215547

DATA DRIVEN LANDSLIDE HAZARD ASSESSMENT
USING GEOGRAPHICAL INFORMATION SYSTEMS
AND REMOTE SENSING

119439

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

119439

MEHMET LÜTFİ SÜZEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
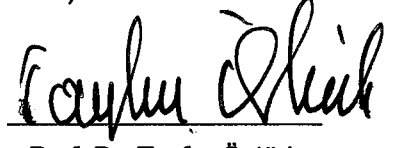
DOCTOR OF PHILOSOPHY

IN

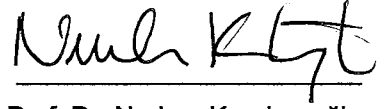THE DEPARTMENT OF GEOLOGICAL ENGINEERING

JANUARY 2002

Approval of the Graduate School of (Name of the Graduate School)
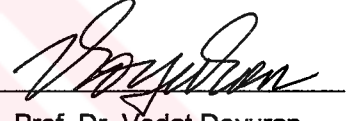
Prof. Dr. Tayfur Öztürk

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Nurkan Karahanoğlu

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 Doctor of Philosophy.
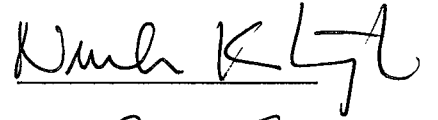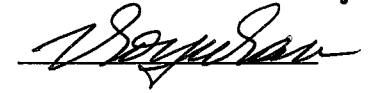
Prof. Dr. Vedat Doyuran
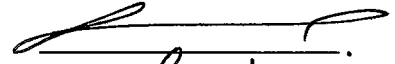
Supervisor

Examining Committee Members
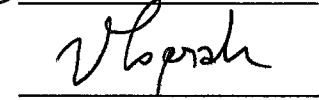
Prof. Dr. Nurkan Karahanoğlu
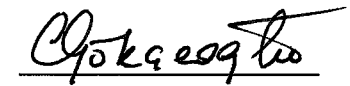
Prof. Dr. Vedat Doyuran

Prof. Dr. Reşat Ulusay

Assoc. Prof. Dr. Vedat Toprak

Assist. Prof. Dr. Candan Gökçeoğlu

ABSTRACT

DATA DRIVEN LANDSLIDE HAZARD ASSESSMENT USING GEOGRAPHICAL INFORMATION SYSTEMS AND REMOTE SENSING

Süzen, Mehmet Lütfi

Ph.D., Department of Geological Engineering

Supervisor: Prof. Dr. Vedat Doyuran

January 2002, 196 pages
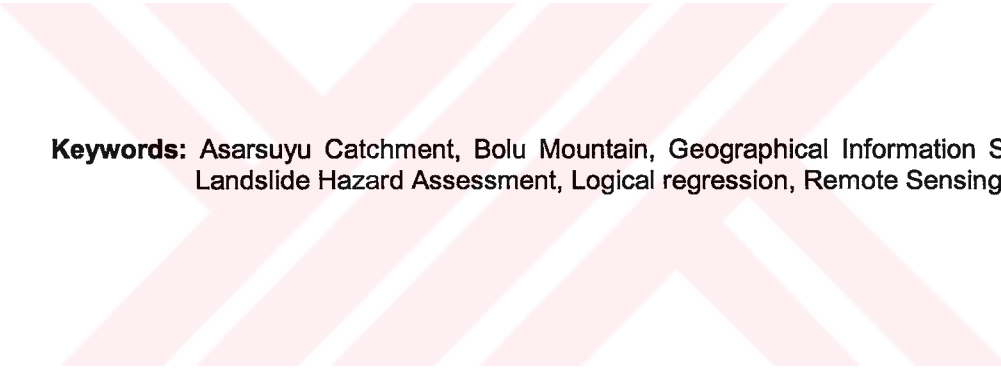
The purpose of this thesis is to generate and evaluate a concise system of data driven landslide hazard assessment procedure. The Asarsuyu catchment area and Bolu Mountain highway pass are selected for this purpose

In this study, three data domains, which are remote sensing products, geological maps and topographical maps are used. The landslides of four different historical periods are interpreted using aerial photographs. Land cover is extracted from satellite images. Thirteen sets of parameter maps are produced from geological and topographical maps and from remote sensing products. In order to store the information of these parameter maps in a concise thematic database a 25x25 meter grid is overlaid to the area. Two different sets of points are defined. First one representing the properties of slided masses and the other set representing the conditions contributing to the sliding phenomena. The information falling on these points are stored in five separate thematic landslide attribute databases, as two main databases are Seed Cells and Slided Mass Databases, with accessory three other as, Polystats, Fuzzystats and photo-characteristics databases. Following the creation of the databases, the

information stored are evaluated and preliminary landslide generating decision rules are extracted.

In conjunction with these preliminary results a hazard assessment procedure, starting from the least detailed and simple ranging up to most complex statistical analyses are initiated. Finally a relative accuracy assessment procedure is carried out using the hazard maps produced. It is found that the most accurate, reliable and realistic results concerning the landslide hazard assessment of Asarsuyu catchment are obtained through logistic regression analyses.

The results of landslide hazard assessment of Asarsuyu catchment shows that: the northern slopes of the Asarsuyu catchment is classified as very low hazard due to the presence of low population, undisturbed dense forest land cover, being very distant to E-5 highway and to the major active fault and the favorable lithological conditions. The southeastern slopes especially the Bolu Mountain Highway pass are definitely on very high hazard class due to the removal of lateral supports by E-5 highway cut slopes, close location to active faults, high disturbance of the land cover, high traffic activity along the highway resulting in extra vibration, and the presence of flyschoidal units

ÖZ

COĞRAFİ BİLGİ SİSTEMLERİ VE UZAKTAN ALGILAMA TEKNİKLERİ KULLANILARAK VERİ KAYNAKLI HEYELAN AFETİ TAYİNİ

Süzen, Mehmet Lütfi

Doktora, Jeoloji Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Vedat Doyuran

Ocak 2002, 196 sayfa

Bu çalışmanın amacı, kısa ve öz veri kaynaklı bir heyelan afeti tayin yöntemi geliştirmek ve değerlendirmektir. Bu amaçla uygulama arazisi olarak Asarsuyu su toplama havzası ve Bolu Dağı otoyol geçişi seçilmiştir.

Bu çalışmada, uzaktan algılama ürünleri, jeolojik haritalar ve topografik haritalar, olmak üzere üç veri alanı kullanılmıştır. Hava fotoğrafları kullanılarak dört ayrı dönemin heyelan envanteri çıkarılmıştır. Arazi kullanım haritası için ise, uydu görüntülerinden yararlanılmıştır. Jeolojik ve topoğrafik haritalardan ve uzaktan algılama ürünleri kullanılarak 13 ayrı parametre haritası üretilmiştir. Bu parametre bilgilerini bir veri tabanına aktarabilmek için, 25x25 metrelik bir ağ arazi üzerine oturtulmuştur. İki ayrı takım nokta belirlenmiştir. Bunlardan ilki kayan kütlelerin özelliklerini, diğeri ise kaymaya neden olan koşulları içermektedir. Bu noktalarda saklanan bilgiler, beş ayrı veri tabanına aktarılmıştır. Bunlardan kayan kütle ve kök hücre veri tabanları iki ana veri tabanını oluşturmaktadır. Diğer üç yardımcı veri tabanı ise, alansal (polystats), şekilsel (fuzzystats) ve foto-karakteristik veri tabanlarıdır. Veri tabanlarının üretilmesinden sonra heyelanlar hakkında ilk karar verme kuralları elde edilmiştir.

İlk sonuçlar ışığında en basit ve ayrıntısız analizlerden başlayarak, en karmaşık istatistiksel yöntemlere kadar uzanan bir heyelan afeti tayin yöntemi izlenmiştir. En son olarak ise, göreceli hata tayin yöntemi kullanılmıştır. Bunların ışığında en makul ve gerçeğe en uygun yötemin mantıksal regresyon olduğu sonucuna varılmıştır.

Heyelan afeti tayininin sonuçlarına gore, düşük nüfus yoğunluğu, bozulmamış arazi örtüsü, E-5 otoyoluna ile aktif fay hatlarına uzaklığı ve sağlam kaya birimleri ile Asarsuyu su toplama havzasının kuzey yamaçları çok düşük riskli afet alanı olarak belirlenmiştir. Güneybatı yamaçları, özellikle Bolu Dağı geçişi ise, E-5 yolunun yarmaları nedeniyle yok olan yanal destekler, aktif faylara olan yakınlığı, E-5 otoyolundaki yüksek trafik yoğunluğu ve fliş litolojisinin varlığı nedenyle çok yüksek riskli afet alanı olarak belirlenmiştir.

Anahtar Kelimeler: Asarsuyu Havzası, Bolu Dağı, Coğrafi Bilgi Sistemleri, Heyelan Afeti Tayini, Mantıksal Regresyon, Uzaktan Algılama

To my Family

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1 Purpose and Scope

Every year, thousands of people all over the world are loosing their lives in natural disasters, which are quite costly to the world economy both in humilian affairs and in economical aspects. Natural disasters have major direct and indirect economic and socio-economic effects in addition to the physical destruction that may occur. Furthermore, the impact of these disasters are higher in developing countries than in developed countries due to the fact that the developed countries had already created disaster mitigation and disaster preparedness programs.

Disasters are natural hazard events in which a natural phenomenon or a combination of natural phenomena such as earthquakes, mass movements, floods, volcanic eruptions, tsunamis etc., can cause many loss of lives and damage to the property. Almost no portion of the earth's surface is free from the impact of natural hazards. Due to increasing population densities and uncontrolled or not well-planned development, especially in developing countries, more and more people are affected by disasters. High precipitation, the deficiencies of infrastructure and lack of disaster plans and hazard maps are paid off by human lives in Turkey.

Disasters have both immediate and long-term implications. Plans formulated for disaster prone areas should therefore cover both these contingencies. It should also be remembered that a disaster might initiate a chain of severe hazards in addition to the direct impact damage.

Over the past two decades, many scientists have attempted to assess landslide hazards and produced maps portraying their spatial distribution. However, up till now there has been no general agreement on the methods or even on the scope of these investigations (Brabb, 1984, Carrara, 1983, 1989, 1993).

Despite the methodological and operational differences, all methods proposed are founded upon a single basic conceptual model. This model requires first the identification and mapping of a set of geological-geomorphological factors which are

directly or indirectly correlated with slope instability. It then involves both an estimate of the relative contribution of these factors in generating slope failure, and the classification of the land surface into zones of different hazard degree (Carrara, 1993).

