aspect among all tests. Range-shrub (C2) is also correlated with elevation and slope to a degree, the correlation between elevation and range-shrub is however, a negative relationship, aspect has poor correlation with C2. Range-herb (C3) has negative relationship with elevation in a reasonable level. However it has very poor correlation with both slope and aspect of the terrain. Forest (C4) has fair degree of negative relationship with slope and a reasonable relationship with aspect but it is randomly correlated with elevation of the terrain with a low significance.

Under these evaluations; aspect was considered to add little or no to classification and incidentally it was excluded from the remaining part of the study associated with the integration of topographical data into classification procedure. As a consequence of the biserial analysis; elevation and slope data were qualified to be used as ancillary topographical data in classification.

4.3.3. Phase 3: Definition of Class Topographical Signatures

A set of samples that describe the spectral characteristics of each land cover class was previously selected as mentioned in section 4.3.1. Since it was realized after the evaluations in previous section 4.3.2. that elevation and slope data are correlated with land cover classes in the study area, the next question is how this data will contribute to classification.

Integration of ancillary data into classification is followed by two approaches. The first is inserting ancillary data into classification as additional band so-called logical channel method, and the second is classifier modification by changing prior probabilities.

Changing prior probabilities either by using information gathered from bands or ancillary data actually makes areal estimations about the final product. Weights to be assigned to each class are derived from these data, and accuracy of the classification is highly dependent to accuracy of estimates. However, for this study any of the topographical data proposed to be integrated into classification can offer such reliable estimates about the areal distribution of classes in the final product, therefore, use of classifier modification was inadequate for this study with the information that can be derived from elevation or slope data.

The other approach, addition of ancillary data as a separate channel seemed quite reasonable, however, simple addition of non-spectral data as input into classification adds little to accuracy of classification as it was stated by Hutchinson (1982). In order to make efficient use of ancillary data in classification procedure, modifying training set so as to guide classification to take ancillary data characteristics for each class into account would be a solution. This operation simply requires, topographical signatures for each land cover class to be included into training set.

Selection of class topographical signatures was aimed just the same as selection of class spectral signatures, however, selection of topographical signatures was rather different than selection of spectral signatures. This is because, topographical samples should be selected according to being more likely to be observed in a specific class. As a consequence of this, topographical signatures cannot be collected via visual interpretation.

In order to gather topographical samples, a kind of stratified random sampling was applied making use of the pixels that satisfy the spectral ranges defined in class spectral signatures for each class but not the class spectral signatures themselves. The reason for using the pixels satisfying the minimum and maximum ranges for spectral signatures instead was the need for collecting more samples to better represent the topographical distribution and different samples to derive unbiased topographical relations.

Fleeming and Hoffer (1979) used stratified TSRS (Topographical Stratified Random Sampling) for determining forest types-topography relationship via 4450 points for 3750 km2 study area. Making a simple ratioing for the study area of 66 km2, a total of approximately 80 points were required. Anyway for this study total number of 1200 samples was gathered, in order to better represent site characteristics (Table 4.5), since topography shows wide variability and there are four different types of land cover classes.

Table 4.5: Number of Stratified Random Samples collected for obtaining class-topography relationship

Class	Number of Samples
Agriculture (C1)	128
Range-shrub (C2)	482
Range-herb. (C3)	540
Forest (C4)	50
TOTAL	1200

These sample sets for each class were than used to stratify elevation and slope data. Subsequent to that, the elevation and slope values corresponding to class signatures were randomly selected. Location of the samples was random and those samples did not necessarily overlap with that of spectral training samples.

Topographic distributions were developed to determine the frequency of occurrence of each elevation or slope value for any given land cover class. Frequency histogram was a valuable supplementary in defining elevation or slope ranges where classes were most likely to occur. Data ranges representing class topographical signatures were determined with the help of histogram graphics.

Histograms showing the distribution of observed topographical values for each class were truncated by removing the observations at the two tails of the histogram so as to exclude deviated region of the distribution profile.

The topographical distribution for land cover classes were represented in Figure 4.5.

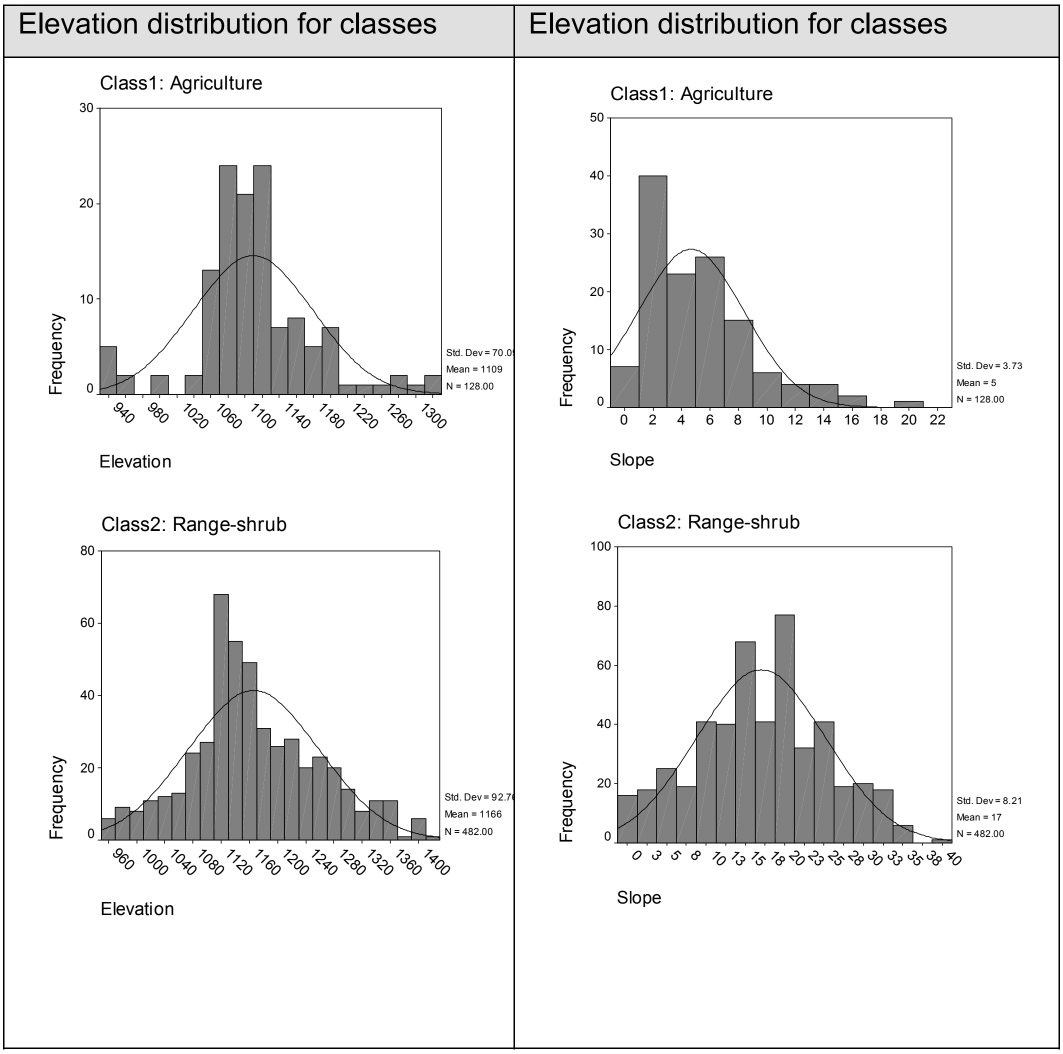


Figure 4.5: Histograms for distribution of topographical values corresponding to spectral signatures of each class

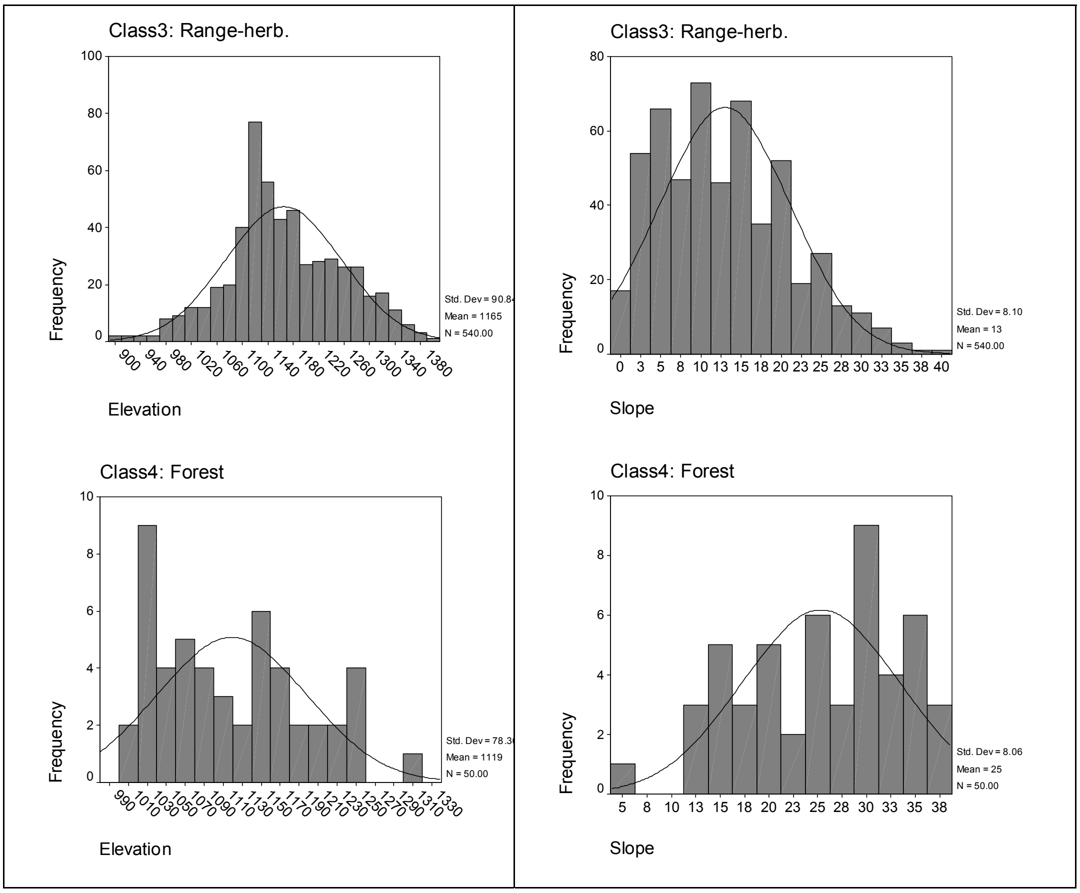


Figure 4.5: Histograms for distribution of topographical values corresponding to spectral signatures of each class (continued)

The topographical distributions are not normal and they have outlier elements which may cause erroneous representations for the classes of interest. The tails where outliers are situated were removed manually from the sets of topographical training data. By this way, minimum and maximum ranges for topographical attributes associated with four classes were redefined. The new minimum and maximum ranges were given in Table 4.6.