Experience gained from hundreds of surveys carried out in different parts of the world has demonstrated that well trained investigators are able to detect and correctly map many or most of the landslides occurring in an area by applying aerial-photograph interpretation techniques and systematic field checks (Rib and Liang, 1978). However, old, dormant landslide bodies, are generally intensively modified by farming activity or covered by dense vegetation and thus, they cannot be readily identified and correctly classified. This introduces a factor of uncertainty that cannot be readily evaluated and explicitly incorporated in the subsequent phases of the analysis, being largely dependent on the skill of the surveyor, and the quality and the scale of aerial photographs and base-maps used. There, onwards the power of spatial data manipulation and use of multivariate statistics in a GIS environment arouses.

A further step towards hazard zonation would require the identification of the conditions leading to the slope failure, their systematic and consistent mapping, and evaluation of their relative contributions to the mass movements in the area. In fact, the causes of each slope failure are many, complex and sometimes unknown. Conversely the geological - geomorphological factors, which are both relevant to the prediction of landslide hazard that are mappable at an effective cost over a wide region are not as many yet (Carrara, 1989).

All these methods have significant advantages and drawbacks. The main advantages of the geomorphological approach lie in the capability of a skilful surveyor to estimate actual and potential slope failure, taking into consideration a large number of factors detected in the field or on the aerial-photographs. In addition, local or unique slope instability conditions can be identified and assessed. The major drawback of the approach concerns the high subjectivity characterizing all phases of the geomorphological investigation. The degree of uncertainty associated to the different phases of the mapping operations cannot be evaluated. Likewise it is difficult or sometimes impossible to compare landslide hazard maps produced by different surveyors.

The purpose of this thesis is to generate and evaluate a concise system of landslide hazard assessment procedure. With the intention of this goal few assumptions are to be made such as, the rainfall and earthquake are considered as triggering factors at which the outcomes of the system proposed should be considered with these triggering events. Furthermore, due to the small size of the area the effects of both precipitation and earthquake will be uniform through the study area, due to this fact they are not considered for the analyses.

2

In landslide hazard analysis it is also aimed that, (a). data set should be easily accessible and simple; (b). the system should be objective; (c). the system should be data-driven with minimal expert interaction and (d). the system should be simple and be easily implemented and used by novice users

The studies are initiated with an extensive review of the previous works carried on Landslide Hazard Assessment, that is presented in Chapter 2. The definitions, methodologies, contributions of remote sensing and review of available models in geographic information systems are considered. The geology of Asarsuyu Catchment constitutes Chapter 3 which is involved in regional geological setting, identification of lithological units, the North Anatolian Fault Zone and the earthquakes of the region. Types of input data and data production stage are given in Chapter 4. Every single parameter map and the methodologies to create these maps are explained. A total 13 sets of parameter maps are produced and used. The foundations of landslide attribute databases are also created in this chapter. Following the data production stages landslide related databases are analyzed in detail in Chapter 5. The comparisons of polystats and fuzzystats databases in 4 different periods, investigation of photo-characteristics database and evaluation of the parameter map values of landslide databases constitute the major issues. The core of this thesis is formed by the investigation of the available and newly produced hazard assessment procedures, which are presented in Chapter 6. Finally the results are discussed and conclusions are given.

1.2. Geographical Setting

The study area is located in the northwestern part of Turkey, in the area between Bolu and Düzce cities (Figure 1.1). The area is geographically defined as the catchment of Asarsuyu stream (Figure 1.2). The extents of the coordinates of Asarsuyu Catchment is defined as 4520500 N, 351500 E in the northwestern edge and 4506750 N, 372500 E in the southeastern edge in Universal Transverse Mercator projection with 36 North zone in European 1950 Mean Datum by International 1909/1924 / Hayford 1910 ellipsoid. The catchment is covered with six 1:25.000 scale topographical maps, that are G26-a3, G26-b3, G26-b4, G26-c1, G26-c2 and G26-d2 sheets. The catchment has an ellipsoidal form extending in east-west, having a maximum length of approximately 20 kilometers and the approximate width is 10 kilometers, also covering approximately 200 square kilometers.
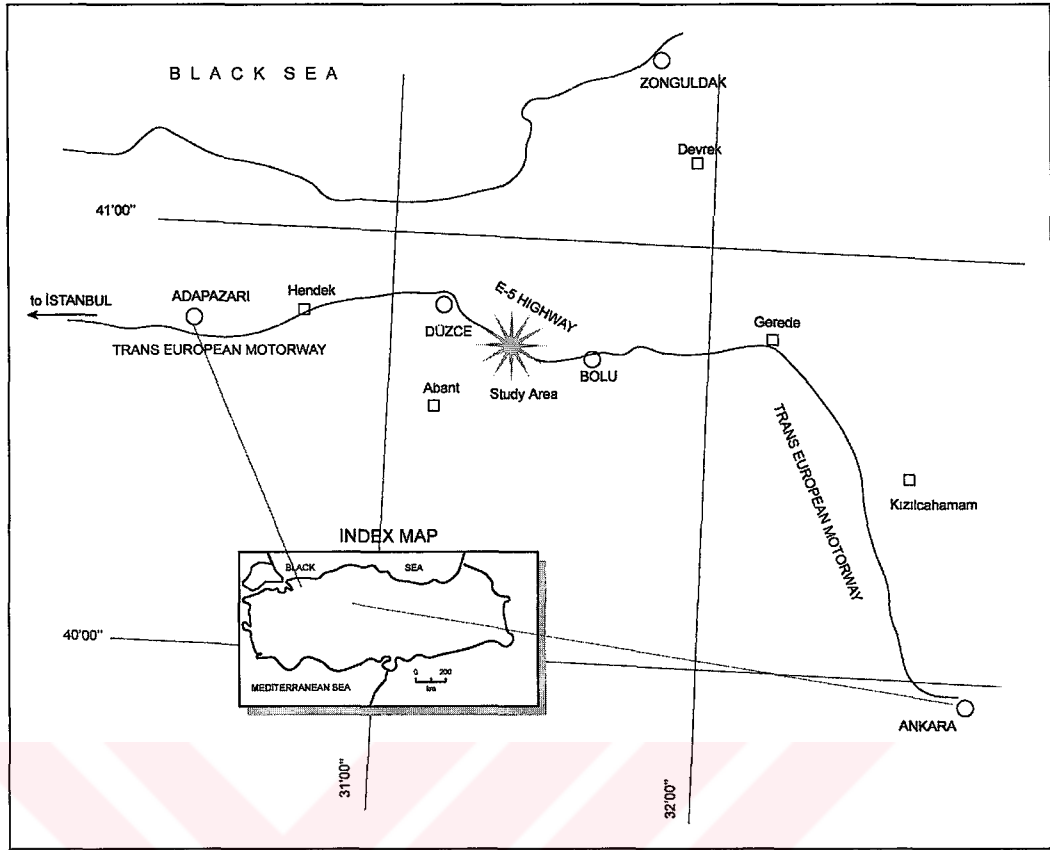
Figure 1.1. The geographic setting of the study area.
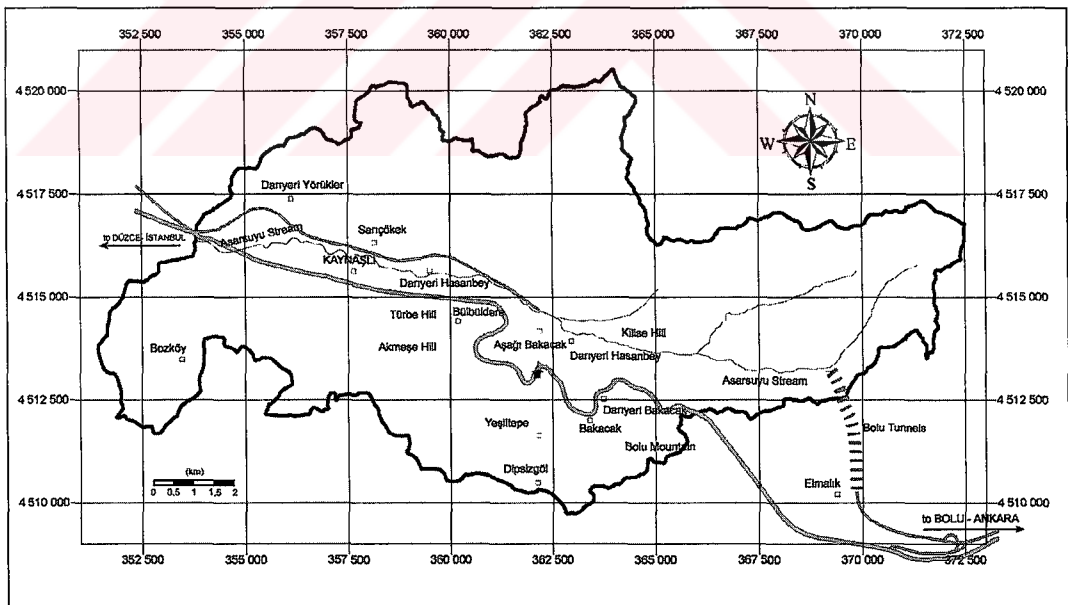
Figure 1.2. The outline of Asarsuyu catchment with important reference locations.

CHAPTER 2

BACKGROUND INFORMATION ON LANDSLIDE HAZARD AND USE OF REMOTE SENSING AND GEOGRAPHICAL INFORMATION SYSTEMS IN LANDSLIDE HAZARD ASSESSMENT

2.1. Definition of Landslide Hazard and Terminology

Mass movement is defined as "the outward and downward gravitational movement of earth material without the aid of running water as a transporting agent" by Crozier (1986), or "the movement of a mass of rock, debris or earth down a slope" by Cruden (1991). These are the internationally accepted and most widely used definitions of the phenomenon. Although they are slightly different from each other considering beyond the scope of inclusion of water, they both point a mass transportation down the slope in which a hazardous activity for humans can occur.

Rather than dealing with the types, activities and definitions, as they are defined by the IAEG Commission on Landslides in the 1990's, a more relational approach is given by Soeters and van Westen (1996) "Slope instability processes are the product of local geomorphic, hydrologic and geologic conditions; the modification of these by geodynamic processes, vegetation, land use practices and human activities; and the frequency and intensity of precipitation and seismicity".

Mass movement or slope instability or landsliding are natural denudational and degradational processes, unless they are affecting human life. Their interference with ongoing human practices in the terrain makes it a landslide hazard. The general accepted terminology in the world is followed below by Varnes (1984):

Natural hazard (H): The probability of occurrence of a potentially damaging phenomenon within a specified period of time and within a given area (Figure 2.1).

Vulnerability (V): The degree of loss a given element or set of elements at risk resulting from the occurrence of a natural phenomenon of a given magnitude. Scale is 0 (no change) to 1 (total loss) (Figure 2.1).

Specific risk (Rs): The expected degree of loss due to a particular natural phenomenon. It may be expressed by the product of H and V.

$$Rs=H^*V$$

Elements at Risk (E): The population, properties, economic activities, including public services and etc. at risk in a given area.

Total Risk (Rt): The expected number of lives lost, persons injured, damage to property or disruption of economic activity due to a particular natural phenomenon (Figure 2.1). It is expressed as

$$Rt=E^*Rs$$
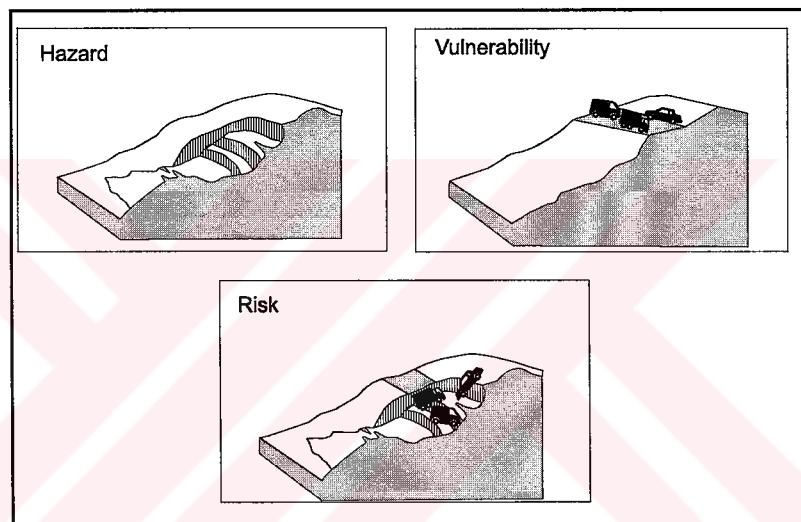
Figure 2.1. Graphical representation of hazard, vulnerability and risk

Based on the above definitions, hazard and risk information are generally represented as discrete maps. The discrete classes represent equal probability classes, which are inturn equal hazard or risk classes. The differentiation of hazard classes and their groupings are called as "zonation". The formal definition is as follows: "The term zonation refers to the division of land into homogenous areas or user defined domains and the ranking of these areas according to their degrees of actual or potential natural hazards" (Varnes, 1984).

The natural hazard zoning/mapping constitutes the first and major task of the earth scientists in natural hazard analysis (Figure 2.2). The zoning of a natural hazard is the vital part of the study strategy in which the whole strategy will be based on. The zonation activities are mutually dependent over some factors as shown in Figure 2.2.

6

These factors can be grouped into magnitude properties of the hazard, frequency of the hazard and the spatial location of the hazard. The next step in hazard mapping is to show the mapped hazard and to classify the hazard map into some homogenous areas regarding the equal attributes of the hazard map.
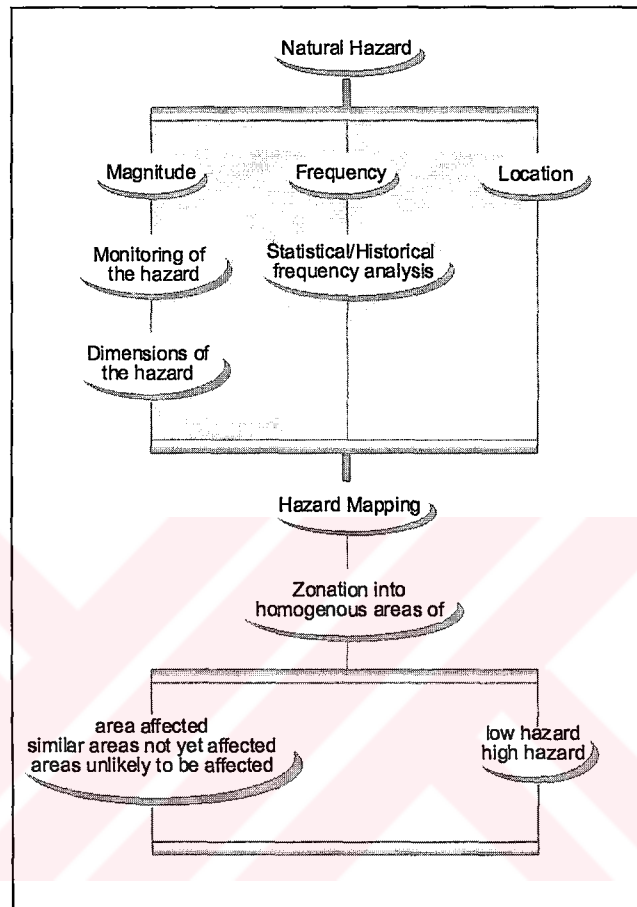
Figure 2.2. An overview of zonation activities

The natural hazard zoning is controlled mainly by two factors, such as: the scale of the zoning or mapping and the knowledge type used in the hazard zoning.

2.1.1. Scale Factor in Analysis

Before starting to any data collection, an earth scientist working on a hazard analysis project should have to answer a number of interrelated questions:

1. What is the aim of the study?
2. What scale and with what degree of precision must the result be presented?
3. What are the available resources in the form of money, data and manpower?

7

As the aim of the study would be previously defined, the scale and the precision are the first parameters to be defined prior to the start of the project. Hence, the scale factor should have to be determined at the first glance as it controls the type of the input data, nature of the analysis and the output data of the project. The outcome precision is also dependent on the scale chosen. Although the precision is dependent on scale, it is also an independent parameter regarding the nature of the project. The necessary adjustments should be made with the scale until the output precision and the desired precision fulfills the project conditions. The resource analysis will be conducted after the aim and scale is fixed.

The following scales of analysis, which were presented in the International Association of Engineering Geologists (IAEG) Monograph on engineering geological mapping (IAEG, 1976) can also be distinguished in general natural hazard zonation (Figure 2.3).

2.1.1.1. National Scale (<1/1.000.000)

The national scale analysis is used only to outline the problem, give an idea about the hazard types and affected hazard prone areas. They are prepared generally for the entire country and the required map detail is very low, even in the best case giving only data based on records in the form of an inventory. The degree of the hazard is assumed to be uniform. These kinds of maps are generally prepared for agencies dealing with regional (agricultural, urban or infrastructure) planning or national disaster prevention / hazard assessment agencies.

2.1.1.2. Regional/Synoptic Scale (< 1/100.000)

The scale is still so small to use in any quantitative method, but these maps are used for regional planning and in early stages of region wise planning activities. The areas to be investigated are still large in the order of thousands of square kilometers and the map detail is low again. Only simple methods are used with qualitative data combination and the zoning is primarily based on regional geomorphological Terrain Mapping Units / Complexes (TMU) or dependent on regional geological units.

2.1.1.3. Medium Scale (1/25.000 - 1/50.000)

These hazard maps are made mainly for agencies dealing with intermunicipal planning or companies dealing with feasibility studies for large engineering works. The areas to be investigated will have areas of several hundreds of square kilometers. At this

scale considerably more detail is required than at the regional scale. These maps do serve especially for the choice of corridors for infrastructure construction or zones for urban development. Statistical techniques are dominantly used in this scale.

1.National Scale
  <1:1.000.000
  Inventory
  Entire Country
  Based on Records

2. Regional Scale
  <1:100.000
  Regional Planning
  Large Areas
  Simple Methods

3. Medium Scale
  1:25.000 to 1:50.000
  Local Planning
  Areas up to 200 sqkm
  Statistical Methods

4. Large Scale
  > 1:10.000
  Detailed Planning
  Small areas
  Stability analysis
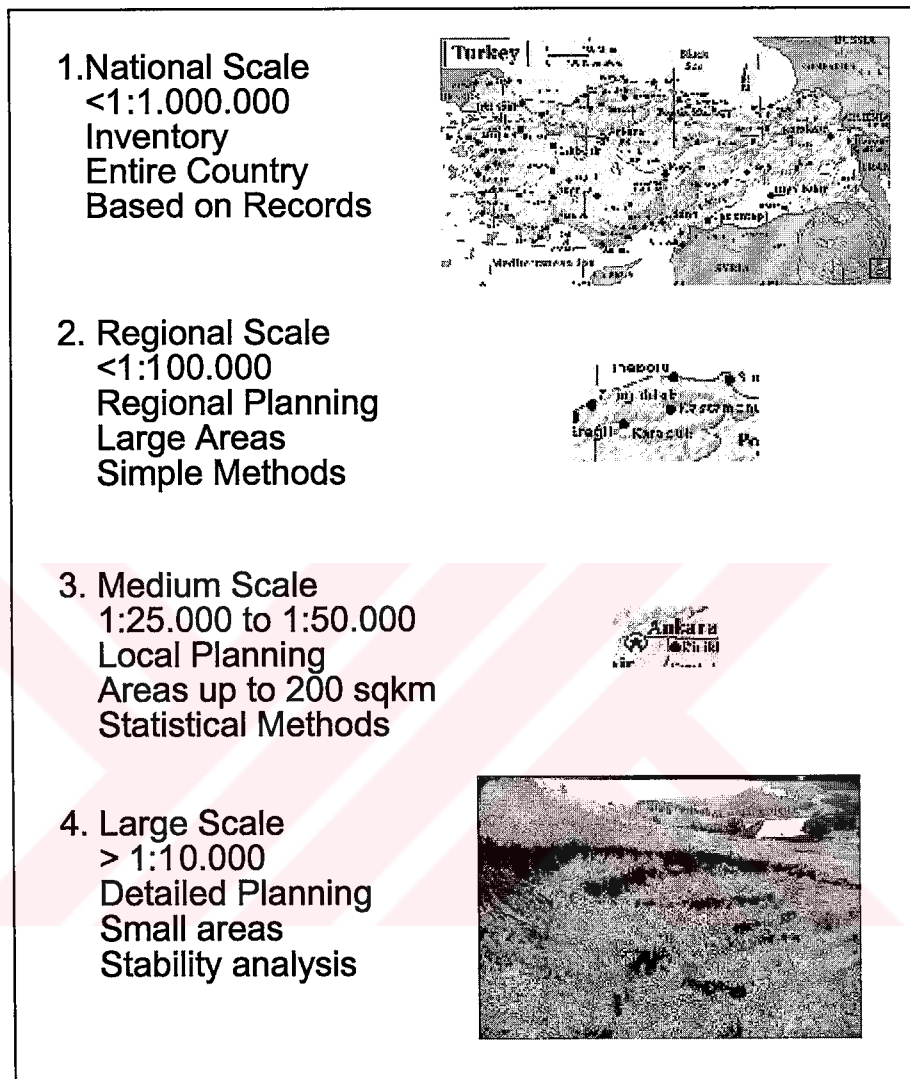
Figure 2.3. The scales of analysis and minor details

2.1.1.4. Large Scale (> 1/10.000)

These hazard maps are produced generally for authorities dealing with detailed planning of infrastructure, housing or industrial projects or with evaluation of risk within a city or within a specified project area. They cover very small areas hence the deterministic hazard analyses become available to be used. The detail level of the maps is set into maximum. They are based on physical numerical models that require extensive data collection in the field and laboratory surveys.