Table 4.6: Minimum, maximum, mean and variance values of topographical signatures for each class

CLASS1: Agriculture	minimum	Maximum	mean	variance
ELEVATION	945	1202	1109	4912
SLOPE	0	8	4,6	13,92

CLASS 2: Range-shrub	minimum	Maximum	mean	Variance
ELEVATION	1000	1363	1165	8603
SLOPE	6	32	17,00	67,34

CLASS 3: Range-herb.	minimum	Maximum	mean	variance
ELEVATION	1110	1375	1164	8251
SLOPE	2	26	13,00	65,66

CLASS 4: Forest	minimum	Maximum	mean	variance
ELEVATION	1023	1240	1119	6140
SLOPE	12	37	25,32	64,96

Boxplot graphics derived from topographical signatures also give valuable information about the coincidence and seperability characteristics of elevation (Figure 4.6) and slope (Figure 4.7) for each class

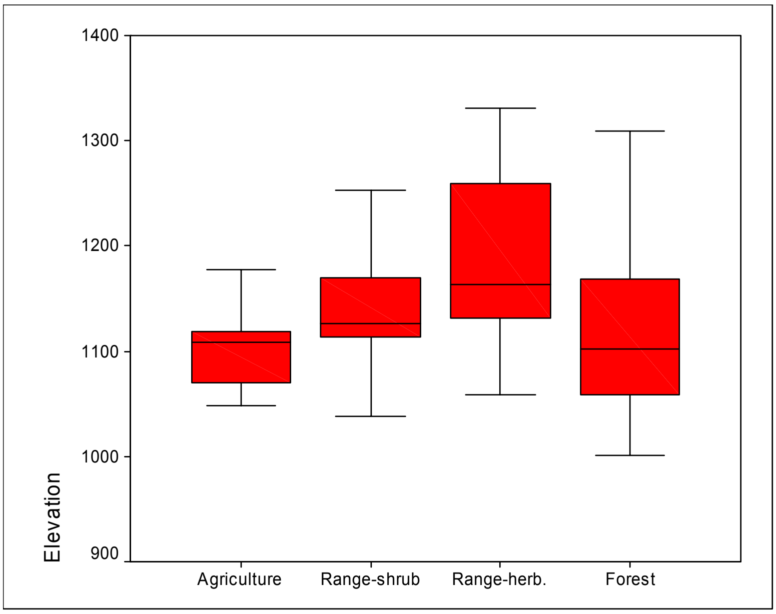


Figure 4.6: Box plot of elevation signatures showing means and 1 standard deviations

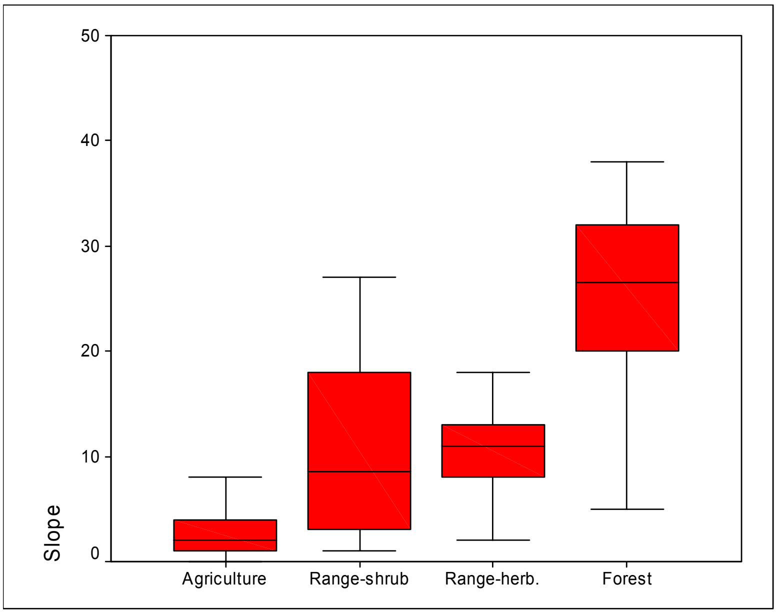


Figure 4.7: Box plot of slope signatures showing means and 1 standard deviations

4.3.4. Phase 4: Redefinition of Training Sets

Selection of the initial training set including both the spectral and the topographical signatures was implemented sequentially as mentioned in section 4.3.1. and 4.3.3.

However the aim of selecting the initial training set was actually to obtain the spectral and topographical range of values that represent land cover classes rather than using this information directly for training of classification procedure. The training set was used as supplementary information to serve as basis for next and the last step in training set definition.

The effect of ancillary topographical attributes on classification accuracy was intended to be tested as a requirement of the study. The test basically depended on a comparison of two products one of which was derived from spectral data and the other from both spectral and the topographical data. This can be achieved by classifying the multispectral image data by means of training set involving class spectral signature only, to yield Product 1, and on the other hand; classifying multispectral image data and topographical data by means of training set involving both class spectral and topographical signatures to yield Product 2.

As a consequence of this, two training sets were redefined to satisfy the afore mentioned criteria; Training Set 1; involving class spectral signatures only and Training Set 2; involving both class spectral and class topographical signatures. Training Set 1 has already been collected as mentioned in part 4.3.1. and was ready to be used to train classification in order to derive Product 1. However Training Set 2 was problematical. The cause of the problem with Training Set 2 was that spectral training samples and the topographical samples did not coincide when merged in a single training raster. This was inherently a result of different manner and method used for collecting spectral and topographical samples. Solution of this problem could be to create a new Training Set 2 with sample pixels every single of which can satisfy both spectral and topographical class signatures. However, this task introduces a new problem of challenging the critical issue to ensure for the two training sets, that is; preventing the class spectral signatures constant in both Training Set 1 and Training Set 2. If this is not provided, we can never make sure that the difference in between Product 1 and Product 2 is due to topographical effect. The question to mention here is "*is it possible to manually select training samples that would also represent topographical signatures, without deforming the class spectral signatures?*" Answer to this question is almost no. Because, collecting samples which can satisfy topographical signatures and do not change the characteristics of spectral signatures is impractical manually. Therefore an automated selection procedure was adopted to solve this problem. In order to implement automated selection, all of the raster attributes were transferred to vector points every one of which represented a standard raster grid of 30x30 meters. Pixel values data for every single raster related with the classification procedure were transferred as data base records into database tables attached to the internal table of point elements (Figure 4.8).

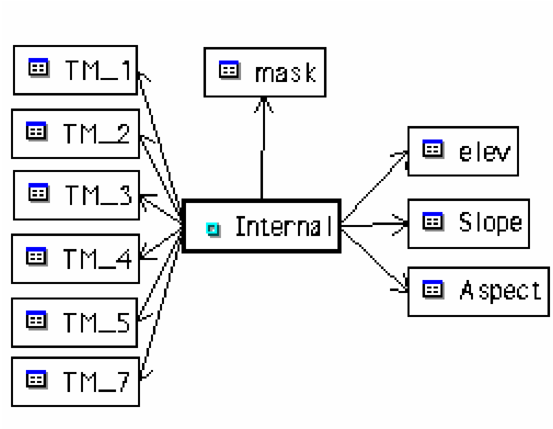


Figure 4.8: Database tables of point elements that are representing raster pixel values

Two queries one of which is for Training Set 1 and other for Training Set 2 were performed with respect to minimum and maximum ranges. Those ranges were previously defined for each input of each class in the initial training set definition stage (section 4.3.1.).

First query given was aimed to select pixels within the minimum and maximum ranges defined for bands only. And the second query given was aimed to select pixels within the minimum and maximum ranges defined for bands, elevation and slope.

Query 1:

```
if      TM_1.RastValue >= 64 and TM_1.RastValue <= 84
        and TM_2.RastValue >= 55 and TM_2.RastValue <= 75
        and TM_3.RastValue >= 41 and TM_3.RastValue <= 73
        and TM_4.RastValue >= 82 and TM_4.RastValue <= 152
        and TM_5.RastValue >= 67 and TM_5.RastValue <= 112
        and TM_7.RastValue >= 34 and TM_7.RastValue <= 70      assign to Class1

if      TM_1.RastValue >= 70 and TM_1.RastValue <= 86
        and TM_2.RastValue >= 57 and TM_2.RastValue <= 77
        and TM_3.RastValue >= 54 and TM_3.RastValue <= 82
        and TM_4.RastValue >= 51 and TM_4.RastValue <= 85
        and TM_5.RastValue >= 82 and TM_5.RastValue <= 135
        and TM_7.RastValue >= 52 and TM_7.RastValue <= 98      assign to Class2

if      TM_1.RastValue >= 75 and TM_1.RastValue <= 95
        and TM_2.RastValue >= 63 and TM_2.RastValue <= 87
        and TM_3.RastValue >= 60 and TM_3.RastValue <= 98
        and TM_4.RastValue >= 64 and TM_4.RastValue <= 91
        and TM_5.RastValue >= 107 and TM_5.RastValue <= 165
        and TM_7.RastValue >= 71 and TM_7.RastValue <= 124     assign to Class3

if      TM_1.RastValue >= 61 and TM_1.RastValue <= 71
```

```
           and TM_2.RastValue >= 46 and TM_2.RastValue <= 58
           and TM_3.RastValue >= 35 and TM_3.RastValue <= 52
           and TM_4.RastValue >= 50 and TM_4.RastValue <= 68
           and TM_5.RastValue >= 43 and TM_5.RastValue <= 89
           and TM_7.RastValue >= 24 and TM_7.RastValue <= 49         assign to Class4
```

pixels satisfying Query 1 compose the Training Set for bands only (T1).

Ouery 2:

```
if         TM_1.RastValue >= 64 and TM_1.RastValue <= 84
           and TM_2.RastValue >= 55 and TM_2.RastValue <= 75
           and TM_3.RastValue >= 41 and TM_3.RastValue <= 73
           and TM_4.RastValue >= 82 and TM_4.RastValue <= 152
           and TM_5.RastValue >= 67 and TM_5.RastValue <= 112
           and TM_7.RastValue >= 34 and TM_7.RastValue <= 70
           and ELEV.RastValue >= 945 and KRIGING.RastValue <= 1202
           and SLOPE.RastValue >= 0 and SLOPE_s.RastValue <= 8     assign to Class1

if         TM_1.RastValue >= 70 and TM_1.RastValue <= 86
           and TM_2.RastValue >= 57 and TM_2.RastValue <= 77
           and TM_3.RastValue >= 54 and TM_3.RastValue <= 82
           and TM_4.RastValue >= 51 and TM_4.RastValue <= 85
           and TM_5.RastValue >= 82 and TM_5.RastValue <= 135
           and TM_7.RastValue >= 52 and TM_7.RastValue <= 98
           and ELEV.RastValue >= 1000 and ELEV.RastValue <= 1363
           and SLOPE.RastValue >= 6 and SLOPE.RastValue <= 32      assign to Class2

if         TM_1.RastValue >= 75 and TM_1.RastValue <= 95
           and TM_2.RastValue >= 63 and TM_2.RastValue <= 87
           and TM_3.RastValue >= 60 and TM_3.RastValue <= 98
           and TM_4.RastValue >= 64 and TM_4.RastValue <= 91
           and TM_5.RastValue >= 107 and TM_5.RastValue <= 165
           and TM_7.RastValue >= 71 and TM_7.RastValue <= 124
           and ELEV.RastValue >= 1110 and ELEV.RastValue <= 1375
           and SLOPE.RastValue >= 2 and SLOPE.RastValue <= 26      assign to Class3

if         TM_1.RastValue >= 61 and TM_1.RastValue <= 71
           and TM_2.RastValue >= 46 and TM_2.RastValue <= 58
           and TM_3.RastValue >= 35 and TM_3.RastValue <= 52
           and TM_4.RastValue >= 50 and TM_4.RastValue <= 68
           and TM_5.RastValue >= 43 and TM_5.RastValue <= 89
           and TM_7.RastValue >= 24 and TM_7.RastValue <= 49
           and ELEV.RastValue >= 1023 and ELEV.RastValue <= 1240
           and SLOPE.RastValue >= 12 and SLOPE.RastValue <= 37     assign to Class4
```

pixels satisfying Query 2 compose the Training Set for bands, elevation and slope (T2).Number of pixels selected to be training samples are given in Table 4.7. Pixels with the attributes 0 for mask table, where 0 is attached to the elements to be masked and 1 is no action, were excluded from both of the training sets.

Table 4.7: Number of Pixels selected for T1 and T2

T1 Training set for bands only

CLASS	Number of training pixels
Agriculture	4141
Range-shrub	43002
Range-herb.	41103
Forest	899

T2 Training set for bands, elevation and slope

CLASS	Number of training pixels
Agriculture	2799
Range-shrub	28443
Range-herb.	34214
Forest	670

The spectral characteristics including mean and variance for the two new training sets are presented in Table 4.7. Minimum and maximum values concerning spectral data are constant for bothT1 and T2. Elevation and slope values were given for Training set for bands only, although it is not used in anyway. That is to show what the elevation and slope values corresponding to spectral training samples in the normal case. A big amount of change is observed in elevation and slope compared to band signatures of T1 and T2.