2.1.2. Knowledge Type Used

Prediction of landslide hazard for areas not currently subject to landslide hazard is based on the assumption that hazardous phenomena that have occurred in the past can provide useful information for prediction of future occurrences. Unlike general educational geological phrases in this case "present is not key to the past but present and past are the keys of future", of which the real value of engineering and its futuristic approaches are represented. Therefore, mapping these phenomena and the factors thought to be of influence is very important in hazard zonation. In relation to the analysis of the terrain conditions leading to slope instability, two basic methodologies can be recognized (van Westen, 1993):

1. The first mapping methodology is the experience-driven applied-geomorphic approach, by which the earth scientist evaluates direct relationships between landslides and their geomorphic and geologic settings by employing direct observations during a survey of as many existing landslide sites as possible. This is also known as direct mapping technology.

2. The opposite of this experience-based, or heuristic approach is the indirect mapping methodology, which consists of mapping a large number of parameters considered to potentially affect landsliding and subsequently analyzing (statistically) all these possible contributing factors with respect to the occurrence of slope instability phenomena. In this way the relationships between the terrain conditions and the occurrence of the landslides may be identified. On the basis of the result of this analysis, statements are made regarding the conditions under which slope failures occur.

Another division of techniques for assessment of slope instability hazard was given by Hartlen and Viberg (1988), who differentiated between relative and absolute hazard assessment techniques.

1. Relative hazard assessment techniques differentiate the likelihood of occurrence of mass movements for different areas on the map without giving exact values.
2. Absolute hazard maps display an absolute value for the hazard such as a factor of safety or a probability of occurrence.

Furthermore the hazard assessment techniques can also be divided into three broad classes based on use of statistical methods (Carrara, 1983; Hartlen and Viberg, 1988; Soeters and van Westen, 1996).

1. White box models: based on physical models (slope stability and hydrologic models), also referred to as deterministic models
2. Black box models: not based on physical models but strictly on statistical analysis
3. Gray box models: partly based on physical models and factual data and partly on statistics.

2.2. Use of Remote Sensing in Landslide Hazard Assessment

The phenomenon, landslide is affecting the earth's surface, hence it also falls in to the research and application areas of both aerial and space born remote sensing. The nature of this phenomenon as it is occurring at the surface of earth lets earth scientists to exploit this fact using remotely sensed data. On the other hand, again the nature of this phenomenon limits the applications, as being dynamic and sometimes being quite small in terms of conservative remote sensing language. Furthermore they reveal very small information when they are observed in planar 2-D, however, they contain large amounts of data when explored in 3-D. Basing on this fact the use of stereo-remote sensing products seems to be indispensable, which reveals the true morphodynamical features of the landslides. These information are providing the diagnostic information regarding the type of the movement (Crozier, 1973). The general application fields of remote sensing in landslide business are monitoring the change of landslide activities through time (change detection) and mapping out where the hazard occurs.

Plenty of researchers have tested the usage of remote sensing products through the last 30 years. Two major groupings could be made upon the investigation of these researches. These are aerial photography and space-born sensor images.

Numerous applications have been carried out which generally define the landslide areas. Chandler and Moore (1989), Chandler and Brunsden (1995) and Fookes et al. (1991) give excellent applications for photogrammetry. For single landslides and for smaller areas, a monitoring scheme is best applied with this technique with large accuracies. In opposition, the applicability of this technique limits the extents of the interest area as the larger areas could be accomplished by classical aerial photographical studies easily. For stereographical aerial photography Rengers (1986), Sissakian (1986) and Mollard (1986) could be counted as single application manuscripts. However, studies with landslides and aerial photographs are as old as the applications of

first stereographical aerial photographs, resulting in plenty of textbooks and textbook sections.

The landslide information extracted from the remotely sensed products is mainly related with the morphology, vegetation and the hydrological conditions of the slope. The slope morphology is best examined with stereographical coverages. Generally the identification of the slope instabilities are indirect, they are identified by associated elements with slope instability process. The advantages of aerial photographs can be listed as:

1. They provide quite older coverages before digital world starts
2. The flight coverages are flexible for new missions
3. The spatial and temporal resolution are very high
4. Stereoscopic coverage
5. Most of the geoscientist are familiar
6. Every country have at least one full coverage of their land due to military reasons

The disadvantages are as follows:

1. Low spectral resolution
2. The nature of photograph as hardcopy
3. Presence of distortions in the images
4. Absence of coordinate information
5. Orthorectification is needed to remove distortion and add coordinate information
6. The resultant map is dependent to the experience of interpreter

The applications with space born images are quite new compared to the others. Furthermore, they are generally defining the landslides indirectly by mapping out other parameters such as land cover. Some examples from the literature could be said of Gagon (1975); Mc Donalds and Grubbs (1975); Sauchyn and Trench (1978); Stephens (1988); Huang and Chen (1991) and Vargas (1992).

In comparison to the aerial photographs, the advantages of satellite images are:

1. Getting the bigger picture
2. Larger spectral range
3. Easily accessible
4. No distortion
5. Only georeference is needed to transfer the coordinates

The disadvantages are:

1. Low spatial resolution
2. More expensive than aerial photographs of the same resolution
3. Limited stereo capability
4. Limited number geoscientists are familiar

Although there are plenty of disadvantages of aerial photographs, they are the most frequently used medium in landslide projects as they have cheaper high-resolution images. The spatial resolution nearly controls everything in landslide hazard assessment. The comparison of spatial resolution of photographic and non-photographic remote sensing product requires the concept of Ground Resolution Cell (GRC) defined by Strandberg (1967) and introduced to landslides first by Rengers et al. (1992). Strandberg (1967) suggested that the formula for GRC is in relation to scale as:

$$GRC=S/1000R$$

Where GRC is ground resolution cell in meters, S in the scale number of photograph and R is the resolution of photographic system (line pairs/mm, normally 40 in conventional systems)

Soeters and van Westen (1996) figured out the necessary minimum number of GRC's namely the pixels in the images, in order to identify and interpret the landslides, which is presented in Table 2.1. They also exploited this information and created a comparison table of photographic images with non-photographic ones that is also presented in Table 2.2.

Table 2.1. The number of GRC needed to identify and interpret object of varying contrast in relation to its background (Soeters and van Westen, 1996).

	The number of GRC	
	For identification	For interpretation
Extreme contrast	20-30	40-50
High contrast	80-100	120-140
Low Contrast	1000-1200	1600-2000

Table 2.2. Minimum object size needed for landslide Identification or Interpretation (Soeters and van Westen, 1996).

	GRC size (m)	Size m² needed for			
		High Contrast		Low Contrast	
		Identification	Interpretation	Identification	Interpretation
Landsat MSS	~80	160000	288000	7040000	11520000
Landsat 5 TM	30	22500	40500	990000	1620000
Spot Multispectral	20	10000	18000	440000	720000
Spot Panchromatic	10	2500	4500	110000	180000
Aerial Photographs					
1:50000	1	25	45	1100	1800
1:15000	0,3	6,5	11,5	300	450

Basing on the above facts, aerial photography is still indispensable in the landslide hazard zonation activities. However, remote sensing is depicting an important role in landslide hazard assessment, though, this role is not the primary role in the game.

2.3. Geographical Information Systems and Landslide Hazard Assessment

A GIS is defined as a "powerful set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world for particular set of purposes" (Burrough, 1986). A more specific definition is given by Bonham-Carter (1996) as follows: "a geographic information system, or simply GIS, is a computer system for managing spatial data. The word *geographic* implies that the locations of the data items are known, or can be calculated, in terms of geographical coordinates. The word *information* implies that the data in GIS are organized to yield useful knowledge, often as colored maps and images, but as also statistical graphics, tables and various on-screen responses to interactive queries. The word *system* implies that a GIS is made up from several interrelated and linked components with different functions. Thus, GIS has functional capabilities for data capture, input, manipulation, transformation, visualization, combination, query, analysis, modeling and output."

These international valid definitions of GIS directly oppose to the belief that GIS is only a CAD software or only a drawing tool. CAD can only constitute a small portion of the whole integrated system, in which an ideal GIS and its possible components are shown in Figure 2.4.