Table 4.8: Mean and variance values for T1 and T2

T1 Training set for bands only

CLASS1 Agriculture	mean	variance
TM1	75,99	14,51
TM2	66,27	16,24
TM3	59,99	48,02
TM4	93,61	103,22
TM5	95,24	119,90
TM7	58,00	73,96
ELEVATION	1112,09	699
SLOPE	7,02	37,69

CLASS 2 Range-shrub	mean	variance
TM1	80,02	12,53
TM2	68,80	14,51
TM3	70,24	37,69
TM4	73,06	44,22
TM5	115,29	150,55
TM7	79,81	100,22
ELEVATION	1174,88	6935,55
SLOPE	11,81	66,74

T2 Training set for bands, elevation and slope

CLASS 1 Agriculture	mean	variance
TM1	75,55	15,44
TM2	66,12	17,13
TM3	59,66	53,58
TM4	94,74	112,78
TM5	92,74	122,98
TM7	56,56	81,72
ELEVATION	1102,61	2078,44
SLOPE	3,81	4,62

CLASS 2 Range-shrub	mean	variance
TM1	79,34	12,96
TM2	68,21	15,84
TM3	69,40	40,44
TM4	72,40	47,88
TM5	114,25	142,56
TM7	78,77	98,40
ELEVATION	1189,50	6263,13
SLOPE	14,60	42,25

Table 4.8: Mean and variance values for T1 and T2 (continued)

CLASS 3 Range-herb.	mean	variance
TM1	86,33	15,52
TM2	76,45	20,07
TM3	81,92	50,55
TM4	79,54	38,44
TM5	129,86	85,19
TM7	93,19	74,82
ELEVATION	1160,84	7044,24
SLOPE	12,84	55,95

CLASS 4 Forest	mean	variance
TM1	64,05	13,91
TM2	51,70	16,97
TM3	45,19	44,35
TM4	57,76	35,88
TM5	62,42	78,49
TM7	39,54	69,72
ELEVATION	1139,92	4385,08
SLOPE	25,20	35,88

CLASS 3 Range-herb.	mean	variance
TM1	84,04	5,10
TM2	73,25	9,36
TM3	77,35	14,64
TM4	77,27	20,88
TM5	127,28	87,98
TM7	90,02	43,82
ELEVATION	1183,19	32490,0
SLOPE	7,90	60,84

CLASS 4 Forest	mean	variance
TM1	66,98	5,10
TM2	51,48	9,61
TM3	44,92	18,57
TM4	57,28	18,74
TM5	61,48	90,06
TM7	38,92	46,64
ELEVATION	1127,92	11657,60
SLOPE	25,01	40,96

As observed from the signature statistics there is an overlap concerning range shrub and range-herb. This overlap was mainly due to the nature of the two rangeland classes, which show a degree of spectral similarity. Since the similarity was not that significant to necessitate merging of the two classes into a single class, overlap was accepted of reasonable extent. Although the overlap is acceptable in the spectral domain, representation of overlap in the training raster is impractical since a pixel may have a single value only. For the case, some pixels were satisfying both T1 and T2 criteria. In order to solve the problem, pixels satisfying more than one criterion were removed. The action attempted to provide that every one of pixels selected as training samples have only one class id, removed the overlapped region from both class 2 and 3

The spectral characteristics after overlap extraction made negligible changes over the training spectra. The number of training samples decreased. Especially the samples of range-shrub and range-herb pixels reasonably decreased due to amount of overlapped region (Table 4.9).

Table 4.9: Number of pixels selected for T1 and T2 after overlap extraction

T1 Training set for bands only

CLASS	Number of training pixels
Agriculture	3169
Range-shrub	11800
Range-herb.	13585
Forest	881

T2 Training set for bands, elevation and slope

CLASS	Number of training pixels
Agriculture	2660
Range-shrub	11630
Range-herb.	17591
Forest	672

Training statistics including, mean and variance for the two ultimate training sets are presented in Table 4.10.

Table 4.10: Mean and variance values for ultimate T1 and T2 (overlap-excluded)

T1 Training set for bands only

CLASS1 Agriculture	mean	variance
TM1	75,11	32,26
TM2	65,93	39,06
TM3	58.93	94,09
TM4	95.71	121,22
TM5	94.05	228,91
TM7	56.62	132,02
ELEVATION	1115.02	5584.57
SLOPE	6.28	32,83

CLASS 2 Range-shrub	mean	variance
TM1	77.08	11,83
TM2	65.53	13,54
TM3	64.87	35,76
TM4	68.65	44,48
TM5	100.57	122,54
TM7	69.00	79,56
ELEVATION	1162.71	8302,85
SLOPE	16.52	148,10

CLASS 3 Range-herb.	mean	variance
TM1	86.26	28,72
TM2	76.35	38,44
TM3	81.82	90,06
TM4	79.44	49
TM5	130.18	99,80
TM7	93.33	83,53
ELEVATION	1164.41	8510,06
SLOPE	12.86	55,50

T2 Training set for bands, elevation and slope

CLASS 1 Agriculture	mean	variance
TM1	75.44	55,50
TM2	66.02	63,20
TM3	59.36	172,65
TM4	95.02	173,97
TM5	92.41	225,30
TM7	56.26	159,01
ELEVATION	1102.72	2489,01
SLOPE	3.52	10,69

CLASS 2 Range-shrub	mean	variance
TM1	77.47	25,70
TM2	66.14	35,52
TM3	65.78	82,62
TM4	69.46	78,67
TM5	104.99	187,96
TM7	71.96	146,16
ELEVATION	1157.76	11406,24
SLOPE	18.11	52,99

CLASS 3 Range-herb.	mean	variance
TM1	84.02	26,11
TM2	73.21	34,10
TM3	77.29	84,45
TM4	77.22	44,75
TM5	127.34	92,54
TM7	90.02	79,38
ELEVATION	1183.38	5301,29
SLOPE	7.89	41,99

Table 4.10: Mean and variance values for ultimate T1 and T2 (overlap-excluded) (continued)

CLASS 4 Forest	mean	variance
TM1	67.01	8,82
TM2	51.64	16,48
TM3	45.14	29,92
TM4	57.72	34,57
TM5	62.28	131,79
TM7	39.46	72,59
ELEVATION	1141.30	5838,48
SLOPE	25.35	148,10

CLASS 4 Forest	mean	variance
TM1	66.95	9,48
TM2	51.43	17,05
TM3	44.86	30,80
TM4	57.27	30,36
TM5	61.34	132,48
TM7	38.82	74,47
ELEVATION	1128.03	5329
SLOPE	25.06	68,72

4.3.5. Phase 5: Classification

Classification is simply the procedure of assignment of each pixel within the image to a particular class. Maximum likelihood classification for the study follows the selection of training samples and generation of training statistics, and is followed by accuracy assessment.

Aim of the classification is to yield four pre-defined information classes which were; Agriculture, rangeland-shrub, rangeland-herb. and forest. Thus, other classes: Urban built-up land and water bodies were subject to being excluded from the classification procedure. The way to exclude mentioned classes from the classification is masking the regions associated with those classes from the analysis. Masking is generally known as visually excluding the regions of interest from the image or map. However, masking is offered as a utility of classification within the image processing system for excluding unwanted regions from the classification analysis. Classification procedure uses a binary raster where O indicates the pixels to be masked and 1 indicates pixels to be involved into classification. The mask raster was derived from the previously produced ground truth information data (Figure 4.9).

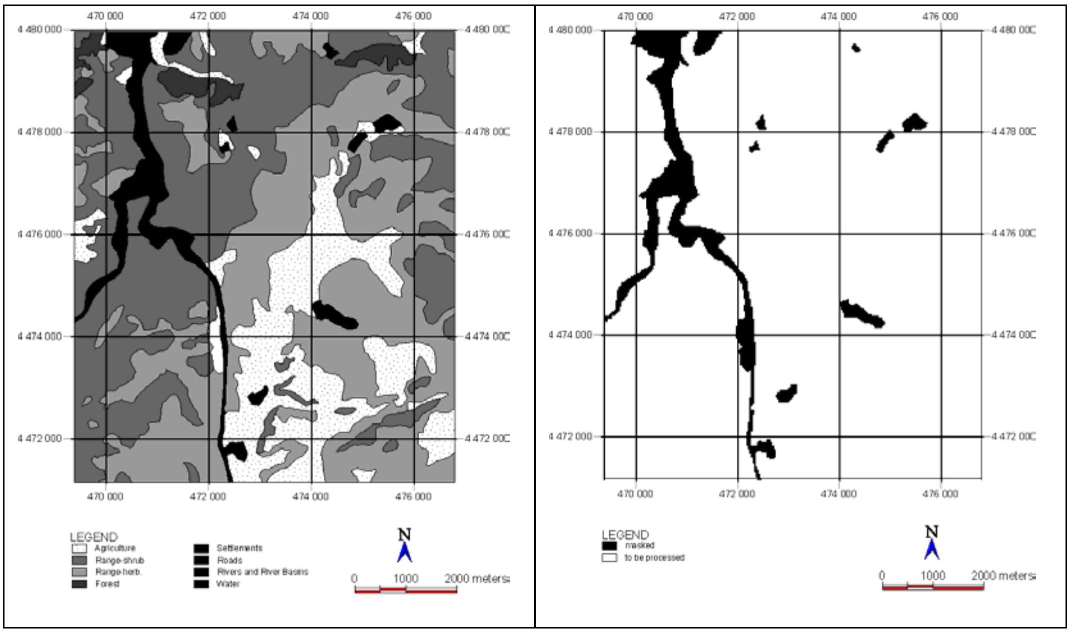


Figure 4.9: Ground Truth data and mask raster produced from ground truth data

Urban built-up land, water bodies and their subclasses were all assigned to zero. After the mask raster was introduced to classification procedure, Classification was performed. Maximum likelihood classification was performed for two times; first to yield Product 1 (Figure 4.10),

which is the result of classification of spectral data only by means of Training 1 (Training set for spectral data only), second to yield Product 2 (Figure 4.11), which was the result of classification of both spectral and topographical data by means of Training 2 (Training set for spectral and topographical data). Product 3 and product 4 was also produced to verify that the results of Product 1 and Product 2 were reasonable. Product 3 was the result of classification of spectral data only by means of Training2 and Product 4 was the result of classification of spectral and topographical data by means of Training1.

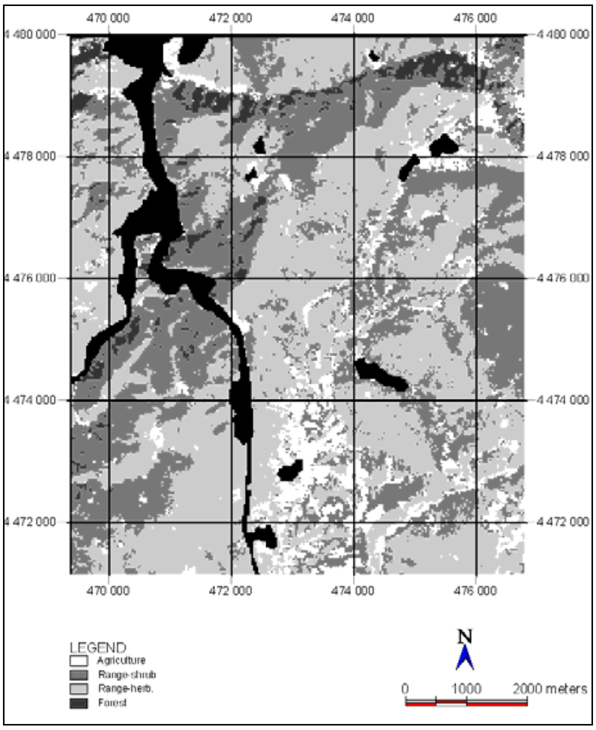


Figure 4.10: Product 1: Classification Product of bands used as input and trained by T1 (Training set for bands)

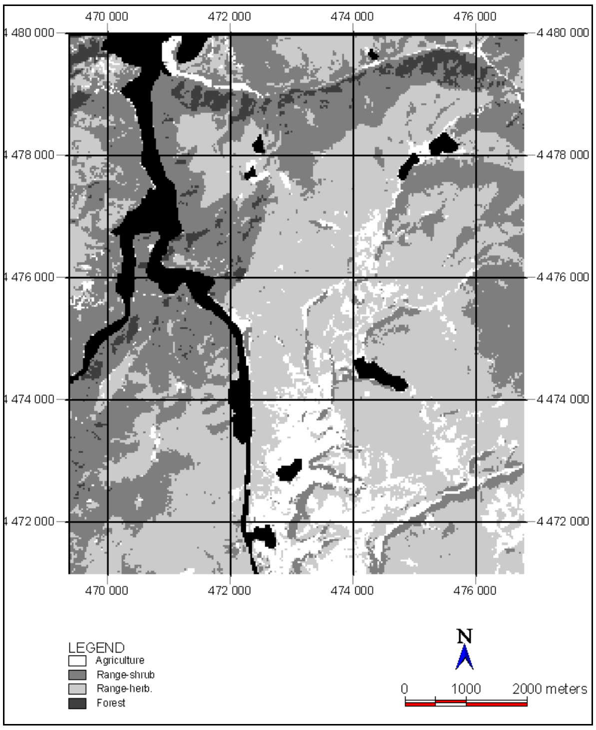


Figure 4.11: Product 2: Classification Product of bands, DTM and slope used as input and trained by T2 (Training set for bands, DTM and Slope)

There is a certain amount of difference between classification the product derived from spectral information and the product derived from both spectral and the topographical information. Distinction between the two products can even be visualized. Product 2 is smoother than Product 1. Land cover classes in Product 1 are rather mixed, whereas in Product 2 classes show homogeneity. However to understand the precise amount of disparity between the two products, and their association with the real world; statistically assessing the product data is needed.

4.4. Accuracy Assessment

Assessing the accuracy of classification product is of great importance to make use of the derived thematic map and /or associated statistics. What's more, classification is not regarded to be completed until its accuracy is assessed (Lillesand and Kiefer, 1996).