14
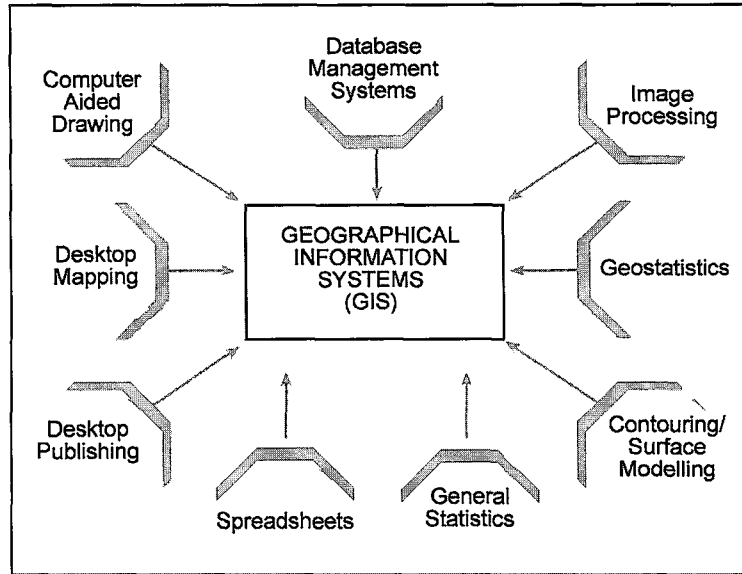
Figure 2.4. GIS and its related software systems as components of GIS (modified from Bonham-Carter, 1996)

Generally a GIS consists of the following phases (Figure 2.5).
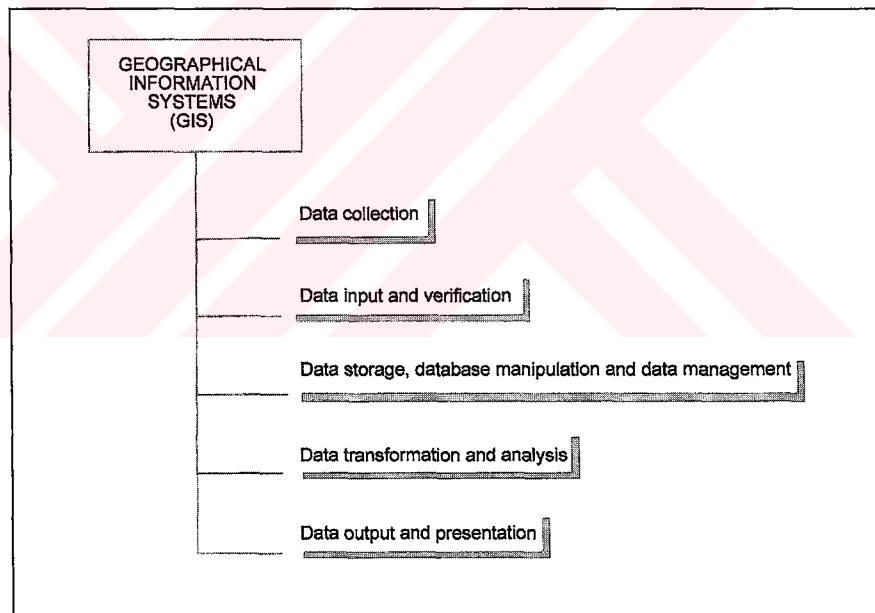
Figure 2.5. The phases of a GIS

A GIS if based over the former components should answer the following questions (Figure 2.6):

More and more the products of mapping and inventory are being stored in data banks for their ultimate retrieval or combination with data from other sources. Often they are incorporated is GIS or LIS (Land Information Systems) which serve as a base for

programmable data manipulation and selective information extraction for planning and project assessment.
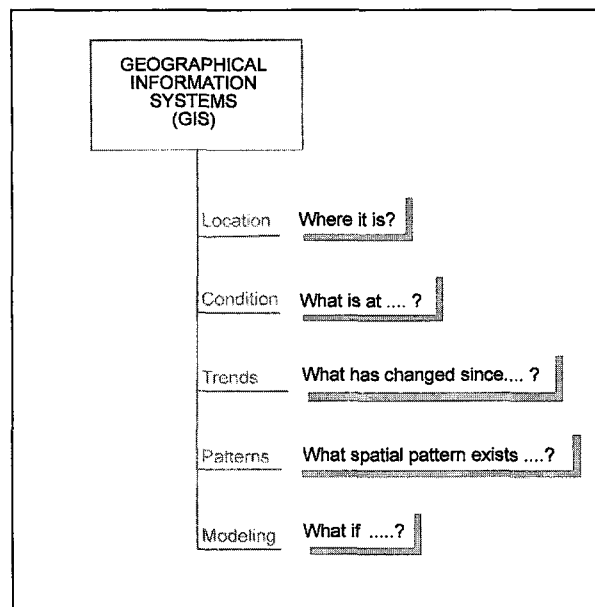
Figure 2.6. The questions of a well-built GIS should answer.

The development of GIS and LIS systems is of considerable interest in the context of satellite surveying, change detection and monitoring. The flexibility of digital data processing, combined with quick input of new data (possible from updating on the basis of satellite remote sensing records) offers new possibilities to the surveyor, cartographer and planner.

It is clear that in a rapidly developing society, change detection is of great importance. In modern society, mapping suffers from high rate of change: change in land use in rural and urban areas; change in requirements for maps and inventories; change in concepts in the various disciplines of earth and social sciences, leading to different interpretations of the same data and change in the economical and technical factors on which mapping methods were based.

In order to refine the discussion around landslide hazard one can say that, the occurrence of slope failure depends generally on complex interactions among a large number of partially interrelated factors. Analysis of landslide hazard requires evaluation of the relationships between various terrain conditions and landslide occurrences. An experienced earth scientist has the capability to mentally assess the overall slope conditions and to extract the critical parameters. However, an objective procedure is often desired to quantitatively support the slope instability assessment. This procedure requires the evaluation of the spatially varying terrain conditions as well as the spatial representation of landslides. A GIS allows for the storage and manipulation of

information concerning the different terrain factors as distinct data layers and thus provides an excellent tool for slope stability hazard zonation.

The advantages of GIS for assessing landslide hazard include the followings:

1. A much larger variety of hazard analysis techniques become attainable. Because of the speed of calculation, complex techniques requiring a large number of map overlays and table calculations become feasible.
2. It is possible to improve models by evaluating their results and adjusting the input variables. Users can achieve maximum results by a process of trial and error, running the models several times, whereas it is difficult to use these models even once in the conventional manner. Therefore, more accurate results can be expected.

The disadvantages of GIS for assessing landslide hazard include the followings:

1. A large amount of time is needed for data entry. Digitizing is especially time consuming
2. There is a danger in placing too much emphasis on data analysis as much as the expense of data collection and manipulation based on professional experience. A number of different techniques of analysis are theoretically possible, but often the necessary data are missing. In other words, the tools are available but cannot be used because of the lack or uncertainty of the data.

2.3.1. Phases of Natural Hazard Analysis in GIS

The following phases can be distinguished in the process of a hazard analysis using GIS (van Westen, 1993); there is a logical order in the sequence though sometimes they may be overlapping. The time schedule of these phases is listed in Table 2.3.

1. Choice of the working scale and the methods of analysis which will be applied
2. Collection of existing maps and reports with relevant data
3. Interpretation of images and creation of new input maps
4. Design of the database and definition of the way in which the data will be collected and stored.
5. Fieldwork to verify the photo-interpretation and to collect relevant quantitative data
6. Digitizing of maps and attribute data
7. Validation of the entered data

8. Manipulation and transformation of the raw data to a form which can be used in the analysis

9. Analysis of data for preparation of hazard maps.

10. Evaluation of the reliability of the input maps and inventory of the errors which may have occurred during the previous phases

11. Final production of hazard maps and adjoining report

Table 2.3. Time schedule comparison of phases of landslide hazard assessment of conventional methods and GIS based methods based on scale (numbers are in percents of the total project time) (van Westen, 1993).

PHASES	Regional Scale		Medium Scale		Large Scale	
	Conventional Methods	GIS Based Methods	Conventional Methods	GIS Based Methods	Conventional Methods	GIS Based Methods
1. Choice of scale and methods	<5	<5	<5	<5	<1	<5
2. Collection of existing data	<5	<5	<5	<5	8	8
3. Image Interpretation	50	50	30	30	10	20
4. Database design	0	<5	0	<5	0	<5
5. Fieldwork	<5	<5	7	7	10	20
6. Data Entry	0	20	0	30	0	15
7. Data Validation	0	<5	0	5	0	5
8. Data Manipulation	0	<5	0	5	0	5
9. Data Analysis	30	10	48	10	61	10
10. Error Analysis	0	<5	0	<5	0	<5
11. Final Map Production	10	<5	10	<5	10	<5

2.3.2. GIS Based Landslide Hazard Zoning Techniques

An ideal map of slope instability hazard should provide information on the spatial probability, temporal probability, type, magnitude, velocity, runout distance and retrogression limit of the mass movements predicted in a certain area (Hartlen and Viberg, 1988). A reliable landslide inventory defining the type and activity of all landslides, as well as their spatial distribution, is essential before any analysis of the occurrence of landslides and their relationship to environmental conditions are undertaken. Even the inventory of historical periods are of great use in the final analyses. The differentiation of slope instability according to type of movement is important, not only because different types of mass movement will occur under different terrain conditions, but also because the impact of slope failures on the environment has to be evaluated according to type of failure.

18

2.3.2.1. Trends in Landslide Hazard Zonation

A large amount of research on hazard zonation has been done in last 30 years, as the consequence of an urgent demand for slope instability hazard mapping. Overviews of the various slope instability hazard zonation techniques can be found in Hansen (1984), Varnes (1984), Hartlen and Viberg (1988). The general trends in landslide hazard zonation are given in Table 2.4. The distribution analyses and qualitative analyses are generally used for very large areas with very low detail such as national hazard maps. The deterministic and frequency analyses are used generally for very small areas such as specific large engineering projects like dams, nuclear power plants, highway strips, open pit mine slopes and spoils. Monitoring and laboratory analyses are indispensable for these analyses. Good reviews of these initial deterministic methods can be found in Lambe and Whitman (1969), Hoek and Bray (1981), Graham (1984), Bromhead (1986) and Anderson and Richards (1987). The limited GIS examples of these methods could be cited as Ward et al. (1982), Okimura and Kawatani (1986), Mulder and van Asch (1988), Mulder (1991) and Hammond et al. (1992). The statistical analyses have the most flexibility in scale and in data type and will be investigated in detail in the following sections.

Table 2.4. The trends in landslide hazard zonation (van Westen, 1993).

Type of landslide hazard analysis	Main characteristics
A. Distribution analysis	Direct mapping of mass movement features resulting in a map, which gives information only for those sites where landslides have occurred in the past.
B. Qualitative analysis	Direct, or semi-direct, methods in which the geomorphological map is re-numbered to a hazard map, or in which several maps are combined into one using subjective decision rules, based on the experience of the earth scientist.
C. Statistical analysis	Indirect methods in which statistical analyses are used to obtain predictions of the mass movement hazard from a number of parameter maps.
D. Deterministic analysis	Indirect methods in which parameter maps are combined in slope stability calculations.
E. Landslide frequency analysis	Indirect methods in which earthquake and/or rainfall records or hydrological models are used for correlation with known landslide dates, to obtain threshold values with a certain frequency.

2.3.2.2 Direct Mapping in Landslide Hazard Analysis

2.3.2.2.1 Landslide Distribution Analysis

The most straightforward approach to landslide hazard zonation is a landslide inventory, based on any or all of the following; aerial photo interpretation, ground survey, and a database of historical occurrences of landslides in an area. The final product gives the spatial distribution of mass movements, which may be represented on a map either as affected areas to scale or point symbols (Wieczorek, 1984).

Such mass movement inventory maps are the basis for most other landslide hazard zonation techniques. However, they can be used as an elementary form of hazard map because they display the location of a particular type of slope movement. They provide only information for the period shortly preceding the date that aerial photographs are taken or the fieldwork conducted. They provide no insight into temporal changes in mass movement distribution. Many landslides occurred some time before photographs are taken my have become undetectable. Therefore, a refinement is the construction of landslide activity maps, which are based on multitemporal aerial photo interpretation (Canuti et al., 1979, 1985, 1986). Landslide activity maps are indispensable to study the effects of temporal variation of a factor such as land use on landsliding. Landslide distribution can also be shown in the form of a density map (Wright et al, 1974). The resulting density values are interpolated and used as landslide isopleths. They can also be used to cite out the current situation of the landslide density per terrain mapping unit or catchment or a predefined geological unit. This method may also be used to test the importance of each individual parameter for predicting the occurrence of mass movements. If the method is used to test the importance of specific parameter classes, the user decides, on the basis of his/her field experience, which individual maps or combination of parameter maps will be used. The method is most appropriate at medium or large scales. At the regional scale the construction of a mass movement distribution or activity map is very time consuming and too detailed for procedures of general zoning.

However, the selection of terrain mapping unit and the conversion of continuous parameter maps into discrete parameter maps involve a quite large subjectivity into the analysis. Furthermore, this analysis should have to be done for each parameter map and for different parameter classes. The effects of the separate parameters with respect to each other are not implemented, hence it is still expert dependent who will be on the charge to define the parameter classes and the parameter maps to be used. The analysis is similar to general bivariate analyses, but does not end up with a hazard score

of any sampling frame in the area. Basically, a simple density per kilometer or per sample area will provide much more objective results about the factual data.

2.3.2.2.2 Heuristic Approach (Geomorphic Analysis)

In heuristic methods the expert opinion of the earth scientist making the survey is used to classify the hazard. These methods combine the mapping of mass movements and their geomorphologic setting as the main input factor for hazard determination.

The basis of geomorphic analysis was outlined by Kienholz (1977), who developed a method for producing a combined hazard map based on the mapping of "silent witnesses" (Stumme Zeugen). The geomorphic method is also known as the direct mapping method. The hazard is determined directly either in the field or by photo or satellite image interpretation by the earth scientist. The process is based on individual experience and the use of reasoning by analogy. The decision rules are, therefore, difficult to formulate because they vary from place to place, yielding as unformalized applicable rules that vary from polygon to polygon. This method is totally subjective and dependent on the skill and experience of the earth scientist. However, GIS serves as an undeniable tool for reproduction and querying the entered data. This method can be applied at all scales in a relatively short period. Some examples of geomorphic analyses can be found in Carrara and Merenda (1974), Brunsden et.al. (1975), Stevenson (1977), Malgot and Mahr (1979), Kienholz (1977,1978,1980,1984), Kienholz et al. (1983,1988), Grunder (1980), Ives and Messerli (1981), Rupke et al. (1987,1988), Perrot (1988), Hermelin (1990,1992), Hearn (1992) and Seijmonsbergen (1992). A weighting scheme is also present in this type of analysis, however this weighting scheme is also quite subjective and "blind weighting" is suggested for this type of weighting by Gee (1992).

2.3.2.3. Indirect Mapping in Landslide Hazard Analysis

2.3.2.3.1. Statistical Methods in Landslide Hazard Analysis

Aiming at a higher degree of objectivity and better reproducibility of the hazard zonation, which is important for legal reasons, statistical techniques have been developed for assessment of landslide hazard.

In statistical landslide hazard analysis the combinations of factors that have led to landslides in the past are determined statistically, and quantitative predictions are made for areas currently free of landslides but where similar conditions exist. Furthermore, overlying of parameter maps and calculation of landslide densities form the

core of the analysis. Most of the analyses are based on the relationship between the landslide densities per parameter class compared with the landslide density over the entire area. Each method has its own specific rules for data integration required to produce the total hazard map. Two different statistical approaches are used in landslide analyses: bivariate and multivariate approaches.

Although the statistical techniques can be applied at different scales, their use becomes quite restricted at the regional scale, where an accurate input map of landslide occurrences may not be available and where most of the important parameter cannot be collected with appropriate accuracy. At large scales, different factors will have to be used, such as water table depth, soil layer sequences and thicknesses. These data are very difficult to obtain even for relatively small areas. Therefore, the medium scale is considered most appropriate for this technique.

2.3.2.3.1.1. Bivariate Statistical Methods in Landslide Hazard Analysis

In this method, overlay of parameter maps and calculations of landslide densities form the core of the analysis, the importance of each parameter, or specific combinations of parameter can be analyzed individually. Using normalized values (landslide density per parameter class in relation to the landslide density over the whole area), a total hazard map can be made by addition of the weights for individual parameters. The weight values can also be used for design decision rules, which are based on the experience of the earth scientist. It is also possible to combine various parameter maps into a map of homogenous units, which is then overlaid by the landslide map to give a density per unique combination of input parameters.

It should be stressed that the selection of parameters has also an important subjective element in this method. The following GIS procedures are used (van Westen, 1993).

1. Classification of each parameter map into a number of relevant classes.
2. Combination of the selected parameter maps with landslide map via map overlay
3. Calculation of weighting values based on the cross table data
4. Assignment of weighting values to the various parameter maps or design of decision rules to be applied to the maps, and classification of the resulting scores in a few hazard classes.

As it is seen from the procedure list the first and the last item contains quite large subjectivity, it is not clear to how to divide the parameter maps into classes and

how many classes should there be? Furthermore, the division of the final hazard map into hazard classes inherits the same problem. This problem limits these methods, as the start and the end directly depends on the expert, which means a degradation in the final hazard map and also limits the reproducibility of the hazard maps under different conditions.

The medium scale is most appropriate for this type of analysis. The method is not detailed enough to apply at the large scale, and at the regional scale the necessary landslide occurrence map is difficult to obtain.

Bivariate statistical analysis deals with one dependent variable (in this case the occurrence of mass movements) and one independent variable. The importance of each factor is analyzed separately. Specific combinations of variables can also be tested by treating the combination map as a new variable. The methods are based on the assumption that the important factors leading to mass movements can be quantified by calculating the density of mass movements for each variable class. However, the new parameter map production as crossing the available parameter maps in fact carries this bivariate analysis procedure into somewhat multivariate domain; as the factor analysis in the multivariate domain also bases its core on new parameter maps with different factor loadings from the initial parameter maps.

In bivariate statistical analysis, each factor map is combined with the landslide distribution map, and weighting values based on landslide densities are calculated for each parameter class. Several statistical methods have been applied to calculate weighting values; these have been termed the *landslide susceptibility method* (Brabb, 1984; van Westen, 1992, 1993), *Information value method* (Kobashi and Suzuki, 1988; Yin and Yan, 1988), *weight of evidence modeling method* (Spiegelhalter, 1986; Bonham-Carter, 1996). Furthermore, there still exist not enough exploited methods as Bayesian combination rules, certainty factors, Dempster and Shafer belief method and fuzzy logic. The three of the methods will be further investigated in the following sections, which are landslide susceptibility analyses, information value analyses and weights of evidence methods. The first two of them depend on the density of landslides in parameter classes; even though they result in approximately same hazard maps, the calculation schemes are different from each other. The weights of evidence method utilize the usage of binary dumb variables, and again based on the probability of occurrence of landslides in parameter classes. However, the usage of binary dumb variables turns a simple GIS in to a small scale chaos especially in landslide hazard projects. As this method is first created to assess the locations of ore deposits, the variables of its initial version were well defined and directly depend on factual data, where all of the answers of "what if?" questions were known.

2.3.2.3.1.1.1. Landslide Susceptibility Analysis

A simple and useful method in statistical analysis to determine the importance of different variables for the occurrence of mass movements is the use of pair wise map crossing. In order to evaluate the importance of the individual maps, a cross between these maps and a landslide occurrence map is prepared. For each variable class and landslide type, two types of densities can be calculated.

1. Area density: the density expressed as the number of pixels with landslides divided by the total number of pixels within the variable class. This can be displayed as a percentage or permillage contents.

$$D_{area} = 1000 \frac{Npix(SX_i)}{Npix(X_i)}$$

where D_{area} : Areal density per millage

$Npix(SX_i)$: number of pixels with mass movements within variable class X_i.

$Npix(X_i)$: number of pixels within variable class X_i.

2. Number density: the density expressed as the number of landslide occurrences per square kilometer of the area of the variable class.

$$D_{number} = \frac{1*10^6}{Area(X_i)} Number(SX_i)$$

where D_{Number} : Number Density (Number/km^2)

$Area(X_i)$: Area in square meters of variable class X_i.

$Number(SX_i)$: Number of mass movements within variable class X_i.

To evaluate the influence of each variable, weighting factors should have to be introduced, which compare the calculated density with the overall density in the area. The formula for the density-based area is:

$$W_{area} = 1000 \frac{Npix(SX_i)}{Npix(X_i)} - 1000 \frac{\sum Npix(SX_i)}{\sum Npix(X_i)}$$

and for the density based on number/km^2

$$W_{number} = \frac{1*10^6}{Area(X_i)} Number(SX_i) - \frac{1*10^6}{\sum Area(X_i)} \sum Number(SX_i)$$

2.3.2.3.1.1.1.1. Production of the Susceptibility Map

The weight values for the variable classes are added to produce a hazard map. With the number of input maps and different combinations of type and activity, a number of different susceptibility maps can be made. The optimal combination of variables is generally a problem, however selection of a small set of maps incorporate the most relevant variables. Two methods have been applied in literature. First method is selection of maps based on field experience in which the variables that are considered, on the basis of field experience, to be relevant for the occurrence of mass movements are selected and summed. The other method is called the stepwise map combination, adding the various input maps one by one. After the addition of another map, the resulting scores are analyzed by crossing with the map showing active landslides. The percentage of pixels with landslides and a total score larger than zero is calculated (correctly classified pixels). If this percentage decreases after the addition of another map, such a map is rejected. If the percentage increases, the map is included (van Westen, 1993). However, the sequence of this summation changes everything which is a major drawback of this method and needs to be justified.

2.3.2.3.1.1.2. Information Value Method

The use of a combination of numerical variables (such as slope angle values) and alphanumerical variables (such as lithological variables) in a statistical analysis is generally problematic. This can be solved by treating each variable class as a separate variable, which can only one of the two states: present (1) or absent (0). It can be determined whether a variable class is present or absent. The information value method can be applied both to land units as well as on a pixel basis. The hazard information method, developed by Yin and Yan (1988) is based on the following simple formula for calculating the information value I_i for variable X_i:

$$I_i = \log \frac{S_i/N_i}{S/N}$$

where:

S_i: the number of land units or pixels with mass movements and the presence of variable X_i,

N_i: The number of land units or pixels with variable X_i

S: The total number of land units or pixels with mass movements

N: The total number of land units or pixels.