One of the most common methods for quantitatively assessing the classification accuracy is to make use of an error matrix. An error matrix determines the accuracy of a classification product based on comparison between classification product and the ground truth test information. In an error matrix, there are equal number of rows and columns. Rows normally represent the classification results from remotely sensed data where columns represent the reference data. Error matrix is an effective way to represent the accuracy of classification and it is very useful since it provides accuracy information for each class. Moreover it provides both inclusion (commission error) and exclusion (omission error). Commission error occurs when a pixel is identified as class A while in fact it is not. Omission error occurs when a point is identified as a member or another class while in fact it is class A.

Accuracy assessment is often performed through a group of sample from the product. Thus, prior to accuracy assessment sample size and sampling strategy must be determined. Widely used sample selection methods for classification products are random sampling and stratified random sampling (SRS) (Richards, 1996). In random sampling method, test samples are randomly distributed all over the area of interest, but this method may cause undersampling for classes that are relatively small but important as well. For the study SRS method was used, since it enables sufficient quantity of sample selection for forest agriculture classes whose size were relatively small compared to other two classes in the study area. The number of samples was offered as 3% of the total number pixels by Harris and Ventura (1995), 3% was adopted for the study which corresponds to 2178 observations approximately (Table 4.11).

Table 4.11: Number of Stratified Random Samples of Ground Truth for Accuracy Assessment

Class	Number of Samples
Agriculture (C1)	387
Range-shrub (C2)	864
Range-herb. (C3)	869
Forest (C4)	60
TOTAL	2180

Two products of the two classifications were tested with the ground truth sample data. Figure 4.12 is the error matrix for Product 1 where only bands are given as input and Training set for bands only was used to train classification. Figure 4.13 is for Product 2 where input data are; six Landsat TM bands, elevation and slope and the Training set is T1 that is prepared for bands, elevation and slope data.

Classif	Ground Truth Data						
	Name	G_1	G_2	G_3	G_4	Total	Accuracy
	1	120	41	34	3	198	60.61%
	2	97	514	125	11	747	68.81%
	3	170	265	710	3	1148	61.85%
	4	0	44	0	43	87	50.57%
	Total	387	864	869	60	2180	
	Accuracy	31.01%	64.58%	81.70%	0.00%		

Overall Accuracy = 63.67% Khat Statistic = 41.62%

Figure 4.12: Error matrix for Product 1

Classif	Ground Truth Data						
	Name	G_1	G_2	G_3	G_4	Total	Accuracy
	1	187	19	32	1	239	78.24%
	2	25	646	107	14	792	81.57%
	3	175	163	730	2	1070	68.22%
	4	0	36	0	43	79	54.43%
	Total	387	864	869	60	2180	
	Accuracy	48.32%	74.77%	84.00%	71.67%		

Overall Accuracy = 73.67% Khat Statistic = 58.85%

Figure 4.13: Error matrix for Product 2

Product 1 and product 2 are test products to better understand the effect of topography and classification of land cover classes. Product 3 and Product 4 are cross-products actually as mentioned before. Figure 4.14 is the error matrix for Product 3 where input data were bands, elevation and slope and training data is however Training set 1 prepared for band only. And likewise, Figure 4.15 is the error matrix for Product 4 where input data were bands only, and training data is however Training set 2 prepared for bands, elevation and slope.

Classif	Ground Truth Data						
	Name	G_1	G_2	G_3	G_4	Total	Accuracy
	1	127	35	33	2	197	64.47%
	2	91	520	117	11	739	70.37%
	3	169	265	719	4	1157	62.14%
	4	0	44	0	43	87	50.57%
	Total	387	864	869	60	2180	
	Accuracy	32.82%	65.28%	82.74%	0.00%		

Overall Accuracy = 64.68% Khat Statistic = 43.23%

Figure 4.14: Error matrix for Product 3

	Ground Truth Data						
Classif	Name	G_1	G_2	G_3	G_4	Total	Accuracy
	1	130	49	54	6	239	54.39%
	2	78	506	97	13	694	75.29%
	3	179	278	718	1	1176	62.22%
	4	0	31	0	40	71	54.79%
	Total	387	864	869	60	2180	
	Accuracy	33.59%	58.56%	82.62%	66.66%		
Overall Accuracy = 62.87% Khat Statistic = 43.65%							

Figure 4.15: Error matrix for Product 4

The Error Matrix shows two measures of accuracy for individual classes. The accuracy values for each column indicate the percentage of cells in that ground truth class that were correctly classified. Values less than 100% indicate errors of omission (ground truth cells omitted from the output class). This value is sometimes called the producer's accuracy. Conversely, the accuracy values for each row show the percentage of sample cells in each output class that were correctly classified. Values less than 100% indicate errors of commision (cells incorrectly included in the output class). This value is sometimes termed the user's accuracy

CHAPTER 5

DISCUSSION AND CONCLUSIONS

After the classification procedure two primary products which are P1 and P2, and two test products which are P3 and P4 were derived.

The common approach for assessing the success of this study would be to emphasize the increase of accuracy for P2 relative to P1. Admitting that this comparison gives the critical information about the success of the study, it is considered to be insufficient. Comparison of P1 and P2 with cross products as well was preferred instead. The four products derived after the classification procedure were P1, P2, P3, P4, where;

P1 input: **bands**
Training set: T1 (Training set for **bands** only)

P2 input **bands**, **elevation**, and **slope**
Training set: T2 (Training set for **bands**, **elevation**, and **slope**)

P3 input **bands**, **elevation**, and **slope**
Training set: T2 (Training set for **bands** only)

P4 input **bands**
Training set: T2 (Training set for **bands**, **elevation**, and **slope**)

P1 pertains to the product lacking topographical information, P2 is the result of classification with integrated topographic inputs and training signatures, P2 actually represents the ultimate product of the method. P3 and P4 are minor but essential test products.

Comparison of particular pairs among the four products is thought to supply very useful information about the success of the study. Pairs that were compared by means of accuracy measures are given sequentially.

P1-P2: Reasonable amount of accuracy was obtained in P2 compared to P1.The basis for this improvement is simply the integration of topographical ancillary information both as input and training information. This improvement under the scope of the study is the evidence of the notion that, use of topographical data as ancillary information in classification procedure improves accuracy.

P2-P3: Reasonable accuracy was obtained in P2, compared to P3. These two products both use spectral and topographical data as input, however, P3 which is derived by simple addition of ancillary data without any modification in the training set cannot compare to P2. Also see P1-P3 comparison.

P2-P4: P2 is inherently superior. P2 and P4 both use the same training set prepared for bands and elevation data (T2), however, input data for P4 do not involve topographical attributes. T2 is essentially prepared for use with band, elevation and slope data. Since the elevation and slope data is missing in the input set, T2 doesn't work.

P1-P3: Slight amount of improvement is observed in P3 compared to P1. Both products used training set for bands (T1), where P3 utilized additional elevation and slope data. Result is the evidence for simply adding ancillary data may improve the classification result, but to one step slightly beyond.

P1-P4: Both of the products seem equal in accuracy. It can hardly be affirmed that one is more accurate than the other because the indicators of accuracy are equivalent, anyway, although assumed it is, the improvement may be random as well. Both of these products used bands as input; however P4 is trained with T2, which is for band, elevation and slope. Actually, a better result that is obtained by another training set would be the evidence for the fact that T1 is not the best training set for classifying bands. And this would be an unfortunate for the study since; T1 for bands is principally invented to be the best set to train the input spectral data.

Comparison between P1 and P2 in detail is also essential. P2 accomplishes overall accuracy of 73,67%; 10% greater than P1. The improvement can be observed in each single class. Differences of accuracy for each class seems reasonable in relation to the over all improvement. If the classes were put into a descending order to comprehend relative improvement based on the integration of topography, the list would be as presented in Table 5.1.

Table 5.1: P2 Classes in descending order according to improvement in accuracy compared to P1

order	Class	P2 omission- P1 omission	P2 commission- P1 commission
1st	Agriculture (c1)	%17,31	%17,63
2nd	Range-shrub (c2)	%10,19	%12,76
3rd	Range-herb. (c3)	%2,30	%6,37
4th	Forest (c4)	%0,00	%3,86

Effectively, the same sequence can be obtained when the classes were put into descending order according to their magnitude of correlation with topographical data (Table 5.2).

Table 5.2: Classes in descending order according to their correlation with elevation and slope

order	class	Topographic Attribute	Correlation Coefficient
1st	Agriculture (c1)	Elevation	**+0,62**
		Slope	**+0,65**
2nd	Range-shrub (c2)	Elevation	**-0,34**
		Slope	**+0,48**
3rd	Range-herb. (c3)	Elevation	**-0,41**
		Slope	**+0,08**
4th	Forest (c4)	Elevation	**+0,1**
		Slope	**-0,5**

Identical sequence of classes listed according to improvement in accuracy and correlation with topography presents precious information that magnitude of class-topography correlation is highly related to the degree of accuracy of classes. This can yield the hypothesis; if the magnitude of correlation between class and topographical attributes is high, the product of classification with the integration of topographical data is more likely to be accurate.

When P2 is examined for class accuracy, degree of increase in accuracy for each class compared to P1 seems reasonable in relation to the correlation of each class with topographical data.

Something out of ordinary with the assessments for four products is that agriculture shows high omission error in all of the products. That is because the content of agriculture ground truth information is different than that of agriculture training signatures. Agriculture in ground truth information involves all of the agricultural land, where agricultural training signatures only represent vegetated agricultural land. As a consequence of this, classification normally ends up with smaller area off agriculture than the actual. The low accuracy is an issue on incorrespondance in class definition the ground truth information, rather than a misclassification problem.

In this study, a method that was primarily based on integrating topographical data into classification procedure as a component in addition to the spectral data, was presented. The results of the classification with the integration of ancillary topographical data verified that the method worked well and it provided a reasonable amount of improvement compared to classification based solely on remotely sensed spectral data.

Although the method presented successful results of improvement, these improvements were obtained under several conditions that may well have effect on the results. Thus, the method should be evaluated considering the cases where valid conditions change. Those conditions or parameters that may have effect on the success of the method are given:

The method with all rules and restraints kept constant may pose different amount of improvement, probably a loss of improvement in accuracy,

- when applied to a different site.

- when applied to a wide extent of area
- when applied to a data set including remotely sensed imagery of a different season.
- when applied to a data set including remotely sensed imagery of a different sensor system lacking some bandwidths present in Landsat TM imagery or having larger of bandwidths than the Landsat TM imagery.

- when applied to a data set including topographical attributes derived from elevation data with less amount of quality or quantity of information.
- when applied to a data set including remotely sensed imagery subject to substantial radiometric errors due to environmental influences.

If the method is proposed to be applied in the existence of any aforementioned case; definition of classes, class spectral and topographical signatures, investigation of correlation between classes and topographical components should be revised and reconstructed if needed for the current case.

The method, besides being successful for the case; also has some potential for further improvement and refinement. Points that are subject to improvement and refinement are basically some assumptions and generalizations admitted as drawbacks for the study. These drawbacks are grouped into subtitles as preprocessing, definition of class signatures, classification and accuracy assessment.

5.1. Size of the Unit Area

Size of the unit area which was defined as one pixel size (900 m^2) restricts any generalization on the final product. However this concept is not equivalent to the minimum mapping unit of a thematic map, which is directly related with the scale of map. Size of the minimum mapping unit is usually accepted as 5x5 mm on a thematic map. In CORINE program, 25 ha was set as the minimum mapping unit of 1/100 000 scaled thematic map (CORINE, 1993). This area was corresponding to 5x5 mm on the 1/100 000 scaled map. Accordingly, 900 m^2 unit area presents a scale of 1/6000. However, possible scale that can be derived from Landsat TM imagery is submitted as 1/20 000 to 1/55 000 (CORINE, 1993). Under these circumstances, size of the unit area should not be considered as minimum mapping unit of a thematic map, instead it should be recognized as a restriction to generalizations and that the final product is kept originally.

5.2. Preprocessing

The radiometric corrections for the study were regarded to be of little importance. The main reason for disregarding radiometric corrections is that those corrections are problematic of improving the quality of a specific product. The study concerns the approval of the effect of topography in image classification, which is presented by the comparison of two classified image. Radiometric correction may add to the raw image but there will be no relative improvement for the two products that are compared, as they will use the same corrected input. Hence during the progress for the study, the products derived from image classification one of which used raw image as input and the other which used the topographically corrected image as input was compared with the ground truth data and it was observed that the product in which the raw (uncorrected) image was used appeared to present better match with the ground truth information.

5.3. Redefinition of Class Signatures

Redefinition of Training sets mentioned in section 4.3.4 causes slight amount of changes in class spectral signatures (tables 4.8 and 4.10), even it was desired to keep class spectral signatures constant to test the effect of topography as a necessity of the study. However this change was assumed not to pose problems if the direction of the change for each class is the same. In other words the spectral signature mean values slightly shifting in the same direction for all classes with almost identical magnitudes do not end up with extraneous results compared to that derived using the original training set. For the study a relative change at amount of %1-2 on the positive direction was accepted reasonable.