The degree of a hazard for a land unit or pixel j is calculated by the total information value I_j

$$I_j = \sum_{i=0}^{m} X_{ij} I_j$$

where:

m   : number of variables,

X_{ij} : 0, if the variable X_i is not present in the land unit or pixel j and 1, if the variable is present.

For the assessment of precision of the classification, Yin and Yan (1988) presented the following equation

$$A = \frac{M_i}{N_i} \sqrt[3]{\left(1 - \frac{M - M_i}{N - N_i}\right)}$$

in which:

     A: precision of the predicted result

     N: total number of terrain units (catchments in this case) in this area

     N_i: total number of units with landslides

     M: number of terrain units predicted as unstable

     M_i: number of terrain units predicted as unstable which have landslides.

The information value method applied on a pixel basis is in fact very similar to the susceptibility determination. The only difference is that in the information value method the log value of the quotient of class density over map density is entered, whereas in the susceptibility method the difference in densities was used. The information values are always smaller than the weight values.

2.3.2.3.1.1.3. Weights of Evidence Modelling

This method was developed at the Canadian Geological Survey (Agterberg et al., 1990; Bonham-Carter et al, 1990) and was applied to the mapping of mineral potential. Sabto (1991) applied the method for landslide hazard analysis. The method consists of reducing each set of landslide-related factors on a map to a pattern of a few discrete states. In its simplest form, the pattern for a feature is binary, representing its presence or absence within a pixel. According to Bonham-Carter et al. (1990), the first step is determining the prior probability of landslides, which is given by the density of pixels with landslides within the study area.

$$P_{prior} = \frac{Npix(slide)}{Npix(total)}$$

in which

P_{prior} : prior probability,

Npix (slide) : the number of pixels with a landslide occurrence,

Npix (total) : the total number of pixels in the map

for mathematical reasons it is more convenient to use the odds (O):

$$O_{prior} = \frac{P_{prior}}{1 - P_{prior}} = \frac{Npix(slides)}{Npix(total) - Npix(slides)}$$

Considering the relationship between a binary variable map (b_i) and a landslide map (S), the following combinations are possible:

$$B_i: \quad Npix(B_i)/Npix(total)$$

$$\overline{B_i}: \quad Npix(total) - Npix(B_i)/Npix(total)$$

four combinations of B_i and S are possible in the map: $B_i \cap S$, $\overline{B_i} \cap S$, $B_i \cap \overline{S}$, $\overline{B_i} \cap \overline{S}$.

The conditional probability of choosing a pixel with a landslide, given that the cell contains pattern B_i, is:

$$P\{S|B_i\} = \frac{B_i \cap S}{B_i}$$

and the three other conditional probabilities are:

$$P\{\overline{S}|B_i\} = \frac{B_i \cap \overline{S}}{B_i} \qquad P\{S|\overline{B_i}\} = \frac{\overline{B_i} \cap S}{\overline{B_i}} \qquad P\{\overline{S}|\overline{B_i}\} = \frac{\overline{B_i} \cap \overline{S}}{\overline{B_i}}$$

According to Bayes rule:

$$P\{S|B_i\} = \frac{P\{B_i|S\}P\{S\}}{P\{B_i\}}, \quad P\{S|\overline{B_i}\} = \frac{P\{\overline{B_i}|S\}P\{S\}}{P\{\overline{B_i}\}}$$

Bonham-Carter et al. (1990) defined positive and negative weights (W_i^+ and W_i^-), which combine these conditional probabilities:

$$W_i^+ = \log_e \frac{P\{B_i|S\}}{P\{B_i|\overline{S}\}} \quad \text{and} \quad W_i^- = \log_e \frac{P\{\overline{B_i}|S\}}{P\{\overline{B_i}|\overline{S}\}}$$

In GIS the method can be implemented rather easily. It is considered as the simple crossing of a binary landslide map with a binary variable map. The four possible resulting combinations are given Table 2.5.below.

Table 2.5. The possible combinations after map crossing

LANDSLIDES	Variable Class represented as binary pattern	
	1 (present)	0 (absent)
Present 1	$Npix_1$	$Npix_2$
Absent 0	$Npix_3$	$Npix_4$

The weights of evidence can be written in numbers of pixels as follows:

$$W_i^+ = log_e \frac{\dfrac{Npix_1}{Npix_1 + Npix_2}}{\dfrac{Npix_3}{Npix_3 + Npix_4}} \qquad W_i^- = log_e \frac{\dfrac{Npix_2}{Npix_1 + Npix_2}}{\dfrac{Npix_4}{Npix_3 + Npix_4}}$$

If more binary maps are used, the weights can be added, provided that the variable maps are conditionally independent with respect to landslide occurrence. The logarithm of the posterior odds can be calculated as follows:

$$log_e O\left\{S \middle| B_1^k \cap B_2^k \cap B_3^k \ldots\ldots B_n^k\right\} = \sum_{i=1}^{n} W_i^k + log_e O_{prior}\left\{S\right\}$$

and the posterior probability as:

$$P\{S\} = \frac{O}{(1+O)}$$

The contrast C=W$^+$ -W$^-$ gives a useful measure of the correlation between the variable map and the landslide occurrence. C becomes zero when a map has a distribution which is spatially independent of the points.

The main assumption for univariate statistical methods is that the maps should be conditionally independent. To test this independence a pairwise test can be executed (Bonham-Carter et al, 1990). All possible pairs of variable maps should be evaluated separately. The pairwise test includes the calculation of observed and expected frequencies of landslides. Therefore, the maps are crossed pairwise, and the resulting cross map is then crossed again with the mass movement map. The combinations obtained from crossing two binary maps and a landslide map is given in the Table 2.6.

Table 2.6. The possible combinations after crossing of two binary maps

Landslides	possible combinations of binary maps			
	$B_i \cap B_2$	$B_1 \cap \overline{B_2}$	$\overline{B_i} \cap B_2$	$\overline{B_i} \cap \overline{B_2}$
Present	$Npix_1$	$Npix_2$	$Npix_3$	$Npix_4$
Absent	$Npix_5$	$Npix_6$	$Npix_7$	$Npix_8$

Using the weight of evidence modeling, the logarithm of the odds for each unique overlap of two variable classes, is calculated by:

$$\log_e O\left(S|B_1 B_2\right) = W_1^+ + W_2^+ + \log_e O(S)$$

$$\log_e O\left(S|B_1 \overline{B_2}\right) = W_1^+ + W_2^- + \log_e O(S)$$

$$\log_e O\left(S|\overline{B_1} B_2\right) = W_1^- + W_2^+ + \log_e O(S)$$

$$\log_e O\left(S|\overline{B_1 B_2}\right) = W_1^- + W_2^- + \log_e O(S)$$

The predicted number of pixels in each unique overlap can be calculated using:

$$m_i = P_i \ Npix_i$$

in which :

m_i : the number of predicted landslides for the overlap of two classes

P_i : The calculated probability for the overlap of the two classes

Npix$_i$: The number of pixels in each overlap (for $B_i \cap B_2$ this will be Npix$_1$ + Npix$_5$)

The conditional independence is tested with the following formula:

$$G^2 = -2\sum_{i=1}^{8} X_i \log\frac{m_i}{x_i}$$

in which

x$_i$: the number of mass movement occurrences for the overlap of two classes (for $B_i \cap B_2$ this will be Npix1)

The function G^2 has a χ^2 distribution with 2 degrees of freedom (Bonham-Carter et al., 1990). On the basis of the result of the χ^2 test the selection of the variable maps is made. The weight of evidence values are added and the posterior probability is calculated. After classification of the posterior probability, the expected number of landslide occurrences per probability class is calculated for each class and compared

with the observed number of occurrences per probability class. The expected frequency per class is given by:

$$f_{i(e)} = P_i Npix_i$$

in which

$f_{i(e)}$:   expected number of occurrences per probability class i

P_i :   the probability per class i

$Npix_i$:the area (in pixels) of probability class i.

By crossing the predictor map with the mass movement map the actual number of mass movements can be calculated, and the χ^2 test can be applied

$$\chi^2 = \sum \frac{\left(f_{i(0)} - f_{i(e)}\right)^2}{f_{i(e)}}$$

where, $f_{i(0)}$ is the observed frequency of landslides.

2.3.2.3.1.2. Multivariate Statistical Methods in Landslide Hazard Analysis

Multivariate statistical analyses of important causal factors controlling landslide occurrence may indicate the relative contribution of each of these factors to the degree of hazard within a defined land unit. The analyses are based on the presence or absence of stability phenomena within these units (van Westen, 1993).

Multivariate statistical analysis models for landslide hazard zonation were developed in Italy, mainly by Carrara (1983, 1988) and his colleagues (Carrara et al., 1990, 1991, 1992). In their applications, all relevant factors are sampled either on a large-grid basis or in morphometric units. For each of the sampling units, the presence or absence of landslides is also determined. The resulting matrix is then analyzed using multiple regression or discriminant analysis. With these techniques good results can be expected in homogenous zones or areas with only a few types of slope instability processes. When complex statistics are applied, as was done by Carrara (1983, 1988) and his colleagues (Carrara et al., 1990, 1991, 1992) or by Neuland (1976) or by Kobashi and Suzuki (1988), subdivision of the data according to the type of the landslide should be also made as well. Therefore, large data sets are needed to obtain enough cases to produce reliable results. The use of complex statistics implies laborious efforts

31

in collecting large amounts of data, because these methods do not use selective criteria based on professional experience. Multivariate statistical analyses of important factors related to landslide occurrence give the relative contribution of each of these factors to the total hazard within a defined land unit. The analyses are based on the presence or absence of mass movement phenomena within these land units, which may be catchment areas, interpreted geomorphic units, or other kinds of terrain units.

The following GIS procedures are used to evaluate multivariate statistics in landslide hazard zonation:

1. Determination of the list of factors that will be included in the analysis. As many input maps are of alphanumeric type, they must be converted into numerical maps. These maps can be converted to presence/absence values for each landunit or presented as percentage cover or the parameter classes can be ranked according to increasing mass movement density. By overlaying the parameter maps with the land-unit map, a large matrix is created.
2. Combination of the land unit map with the mass movement map via map overlay and dividing the stable and unstable units into two groups.
3. Export of the matrix to a statistical package for subsequent analysis.
4. Importation of the results per land-unit into the GIS and recoding of the land units. The frequency distribution of stable and unstable classified units is checked to see whether the two groups are separated correctly.
5. Classification of the map into a few hazard classes.