In section 4.3.4 associated with redefinition of training sets, overlapping spectral regions were removed to end up with the ultimate training raster. Overlap extraction was actually a necessity of the system rather than the theoretical context because;

the classifier algorithm uses a raster training set where each training pixel may have only one attribute. This restricts overlap physically since, a pixel value cannot satisfy more than one class training criteria. Any way, exclusion of the overlapped part that concerns class 2 and 3 only, is not considered to pose a representativity problem. Because; pixels in the whole scene subject to the overlapped region can be assigned to one of class 2 or 3 according to its probability of being a member of one, within the capability of Maximum Likelihood classifier.

Actually, if the system could enable the use of raster independent spectral training set there would not be any physical restrictions similar to that in 2D array raster, and hence, there would be no need to make such modification in the training set, which is probably making a slight change in the output.

Redefined training sets are rather different than the standard training sets composed of samples that are only a very small proportion of the total number of pixels. However, number of samples redefined is of large amount compared to initial training set, because, redefined samples were collected by means of queries that select every single one of pixel that satisfy the criteria defined in the query. This is not commonplace for a standard training set. Nevertheless it was adopted for the study since a limit of maximum number sample size is not mentioned in the literature although there are suggestions about minimum sample size. Hence, increasing the training sample size yielded a training set more likely to be normal distribution compared to initial training set.

Another issue is selection of class topographical signatures. Those signatures were defined via a sampling strategy. The pixel values of topographical data which were corresponding to class spectral signatures were gathered, consequently the tails of the distributions were excluded based on the justification that those values may cause confusion in topographical training sets and they were finalized as class topographical signatures. Removal of tails makes sense since, by this way the uncertain values are excluded from the training sets and making them more representative. However the way that removal is performed is questionable, because where to cut the tail was supervisor- defined. The data range covering 1 or 2 standard deviation would be adopted easily if the distribution was normal, but the distribution for the topographical data, especially the slope is far from being normal.

A statistical approach that can handle the problem for such data should be considered as an enhancement of the method.

5.4. Classification

Maximum Likelihood being a per-pixel and parametric classifier also has some limitations. The per-pixel techniques of classification use the spectral information of each individual pixel to calculate its likelihood to a class regardless of the observation at neighboring pixels (Sharma and Sarkar, 1998), which is a shortcoming of per-pixel conventional techniques. These techniques do not take spatial position of image samples into account. They are commonly based on the implicit assumption that distribution of samples is random and each observation is independent. Unfortunately, this assumption violates one of the basic tenets of geography; the direct relationship between the distance and likeliness (Miller and Franklin, 2002). Vegetation types in an ecosystem in the real case have spatial dependence; means that, elements of an ecosystem close to one another are more likely to be similar because, they are influenced by the same generating process.

Even though conventional per-pixel classifiers are used extensively with fair amount of success, neglecting spatial dependence may introduce some restrictions (Sharma and Sarkar, 1998; Abkar et al., 2000). To solve this problem the two approaches aimed to introduce spatial dependence are; (1) using contextual information and (2) performing per-field classification techniques. However, in this study integrating ancillary topographical data into classification is maybe a further approach for reducing the effects of the problem due to the lack of spatial context. For instance, neighboring pixels may show divergence on account of probable error sources, and classification may end up with assignment of those pixels to other classes erroneously. Elevation or slope data, which are tended to be constant, or slightly changing for a particular location may force those pixels to be classified into the same class, even though the procedure is pixel based. The improvement in accuracy for P2 is mainly due to redefinition of training set so as to take topographical characteristics of the classes into account, and this was mentioned in detail in the former chapters. But, the contextual effect of the topographical data should not be disregarded; the effect, which causes the improvement mentioned in P1-P3 comparison. Classification of topographical data together with spectral bands as if they were bands, is not a very common application. The elevation and slope data that were used in the classification are inherently different types of data involving a context of information dissimilar to that of spectral bands. Elevation and slope data involve values of magnitude of change in elevation or slope, whereas bands involve the magnitude of spectral reflectance of features on Earth. Spectral bands are experimented to have certain amount of correlation among each other and likewise, topographical data also has an amount of correlation among each other. But the correlation between spectral and the topographical data is poor, and this is the main reason for rejecting the use of different type of data as input together with spectral data in a standard classification procedure. However, the correlation should not be searched within the whole range of these two types of data, because correlation can also be partial. And this correlation may pertain to an information class. The case was that for the study. To be more specific; not for the entire region but, for a certain land cover class, spectral and topographical data may be correlated, and this correlation can be obtained as it was done in section 4.3.2.

Another about per-pixel classifier is that it produces an output where full membership for each individual pixel is represented. This means that a pixel in the classification output strictly belongs to one class. However, full membership of the allocated class is often not the case for this study. For instance there may be some mixed pixels in the products dependent on the spatial resolution of the imagery and the distribution of the classes on the ground, where, most of the geographical phenomena do not exist in discrete classes but in inter-grading continua instead. Accordingly, in the zones where classes inter-grade, land cover tends to exhibit the characteristics of two classes. Classification product especially in inter-grading zones is not a good representation of the actual land cover for the case. Mapping probabilities of class membership may be an approach to represent continua. And for this purpose, probability raster for each classification products is produced in addition to the primary classification output which is hard. A probability raster presents the probability of any pixel for being a member of the class it was assigned. A probability raster does not show partial class membership, but at least it may give an idea about the pixels assigned with low probability which may be subject to a part of inter-grade in the continuum.

The other shortcoming of maximum likelihood classification is its being a parametric classifier, based on statistical parameters assuming that the training set has normal distribution. However the natural phenomenon is often unlikely to show normal distribution. Performing Maximum Likelihood classification on such data introduces an amount of error. For the study ML classifier was used anyway even though training data did not show normal distribution. However it should be admitted that redefinition of training sets in order to derive T1 and T2 had affirmative effects on the distribution of the data by making it closer to normal distribution compared to the initial training set (Ti).

5.5. Accuracy Assessment

The ground truth information was based on various data from different sources. Knowing that those data may not necessarily be produced to serve as ground truth information for particular applications, their use is not ideal. Those data may be inadequate in scale, may have unknown measure of accuracy and may not be up-to-date.

On the other hand, ground truth data was produced by means of visual interpretation, which may also bring out user error. In order to test the accuracy of classification product, perfect solution would be to collect field observation data via GPS form the field. Nevertheless, this solution may be the best but not the optimal. Thus, there is a limited amount of field observation for the study area, and the ground truth data was produced mainly dependent on reference maps and fine resolution image and aerial photograph. Under these circumstances, it’s fair to state that ground truth information might have admissible amount of error.

Ground truth data also has some ambiguities especially in inter-grading regions within the study area. Determining boundaries was a difficult task. Therefore there may be some conflicts in boundaries between two classes that are part of a continuum. When selecting random samples from this data, excluding points that are close to the boundaries is an approach for obtaining a more reliable test set. But it was not preferred for the study since it prevents testing of some amount of pixels that are near boundaries and would artificially yield improved results of accuracy where the fact would be rather different.

Ground truth data and the products have different size of the unit area, which indicates that they are not of the same scale. The spatial detail in ground truth information primarily related to available data used and interpreter skills is coarse compared to that of classification products where the size of the smallest unit is 900 square meters.

REFERENCES

1. Abkar A. A., Sharifi M. A. and Mulder N. J. “Likelihood-based Image Segmentation and Classification: A Framework for ther Integration of Expert Knowledge in Image Classification Procedures“ (2000) *International Journal of Earth Observation and Geoinformation* Vol.2, No. 2, pp. 104-119

2. Anderson J. R., Hardy E. E., Roach J. T. and Witmer R. E. (1976) “A Land Use and Land Cover Classification System for the Use with Remote Sensor Data“ Geological Survey Professional Paper, U.S. Government Printing Office, Washington D.C.

3. Ardiansyah P.O.D. and Yokoyama R. (2002) “DEM Generation Method from Contour Lines based on the Steepest Slope Segment Chain and a Monotone Interpolation Function” ISPRS Journal of Photogrammetry & Remote Sensing (in press)

4. Bolstad, P. V. and Lillesand T. M. (1992) “Rule-Based Classification Models: Flexible Integration of Satellite Imagery and Thematic Spatial Data” *Photogrammetric Engineering and Remote Sensing* Vol. 58, No. 7, pp. 965-971

5. Bruzzone L., Conese C., Maselli F. and Roli F. (1997) “Multisource Classification of Complex Rural Areas by Statistical and Neural-Network Approaches” *Photogrammetric Engineering and Remote Sensing* Vol. 63, No. 5, pp. 523-533

6. Burgess D. W., Lewis P. and Muller J. -P. A. L. (1995) “Topographic Effects in AVHRR Data“*Remote Sensing of Environment* Vol.54, pp. 223-232
7. Campbell J. B., 1996, Introduction to Remote Sensing. 2nd Edition. Taylor and Francis, London.

8. CORINE Technical Guide, 1993, European Commission, Office for Official Publications of the European Community, France

9. Deusen P. C. (1995) “Modified Highest Confidence First Classification“ *Photogrammetric Engineering and Remote Sensing* Vol. 61, No. 4, pp. 419-425

10. Eiumnoh A. and Shresta R. P. (1997) “Can DEM Enhance the Digital Image Classification?“ *Proceedings of Asian Conference on Remote Sensing*

11. Favela J. and Torres J. (1998) “A Two Step Approach to Satellite Image Classification Using Fuzzy Neural Networks and the ID3 Learning Algorithm“ *Expert Systems with Applications* Vol. 14, pp. 211-218

12. Fleeming M. D. and R. M. Hoffer (1979), Machine Processing of Landsat MSS Data and DTA Topographic Data for Forest Cover Type Mapping, Proceedings of the 1979, Symposium on Machine Processing of Remotely Sensed Data, Purdue University, pp. 377-390.

13. Fonseca L. M. G. and Manjunath B. S. (1996) “Registration Techniques for Multisensor remotely Sensed Imagery “*Photogrammetric Engineering and Remote Sensing* Vol. 62, No. 9, pp. 1049-1056

14. Foody G. M., (1992) “A Fuzzy Sets Approach to the Representation of Vegetation Continua from Remotely Sensed Data: An Example from Lowland Heath” *Photogrammetric Engineering and Remote Sensing* Vol. 58, No. 2, pp. 221-225

15. Foody G. M., Campbell N. A., Trodd N. M. and Wood T. F. (1992) "Derivation and Applications of Probabilistic Measures of Class Membership from the Maximum Likelihood Classification" *Photogrammetric Engineering and Remote Sensing* Vol. 58, No.9, pp. 1335-1341

16. Franklin S. E. and Giles P. T. (1995) "Radiometric Processing of Aerial and Satellite Remote Sensing Imagery" *Computers and Geosciences* Vol. 21, No. 3, pp. 413-423

17. Fraser R. S., Bahethi O. P., and Al-Abbas A. H. (1977) "The Effect of Atmosphere on the Classification of Satellite Observation to Identfy Surface Features" Remote Sensing of Environment Vol. 6, pp. 1649-1658

18. Gahegan M. and Flack J. (1996) "A Model to Support the Integration of Image Understanding Techniques within a GIS" *Photogrammetric Engineering and Remote Sensing* Vol. 62, No. 5, pp. 483-490

19. Gallant J. C and and Wilson J. P., (1996) " Tapes–G: A Grid-Based Terrain Analysis Program for the Environmental Sciences Computers and Geosciences Vol. 22, No. 7, pp. 713-722

20. Harris, P. M. and Ventura S. J. (1995), "The Integration of Geographic Data with Remotely Sensed Imagery to Improve Classification in an Urban Area" *Photogrammetric Engineering and Remote Sensing* Vol. 61, No. 8, pp. 993-998

21. Hutchinson, C. F. (1982) "Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement" *Photogrammetric Engineering and Remote Sensing* Vol. 48, No. 1, pp.123-130

22. Irvin B. J., Stephen J. V. and Slater B.K. (1997) “Fuzzy and ISODATA Classification of Land Form Elements from Digital Terrain Data in Pleasant Valley, Wisconsin” Geoderma Vol. 77, pp. 137-154

23. Jain A. K. (1989) “Knowledge-based Segmentation of Landsat Images“ IEEE Transactions on Geoscience and Remote Sensing, Vol.29, pp. 222-232

24. Janssen L. L. F., Jaarsma M. N. and Linden E. T. M. (1990) “Integrating Topographic Data with Remote Sensing for Land Cover Classification” *Photogrammetric Engineering and Remote Sensing* Vol. 56, No. 11, pp. 1503-1506

25. Jensen J. R. (1996). Introductory Digital Image Processing A Remote Sensing Perspective 2nd edition, Upper Saddle River, Pretince Hall, New Jersey

26. Kardoulas N. G., Bird A.C. and Lawran A. I. (1999) “Geometric Correction of SPOT and Landsat Imagery: A Comparison of Map and GPS Derived Control Points“ *Photogrammetric Engineering and Remote Sensing* Vol. 62, No. 10, pp. 1173-1177

27. Lillesand, R. M. and R. M. Kiefer, 1994,. Remote Sensing and Image Interpretation, 3rd edition, John Wiley and Sons, New York.

28. Lunetta R. S., Congalton R. G., Fenstermaker L. K., Jensen J. R., McGwire K. C. and Tinney L. R. (1991) “Remote Sensing and Geographical Information System Data Integration: Error Sources and Research issues (1991) *Photogrammetric Engineering and Remote Sensing* Vol. 57, No. 6, pp. 677-687

29. Maselli F., Conese C., Filippis T. D. and Romani M. (1995) “Integration of Ancillary Data into a Maximum Likelihood Classifier with Nonparametric

Priors” *ISPRS Journal of Photogrammetry and Remote Sensing* Vol. 50, No. 2, pp. 2-11

30. Maselli F., Conese C., Petkov L. and Resti Raffaello (1992) “Inclusion of Prior Probabilities Derived from a Nonparametric Process into the Maximum Likelihood Classifier“ *Photogrammetric Engineering and Remote Sensing* Vol. 58, No. 2, pp. 201-207

31. Mather, P.,1989, Computer Processing of Remotely Sensed Images: An Introduction, John Wiley and Sons, Great Britain.

32. Mather P. M., (1999), “Land Cover Classification Revisited” in: P. M. Atkinson and N. J. Tate (eds.) Advances in Remote Sensing and GIS Analysis, 2000, John Wiley and Sons, New York.

33. Mesev , (1998), “The Use of Census data in Urban Image Classification” *Photogrammetric Engineering and Remote Sensing* Vol. 64, No. 5, pp. 431-438

34. Miller J. and Franklin J. (2002) “Modeling the Distribution of four Vegetation Alliances Using Generalized Linear Models and Classification trees with Spatial Dependence“ *Ecological Modeling* Vol. 157 pp.227-247

35. Mizukoshi H. and Ania M. (2002) “Use of Contour-Based DEMs for Deriving and Mapping Topographic Attributes“ *Photogrammetric Engineering and Remote Sensing* Vol. 68, No. 1, pp. 83-93

36. Novak K., (1992) “Rectification of Digital Imagery, *Photogrammetric Engineering and Remote Sensing* Vol. 58, No.3, pp. 339-344

37. Richards J. A., (1996) “Classifier Performance and Map Accuracy“ *Remote Sensing of Environment* Vol. 57, pp. 161-166

38. Sabins F., 1977, Remote Sensing Principles and Interpretation. W. H. Freeman & Co. New York, USA.

39. Sharma K. M. S. and Sarkar A. (1998) “A Modified Contextual Classification Tehnique for Remote Sensing Data“ (1998) *Photogrammetric Engineering and Remote Sensing* Vol. 64, No. 4, pp. 273-280

40. Shrestha D. P. and Zinck J. A. (2001) “Land Use Classification in Mountainous Areas:Integration of Image Processing, Digital Elevation Data and Field Knowledge (Application to Nepal)” *International Journal of Earth Observation and Geoinformation* Vol. 3, No. 1, pp. 78-85

41. Song C., Woodcock C., Seto K. C., Lenney M. P. and Macomber S. A. (2001) “Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects?“ *Remote Sensing of Environment* Vol.75, pp. 230-244

42. Strahler A. H., T. L.Logan and N. A. Bryant, (1978), Improving Forest Cover Classification Accuracy from Landsat by Incorporating Topographic Information, Proceedings of the Twelfth International Symposium on remote Sensing of Environment, Environmental Research Institute of Michigan, pp. 927-942.

43. Strahler A. H., (1980) “The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely Sensed Data, *Remote Sensing of Environment* Vol.10 pp. 135-163

44. Swain P. H. and Davis S. M. (1978) Remote Sensing: Quantitative Approach McGraw Hill Book Company, NewYork.

45. Vogelmann J. E., (1998) “Regional Characterization of Land cover Using Multiple Sources of Data” *Photogrammetric Engineering and Remote Sensing* Vol. 64, No. 1, pp. 45-57

46. Westmoreland S. and Stow D. A. (1992) "Category Identification of Changed Land-Use Polygons in an Integrated Image Processing / Geographic Information System" *Photogrammetric Engineering and Remote Sensing* Vol. 58, No. 11, pp. 1593-1599

47. CCRS, Canada Center for Remote Sensing, http://www.ccrs.nrcan.gc.ca/ccrs/learn/tutorials/fundam/chapter4/chapter4_4_e.html , Section 4.4 Pre-Processing, 2002

48. Devlet Meteoroloji İşleri Genel Müdürlüğü, http://www.meteor.gov.tr/webler/iklim/turkiyeiklimi.htm, İklim Nedir, 2002

49. HGK, http://www.hgk.mil.tr/, HGK, 2002

50. Microimages, http://www.microimages.com/refman/, Online Reference Manual, 2002

51. RSI, Research Systems Inc. A Kodak Company, http://www.rsinc.com/services/output.cfm?tip_id=2192, Tech Tip: Kriging Description, 2002

52. Statsoft, http://www.statsoftinc.com/textbook/stathome.html, The Statistics Homepage, 2002

53. USGS, http://landcover.usgs.gov, EROS Data Center, Sioux Falls, SD, 2002

54. USGS,
http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/dem/3DEM0897.PDF

55. USGS,
http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/dem/2DEM0198.PDF
http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/dem/2DEM0198.PDF, 2002

56. Vassar, http://faculty.vassar.edu/lowry/pbcorr.html, Point Biserial Correlation Coefficient, 2002

57. Wiseman, http://wiseman.brandonu.ca/article1.htm, Article 1, 2002

APPENDIX A

MINOR DATA USED IN THE STUDY

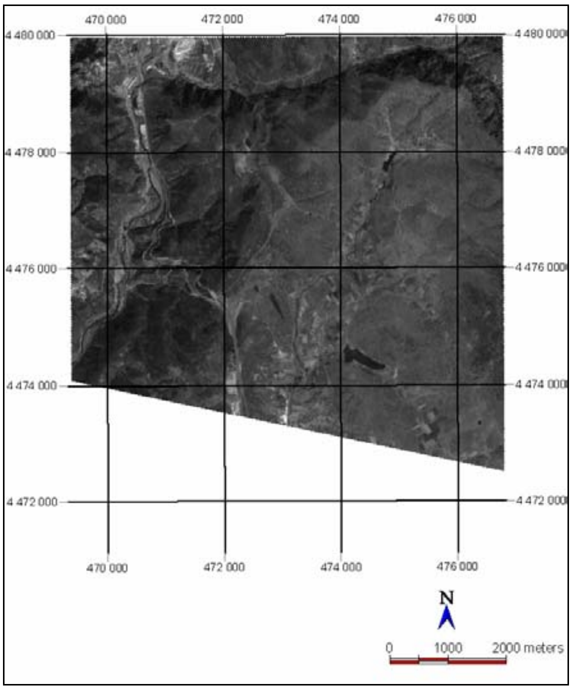


Figure A.1: IRS panchromatic image from 1c sensor

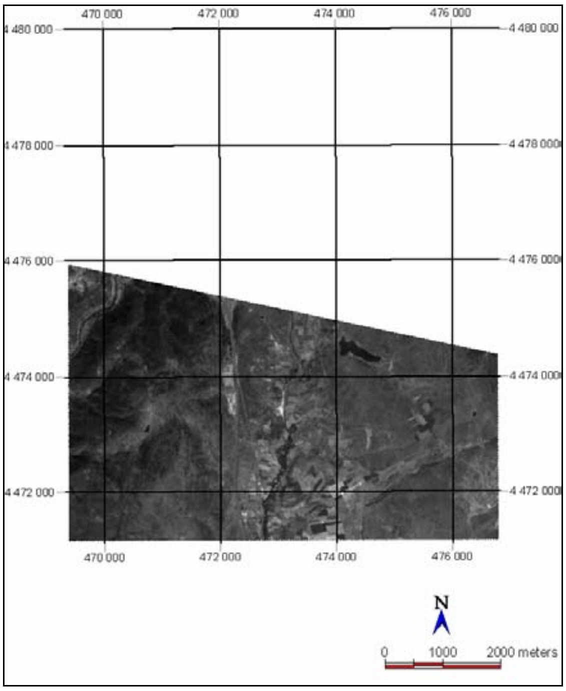


Figure A.2: IRS panchromatic image from 1c sensor

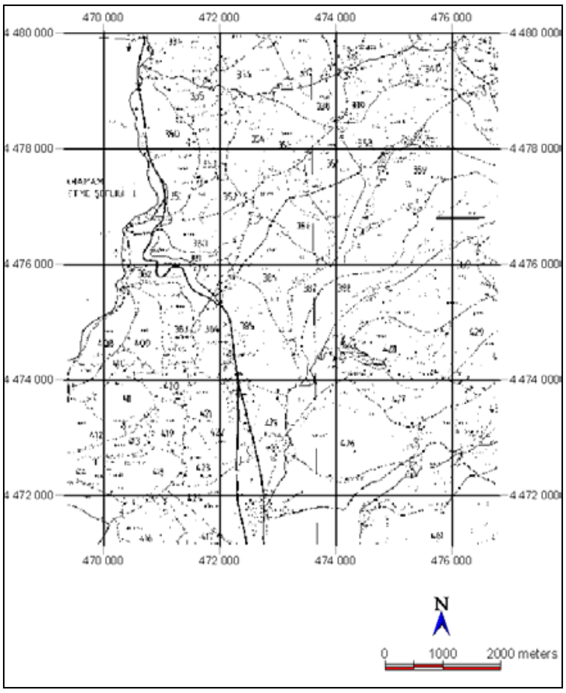


Figure A.3: Forest map: Yıldırım-2 region

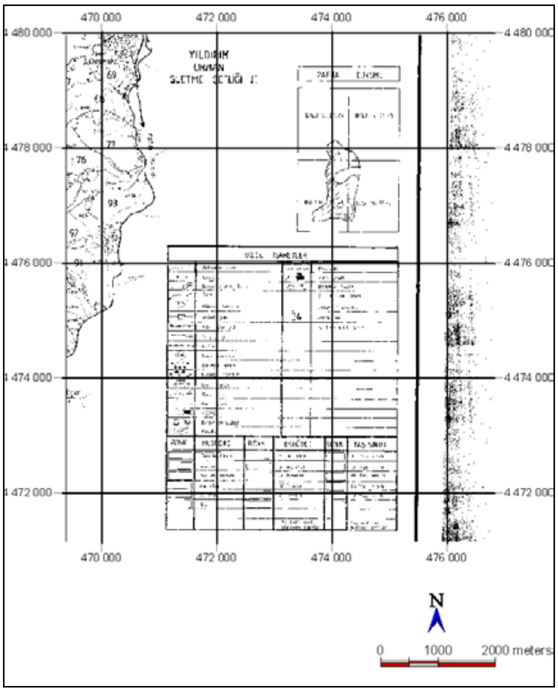


Figure A.4: Forest map: Kızılcahamam-1 region

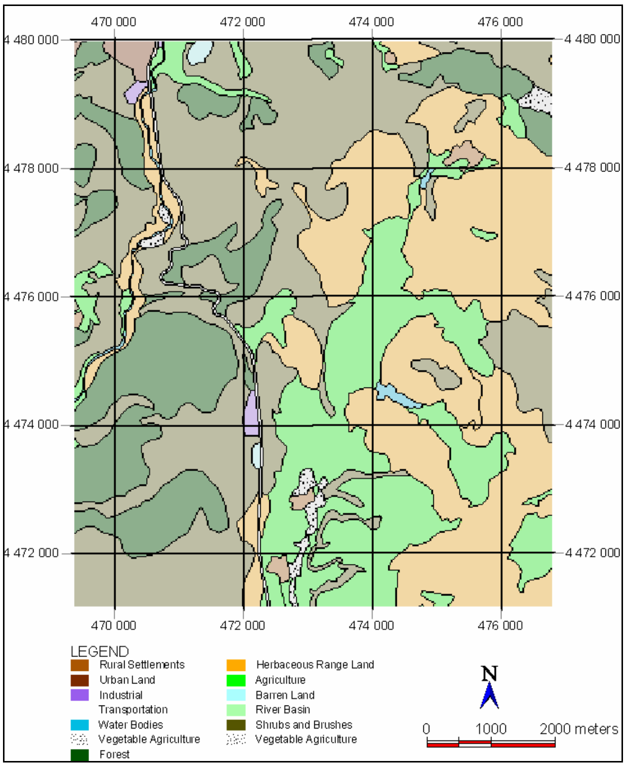


Figure A.5: Digital land use and land cover map

APPENDIX B

GCPs FOR THE STUDY AREA

Table B.1: 1/25000 scaled Topographical Map GCPs

#GCP	Column	Line	North(m)	East(m)	Residual(m)
1	4292.06	251.1	4483000	475000	0.985
2	4918.52	4017.58	4477000	476000	1.042
3	525.04	5272.8	4475000	469000	0.733
4	3036.07	5273.85	4475000	473000	1.077
5	4291.1	5274.12	4475000	475000	1.585
6	3663.56	6528.09	4473000	474000	0.837
7	4290.97	7784.05	4471000	475000	0.618
8	525.95	8411.86	4470000	469000	1.143
RMSE					**1,00**

Table B.2: Landsat Panchromatic band GCPs

#GCP	Column	Line	North(m)	East(m)	Residual(m)
1	1837.87	1139.59	4481486	473366.8	6.15
2	2160.53	1058.13	4482684	478205.4	10.73
3	1857.1	1217.34	4480327	473665.1	16.30
4	1502.87	1387.3	4477774	468362.6	13.37
5	1934.82	1396.59	4477632	474802.8	17.65
6	1722.24	1507.26	4475981	471634.8	10.02
7	1882.22	1597.27	4474616	474032.6	3.10
8	2014.32	1740.08	4472478	476001.9	7.55
9	1889.35	1788.51	4471750	474117.2	18.88
10	1559.96	1673.06	4473473	469192.2	18.83
11	2150.22	1760.77	4472157	478043.4	3.48
12	2051.58	1964.63	4469104	476578.9	13.98
13	1796.22	1783.81	4471822	472740	1.42
14	1606.98	1254.2	4479757	469914.1	11.33
15	1825.49	1426.41	4477188	473192.6	13.73
16	1379.19	1800.03	4471586	466494	3.83
17	1680.37	1915.3	4469865	471015.9	15.20
18	1610.57	1055.48	4482739	469952.3	13.58
19	2148.99	1186.72	4480760	478034.2	8.94
20	2038.27	1683.28	4473312	476372.7	13.74
21	1997.35	1455.13	4476752	475737.6	18.56
22	1784.55	1293.02	4479188	472569.2	7.35
RMSE					**12.45**

Table B.3: Landsat TM 1, 2, 3, 4, 5, 7 bands' GCPs

#GCP	Column	Line	North(m)	East(m)	Residual(m)
1	918.93	569.79	4481486	473366.8	6.18
2	1080.27	529.07	4482684	478205.4	10.74
3	928.55	608.67	4480327	473665.1	16.40
4	751.43	693.65	4477774	468362.6	13.43
5	967.41	698.29	4477632	474802.8	17.72
6	861.12	753.63	4475981	471634.8	9.95
7	941.11	798.63	4474616	474032.6	3.05
8	1007.16	870.04	4472478	476001.9	7.58
9	944.67	894.26	4471750	474117.2	18.86
10	779.98	836.53	4473473	469192.2	18.86
11	1075.11	880.38	4472157	478043.4	3.51
12	1025.79	982.32	4469104	476578.9	13.96
13	898.11	891.9	4471822	472740	1.35
14	803.49	627.1	4479757	469914.1	11.37
15	912.75	713.21	4477188	473192.6	13.75
16	689.6	900.01	4471586	466494	3.82
17	840.18	957.65	4469865	471015.9	15.19
18	805.28	527.74	4482739	469952.3	13.59
19	1074.49	593.36	4480760	478034.2	8.95
20	1019.14	841.64	4473312	476372.7	13.79
21	998.67	727.57	4476752	475737.6	18.54
22	892.27	646.51	4479188	472569.2	7.29
RMSE					**12.47**

Table B.4: IRS B1E15A6D Panchromatic band GCPs according to 2nd order Transformation

	#GCP	Column	Line	North(m)	East(m)	Residual(m)
1	1	6031.12	12931.87	4480213	470217.3	2.58
2	2	6780.7	1664.01	4534556	485507.9	6.71
3	3	7174.9	7007.08	4508015	481917	4.82
4	4	1691.72	7926.81	4509187	454184.7	5.07
5	5	1324.58	5259.31	4522610	455158.1	8.00
6	6	10135.49	4993.19	4514803	498455.6	6.38
7	7	7401.25	14135.23	4472907	475669.5	4.42
8	8	13943	13882.3	4467376	507866	2.61
9	9	4703.2	5582.97	4517530	471319.4	5.45
10	10	7683.76	3823.52	4523061	487688.5	4.38
11	11	14075.26	9930.92	4486563	512583.2	5.28
12	12	921.98	1080.98	4543474	457512.1	6.48
13	14	2951.83	9181.12	4501753	459047.7	6.67
14	15	3228.46	11179.84	4491677	458341	6.70
15	16	13409.06	1868.52	4526689	517670.9	2.99
16	17	8416.7	4786.59	4517588	490280.1	5.93
17	19	9892.05	9391.76	4493532	492713.3	7.47
18	20	8434.99	6920.09	4507130	488161.3	3.27
19	22	4811.07	8863.12	4501373	468454.5	5.48
20	23	3480.56	13171.98	4481680	457508.2	6.68
21	24	3237.48	7033.67	4511960	462656.5	6.08
22	25	8216.32	1593.29	4533418	492597.1	3.10
23	26	12200.75	9985.86	4488239	503373.8	6.28
24	27	12628.96	12854.8	4473759	502511.4	1.68

Table B.4: IRS B1E15A6D Panchromatic band GCPs according to 2nd order Transformation (continued)

25	28	9900.83	12523.18	4478200	489533.3	6.32
26	29	12698.33	7401.14	4500366	508475.3	3.02
27	30	2879.7	3817.56	4528064	464229.9	8.81
28	31	9844.52	1504.16	4532164	500637	4.05
29	32	329.91	14073.82	4480523	441191.8	6.06
RMSE						**5.55**

Table B.5: IRS D1E15A6D Panchromatic band GCPs according to 2nd order Transformation

	#GCP	Column	Line	North(m)	East(m)	Residual(m)
1	1	12704.44	1762.79	4461142	496666.5	7.86
2	2	10343.04	10028.85	4423159	476571.9	4.81
3	4	5577.98	5271.74	4451369	458222.4	8.43
4	5	14439.96	2953.67	4453511	503919.8	6.31
5	6	756.86	8036.12	4442859	431816.9	1.36
6	7	1448.97	5795.01	4453097	437520.2	5.60
7	8	11488.34	6917.04	4437189	485386.6	7.13
8	9	9838.48	3868.07	4453813	480491.8	2.23
9	10	9027.05	3211.66	4457863	477210.6	3.49
10	11	4026.83	8989.53	4434805	446802.5	6.35
11	12	1271.45	9872.29	4433342	432431.6	4.76
12	14	7534.76	10260.26	4424937	462619.2	2.60
13	15	8432.2	13751.04	4406936	463379.5	6.56
14	16	12153.3	10233.46	4420278	485208.3	4.01
15	17	11424.22	9712.97	4423583	482179.5	5.86
16	18	12404.34	13247.84	4405267	483307.4	4.51
17	19	14198.94	10248.55	4418076	495179	3.60
18	20	7705.52	1148.55	4469329	472891.7	1.75
19	21	3146.91	10004.05	4430749	441456.4	4.36
20	22	8539.28	4637.4	4451399	473359.6	8.08
21	23	3055.14	13863.82	4411961	437005.3	5.33
22	24	317.44	13579.28	4416203	423925.1	1.20
23	26	5149.38	12071.42	4418561	449093.9	4.27
24	28	1632.35	1605.07	4473409	442748.2	6.68
25	29	12902.15	11292.02	4414321	487768.9	3.66
26	30	12871.04	3385.96	4453023	495810.6	6.65
27	31	6450.38	7699.07	4438586	459975.9	7.48
28	32	8006.34	8069.52	4435168	467194.5	2.91
29	33	8448.71	3290.44	4458088	474301.2	6.77
30	34	4012.75	7727.68	4440983	448044.6	4.83
31	35	6387.49	3664.8	4458388	463852.2	4.92
32	38	13068.17	5384.18	4443040	494692.5	7.04
33	39	10033.96	1783.02	4463807	483603	5.47
34	40	10134.69	12027.7	4413594	473492.1	5.48
RMSE						**4.85**

APPENDIX C

KRIGING PARAMETERS FOR DTM GENERATION

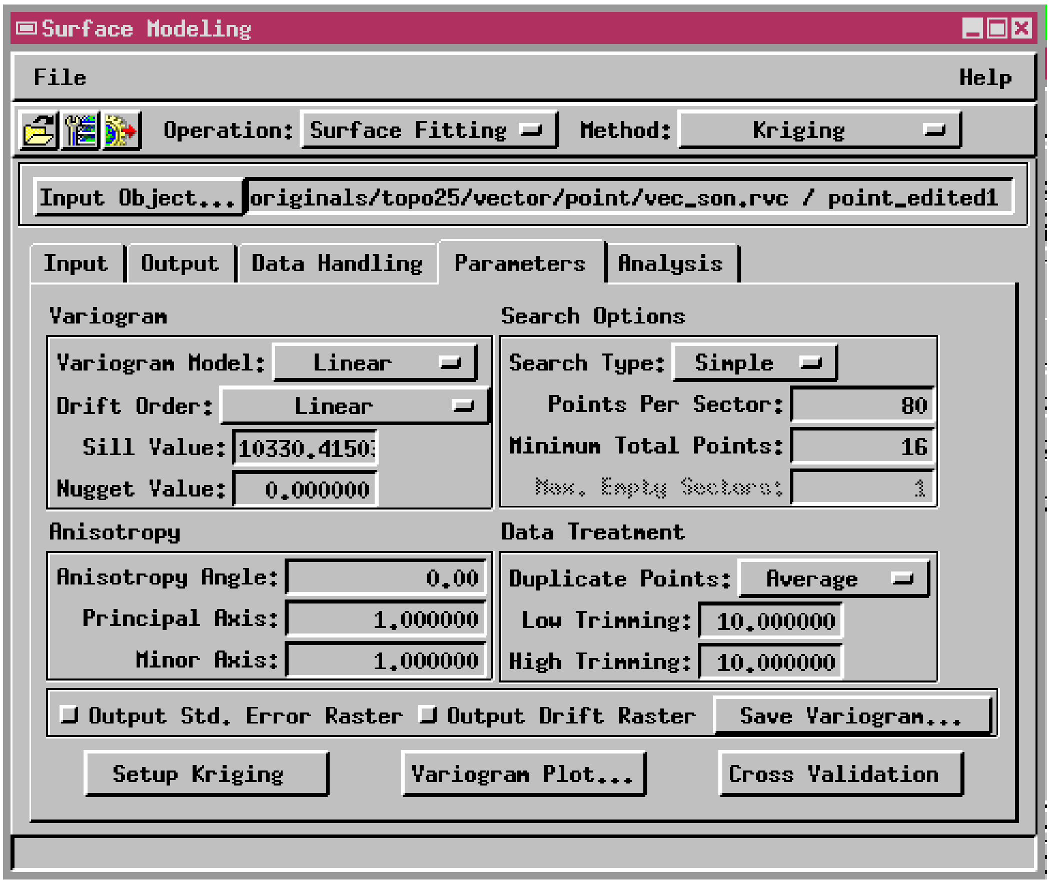


Figure C.1: Kriging Parameters

Input: Elevation points

Variogram Model: Linear
Drift order: Linear
Sill value: 10330 (unadjusted)
Nugget value: 0.0 (unadjusted)

Search type: Simple
Points per Sector: 80
Minimum total points: 16

Anisotropy: unadjusted
Duplicate points: unadjusted

APPENDIX D

FEATURE DOMAIN STRUCTURES

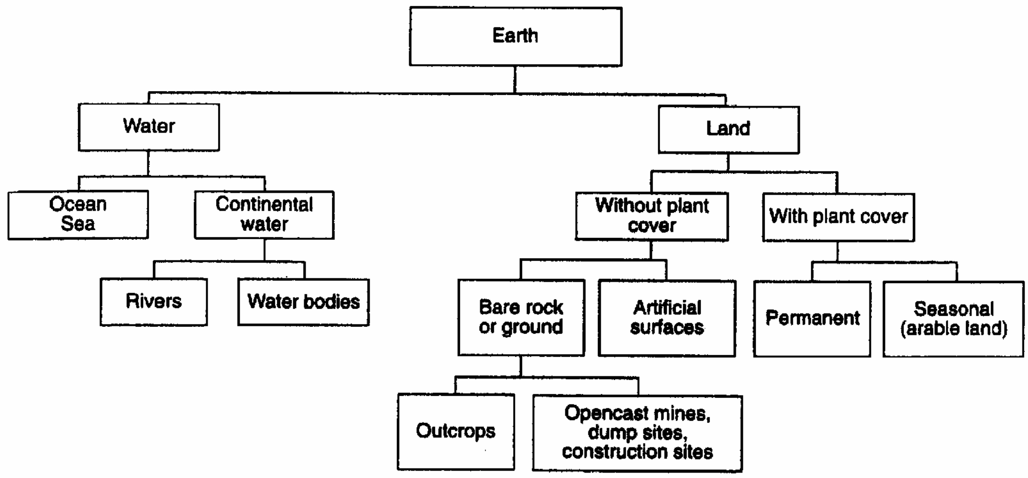


Figure D.1: Hierarchical Tree Structure of Feature Domain (Jain, 1989)

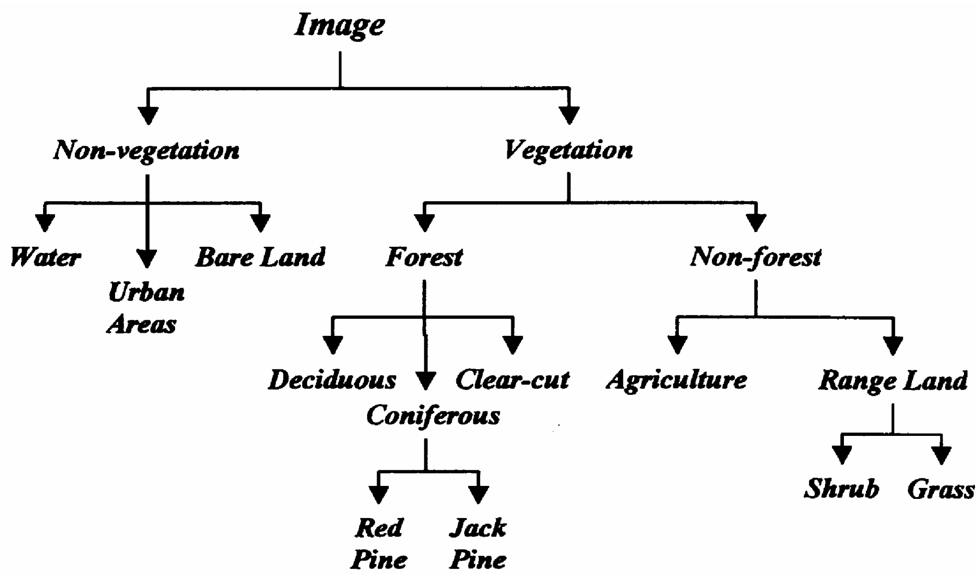


Figure D.2: Theoretical Schematic Construction of Land Cover Nomenclature (CORINE, 1993)

APPENDIX E

CLASSIFICATION SCHEMES

Table E.1: Anderson Level I/II Classification Scheme

Level I	Level II
1. Urban or Built-up Land	1.1. Residential
	1.2. Commercial and Services
	1.3. Industrial
	1.4. Transportation, Communication and Utilities
	1.5. Industrial and Commercial Complexes
	1.6. Mixed Urban or Built-up Land
	1.7. Other Urban or Built-up Land
2. Agricultural land	2.1. Cropland and Pasture
	2.2. Orchards, Groves, Vineyards, Nurseries and Ornamental Horticultural Areas
	2.3. Confined feeding Operations
	2.4. Other Agricultural Land
3. Rangeland	3.1. Herbaceous land
	3.2. Shrub and Brush Land
	3.3. Mixed Rangeland
4. Forest	4.1. Deciduous Forest Land
	4.2. Evergreen Forest land
	4.3. Mixed Forest Land
5. Water	5.1. Streams and Canals
	5.2. Lakes
	5.3. Reservoirs
	5.4. Bays and Estuaries
6. Wetland	6.1. Forested Wetland
	6.2. Non-forested Wetland
	9.2. Glaciers

Table E.1: Anderson Level I/II Classification Scheme (continued)

7. Barren land	7.1. Dry Salt Flats
	7.2. Beaches
	7.3. Sandy Areas other than Beaches
	7.4. Bare Exposed Rock
	7.5. Strip Mines, Quarries and Gravel Pits
	7.6. Transitional Areas
	7.7. Mixed Barren Land
8. Tundra	8.1. Shrub and Brush Tundra
	8.2. Herbaceous Tundra
	8.3. Bare Ground Tundra
	8.4. Wet Tundra
	8.5. Mixed Tundra
9. Perennial Snow or Ice	9.1. Perennial Snowfields
	9.2. Glaciers

Table E.2: The CORINE Land Cover Classification Scheme

Level I	Level II	Level III
1. Artificial Surfaces	1.1. Urban Fabric	1.1.1 Continuous Urban Fabric
		1.1.2. Discontinuous Urban Fabric
	1.2. Industrial, commercial and transportation units	1.2.1. Industrial and Commercial Units
		1.2.2. Road and Rail networks and Associated Land
		1.2.3. Port Areas
		1.2.4. Airports
	1.3. Mine, Dump and Construction Sites	1.3.1. Mineral Extraction Sites
		1.3.2. Dump Sites
		1.3.3. Construction Sites
	1.4. Artificial, Non-Agricultural Vegetated Areas	1.4.1. Green Urban Areas
		1.4.2. Port and Leisure Facilities
2. Agricultural Areas	2.1. Arable Land	2.1.1. Non-Irrigated Arable Land
		2.1.2. Permanently Irrigated Land
		2.1.3. Rice Fields

Table E.2: The CORINE Land Cover Classification Scheme (continued)

2. Agricultural Areas	2.2. Permanent Crops	2.2.1. Vineyards
		2.2.2. Fruit Trees and Berry Plantations
		2.2.3. Olive Groves
	2.3. Pastures	2.3.1. Pastures
	2.4. Heterogeneous Agricultural Areas	2.4.1. Annual Crops Associated with Permanent Crops
		2.4.2. Complex Cultivation Patterns
		2.4.3. Land Principally Occupied by Agriculture with Significant areas of Natural Vegetation
		2.4.4. Agro-Forestry Areas
3. Forests and Semi-Natural Areas	3.1. Forests	3.1.1. Broad-Leaved Forest
		3.1.2. Coniferous Forest
		3.1.3. Mixed Forest
	3.2. Shrub and/or Herbaceous Vegetation Associations	3.2.1. Natural grassland
		3.2.2. Moors and Heathland
		3.2.3. Sclerophyllous Vegetation
		3.2.4. Transitional Woodland Shrub
	3.3. Open Spaces with Little or no Vegetation	3.3.1. Beaches, Dunes and sands
		3.3.2. Bare Rocks
		3.3.3. Sparsely Vegetated Areas
		3.3.4. Burnt Areas
		3.3.5. Glaciers and Perpetual Snow
4. Wetlands	4.1. Inland Wetlands	4.1.1. Inland Marshes
		4.1.2. Peat Bogs
	4.2. Maritime Wetlands	4.2.1. Salt Marshes
		4.2.2. Salines
		4.2.3. Interdinal Flats
5. Water Bodies	5.1. Inland Waters	5.1.1. Water Courses
		5.1.2. Water Bodies
	5.2.Marine waters	5.2.1. Coastal Lagoons
		5.2.2. Estuaries
		5.2.3. Sea and Ocean

APPENDIX F

PHOTOGRAPHS FROM THE STUDY AREA

Figure F.1: Agricultural Land in the Study Area

Figure F.2: Rangeland (Range-shrub) in the Study Area

Figure F.3: Rangeland (Range-herb.) in the Study Area

Figure F.4: Rural Settlement in the Study Area

APPENDIX G

POINT BISERIAL CORRELATION COEFFICIENT OUTPUTS

Data Summary	X=0	X=1	Total
n	174	269	443
ΣY	190077	323522	513599
ΣY2	207928715	390898178	598826893
SS_Y	289255.637	1803440.14	3377834.75
$mean_Y$	1092.3966	1202.684	1159.3657

r_{pb}	t	df
+0.62	+16.46	441

P	
one-tailed	<.0001
two-tailed	<.0001

Figure G1: test between agriculture – elevation

Data Summary	X=0	X=1	Total
n	174	269	443
ΣY	515	3811	4326
ΣY2	2299	71599	73898
SS_Y	774.7184	17607.4721	31653.5847
$mean_Y$	2.9598	14.1673	9.7652

r_{pb}	t	df
+0.65	+17.84	441

P	
one-tailed	<.0001
two-tailed	<.0001

Figure G2: test between agriculture – slope

Data Summary	X=0	X=1	Total
n	174	269	443
ΣY	31656	50004	81660
ΣY2	7586098	12121950	19708048
SS_Y	1826889.17	2826782.65	4655326.55
$mean_Y$	181.931	185.8885	184.3341

r_{pb}	t	df
+0.02	+0.4	441

P	
one-tailed	0.344675
two-tailed	0.689350

Figure G3: test between agriculture – aspect

Data Summary	X=0	X=1	Total
n	116	327	443
ΣY	140201	373398	513599
ΣY2	170184395	428642498	598826893
SS_Y	733357.060	2263090.03	3377834.75
$mean_Y$	1208.6293	1141.8899	1159.3657

r_{pb}	t	df
-0.34	-7.49	441

P	
one-tailed	<.0001
two-tailed	<.0001

Figure G4: test between range-shrub– elevation

Data Summary	X=0	X=1	Total
n	327	116	443
ΣY	2409	1917	4326
ΣY^2	38577	35321	73898
SS_Y	20829.9633	3640.9224	31653.5847
$mean_Y$	7.367	16.5259	9.7652

r_{pb}	t	df
+0.48	+11.38	441

P		
	one-tailed	<.0001
	two-tailed	<.0001

Figure G5: test between range-shrub–slope

Data Summary	X=0	X=1	Total
n	121	322	443
ΣY	1092	3234	4326
ΣY^2	14788	59110	73898
SS_Y	4932.9256	26629.3913	31653.5847
$mean_Y$	9.0248	10.0435	9.7652

r_{pb}	t	df
+0.08	+1.13	441

P		
	one-tailed	0.129545
	two-tailed	0.259090

Figure G8:test between range-herb.–slope

Data Summary	X=0	X=1	Total
n	116	327	443
ΣY	25807	55853	81660
ΣY^2	7086379	12621669	19708048
SS_Y	1344988.92	3081737.47	4655326.55
$mean_Y$	222.4741	170.8043	184.3341

r_{pb}	t	df
-0.11	-4.77	441

P		
	one-tailed	<.0001
	two-tailed	<.0001

Figure G6: test between range-shrub–aspect

Data Summary	X=0	X=1	Total
n	121	322	443
ΣY	21064	60596	81660
ΣY^2	4213542	15494506	19708048
SS_Y	546665.173	4091166.81	4655326.55
$mean_Y$	174.0826	188.1863	184.3341

r_{pb}	t	df
+0.06	+1.29	441

P		
	one-tailed	0.0988635
	two-tailed	0.197727

Figure G9: test between range-herb.–aspect

Data Summary	X=0	X=1	Total
n	121	322	443
ΣY	147271	366328	513599
ΣY^2	179988819	418838074	598826893
SS_Y	742972.380	2079677.77	3377834.75
$mean_Y$	1217.1157	1137.6646	1159.3657

r_{pb}	t	df
-0.41	-9.31	441

P		
	one-tailed	<.0001
	two-tailed	<.0001

Figure G7: test between range-herb.–elevation

Data Summary	X=0	X=1	Total
n	32	411	443
ΣY	36050	477549	513599
ΣY^2	40724964	558101929	598826893
SS_Y	112385.875	3228334.35	3377834.75
$mean_Y$	1126.5625	1161.9197	1159.3657

r_{pb}	t	df
+0.1	+2.21	441

P		
	one-tailed	0.013809
	two-tailed	0.027618

GFgure G10: test between forest.–elevation

Data Summary	X=0	X=1	Total
n	32	411	443
ΣY	802	3524	4326
ΣY^2	21490	52408	73898
SS_Y	1389.875	22192.4866	31653.5847
$mean_Y$	25.0625	8.5742	9.7652

r_{pb}	t	df
-0.5	-12.29	441

P		
	one-tailed	<.0001
	two-tailed	<.0001

Figure G11: test between forest.– slope

Data Summary	X=0	X=1	Total
n	32	411	443
ΣY	3133	78527	81660
ΣY^2	822029	18886019	19708048
SS_Y	515288.718	3882394.35	4655326.55
$mean_Y$	97.9063	191.0633	184.3341

r_{pb}	t	df
+0.24	+5.08	441

P		
	one-tailed	<.0001
	two-tailed	<.0001

Figure G12: test between forest.– aspect