Two types of multivariate analyses have been conducted in the literature extensively, multiple regression and discriminant analyses. There exists plenty of other statistical methods, such as logistic regression or analysis of the parameter maps prior to bivariate analyses by factor analyses. However, these methods require more than entry level statistics and the data manipulation should be done very carefully, as within these methods data manipulation is not a speculative event. Although the multiple regression and discriminant analyses constitute some part of the landslide hazard analysis literature some real big drawbacks are introduced, as the data used for these analyses should have to be distributed normally, which is quite impossible when dealing with natural data. Especially when the data sets of distance to some object is used. Several normality conversion tables could have been used in order to convert the data into normal distribution such as log-log or log-normal coversions, however, these conversions do inherit some critical biases to the natural distribution of the data. Some authors have tried to exploit the data via using dummy binary variables but this had increased the complexity of the data structure and limits the flexibility of the statistical system. Examples of these dummy variables could be seen in Carrara et al. (1990,

1991, 1992) and in Chung et al. (1993). On the other hand, the use of binary logical regression, which is free of data distribution issues, are not so well exploited in the literature only few examples in the last few years are observed, such as: Atkinson and Massari (1998), Dai et al. (2001) and Lee and Min (2001). It is also seen from this fact that logical regression is quite new in this area. In the next sections, only multiple regression and discriminant analyses will be introduced, the application of logical regression and tidbits will be explained in the application chapters.

Although these techniques can be applied at different scales, their use becomes quite restricted at the regional scale, where an accurate input map of landslide occurrences may not be available, and where most of the important parameters cannot be collected with satisfactory accuracy. At large scales, different factors will have to be used (such as water-table depth, soil layer sequences and thickness). These data are very difficult to obtain even for relatively small areas. Therefore, the medium scale is considered most appropriate for these sets of techniques.

2.3.2.3.1.2.1. Multiple Regression

The most common and well-known multivariate statistical method used in earth sciences is multiple regression. It is used to correlate landscape factors and mass movements, according to the following linear equation.

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots\dots\dots + b_n X_n$$

The dependent variable Y represents the presence (1) or absence (0) of a mass movement. It can also be expressed as the percentage of a terrain unit covered by landslides. The variables X_1-X_n are the independent variables, such as slope class, geological units, etc. the symbols b_0-b_n are the partial regression coefficients. The standardized partial regression coefficients, which are the partial regression coefficients expressed in units of standard deviation, indicate the relative contribution of the independent variables to the occurrence of landslides (Davis, 1986). The following statistics are used to evaluate the result of a calculation.

R^2 : amount of variance accounted for by the model. It adjusts for the number of independent variables in the regression

SE: standard error of estimate. The square root of the residual mean square error. It measures the unexplained variability in the dependent variable.

MEA: absolute mean error. The average of the absolute values of the residuals, which is the average error one can expect in a prediction.

The use of terrain units for the sampling of variables in multiple regression analysis is welcomed with a number of problems.

1. Sampling method
2. Size of terrain unit
3. Resultant maps
4. Sample areas / Prediction areas
5. Complexity of the study areas

In order to avoid these cited problems, generally a pixel based approach is used, even in this approach the data requirements of normal distribution could not be achieved. A series of assumptions are made about the assumption of the data normality which in fact degrades the efficiency of the whole system.

2.3.2.3.1.2.2. Discriminant Analyses

A second type of multivariate analysis is discriminant analysis. The objective of the analysis is to find the best discrimination between two groups: units or pixels with and those without mass movements. The analysis results in a discriminant function:

$$D_s = B_0 + B_1 X_1 + B_2 X_2 + \ldots \ldots B_n X_n$$

where X_i are the values of the variables and B_i the calculated coefficients. Before any further analysis can be performed, the success of the formula in separating the two groups must be tested. For this purpose three tests can be used.

1. the variability between the two groups and within the groups, and the total variability of the data, are calculated. The ratio of the variability between the two groups and the variability within the groups is called the eigenvalue. It should be maximized for a good discriminant function.
2. the ratio of the variability between the two groups and the total variability is called "Wilk's λ". A small value indicates strong variation between groups and less variation within groups. A Wilk's λ of 1 indicates that there is equally great variation within groups as between groups (i.e. that the function does not discriminate)
3. the χ^2 test to determine if the two groups are significantly different.

Furthermore, as the slope stability depends on several factors acting at the same time, some efforts have been directed towards the acquisition of simply and quickly determined parameters. Stevenson (1977) using scored factors proposed a method to evaluate relative landslide risk in clayey slopes.

Discriminant analysis provides a more accurate stability assessment. A classical work using statistical techniques is that from Jones et al. (1961) on landslides in Pleistocene terrace deposits of Colombia river. A total of 160 slump-earthflow movement and additional 160 stable slopes were considered. Qualitative and quantitative factors influencing sliding were searched. A final analysis using the discriminant - function method was performed considering as influencing factors: original slope (X1), submergence percentage (X2), terrace height (X3) and groundwater (X4).

2.3.2.3.2. Knowledge Driven Methods in Landslide Hazard Analysis

2.3.2.3.2.1. Qualitative Map Combination

To overcome the problem of the "hidden rules" in geomorphic mapping, other qualitative methods based on qualitative map combination have been developed. In qualitative map combination, the earth scientist uses the expert knowledge of an individual to assign weighting values to a series of parameter maps. The terrain conditions at a large number of locations are summed according to these weights, leading to hazard values that can be grouped into hazard classes. The problem with this method is in determining the exact weighting of the various parameter maps. Often insufficient field knowledge of the important factors prevents the proper establishment of the factor weights, leading to unacceptable generalizations (Soeters and van Westen, 1996).

2.3.2.3.2.2 Favourability Functions

In order to achieve the minimum common factors of expert knowledge dependency and the information derived from the original data, the geographical database is designed under the envelope of some data and expert knowledge dependent functions. This is also used to decrease the subjectivity of expert knowledge and not to deal with the redundant information yielding from pure statistical analyses. In favourability analyses, the data layers are first divided into a number of expert designed classes such as geological, geomorphological or slope classes and etc. For data integration (numeric and alphanumeric databases) each layer is transformed into a number between α and β, where α and β are known constants such as 0 and 1 or -1 and +1. This transformation is the basic step of forming a probabilistic favourability function

from a class to the interval α and β. After completion of this stage, some previously defined decision rules are applied to let the expert earth scientist decide about the factor probability and favourability of the current situation such as standard probability measures, certainty factor, Dempster-Shafer belief method and fuzzy logic interpretation (Soeters and van Westen, 1996; van Westen, 1993).

2.3.2.3.3. Deterministic Modeling in Landslide Hazard Analysis

The methods described so far give no information on the stability of a slope as expressed in terms of its factor of safety, in order to obtain this information these kinds of slope stability deterministic models are necessary.

Despite problems related to collection of sufficient and reliable input data, deterministic models are increasingly used in hazard analysis of larger areas, especially with the aid of GIS techniques, which can handle the large number of calculations involved in determination of safety factors over large areas. Deterministic methods are applicable only when the geomorphic and geologic conditions are fairly homogeneous over the entire study area and the landslide type is simple. The advantage of these white box models is that they are based on slope stability models, allowing the calculation of quantitative values of stability (safety factors). The main problem with these methods is the degree of simplification which is required in the acceptance limits of the assumptions. A deterministic method, usually applied for translational slides is the infinite slope model. These deterministic models generally require the use of ground water simulation models. Stochastic methods are sometimes used to select input parameters for the deterministic models (Mulder and van Asch, 1988; Mulder, 1991; Hammond et al. 1992).

The result is a map showing the average safety factor for a given magnitude of groundwater depth and seismic acceleration. The variability of the input data can be used to calculate the probability of failure in connection with the return period of triggering events. Generally the resulting safety factors and probability factors should not be used as absolute values unless the analysis is done in a small area where all the parameters are well known. Normally they are only indicative and can be used to test different scenarios of slip surfaces and groundwater depths. The method is applicable only at large scales over small areas. At regional and medium scales, the required detailed input data, especially concerning groundwater levels, soil profile, and geotechnical descriptions, usually cannot be provided.

2.3.2.3.4 Landslide Frequency Analysis

The probability of mass movement occurrence at a certain place within a certain time period can only be determined when a relationship can be found between the occurrence of landslides and the frequency of triggering factors, such as rainfall or earthquakes. The most promising technique is the calculation of antecedent rainfall, which is the accumulated amount of precipitation over a specified number of days preceding the day on which a landslide occurred (Crozier, 1986).

The method is most appropriate at medium and large scales. At regional scale, it may be difficult to correlate known landslides at one location with rainfall records from a different location in the area. The spatial component is usually not taken into account in this analysis and therefore the use of GIS is not crucial, however GIS can be used to analyze the spatial distribution of rainfall.

2.3.2.4. Accuracy and Objectivity

The most important question to be asked in each landslide hazard study relates to its degree of accuracy. The terms accuracy and reliability are used to indicate whether the hazard map makes a correct distinction between landslide free and landslide prone areas. The accuracy of landslide prediction depends on a large number of factors the most important of which are:

1. accuracy of the models
2. accuracy of the input data
3. experience of the earth scientist
4. size of the study area

The context of accuracy is a fatal section in disaster management, as the wrong decision of landslide free areas will cause loss of lives, which discloses to the aim of hazard and risk assessment. This fatal section can be checked out by some statistical analyses and trying to find out the possible error component, furthermore the error component should have to be put in all of the maps produced and the knowledge that are made public accessible.

Related to the problem of assessing the accuracy of hazard maps is the question of their objectivity. The terms objective and subjective are used to indicate whether the various steps taken in the determination of the degree of hazard are verifiable and reproducible by other researchers or whether they depend on the personal judgement of the earth scientist in charge of the hazard assessment.

Objectivity in the assessment of landslide hazard does not necessarily result in an accurate hazard map. For example, if a very simple but verifiable model is used or if only a few parameters are taken into account, the procedure may be highly objective but produce an inaccurate map. On the other hand, subjective studies, such as detailed geomorphic slope stability analyses, when made by experienced geomorphologists may result in very accurate hazard maps. Yet, such a good, but subjective assessment may have a relatively low objectivity because its reproducibility will be low. This means the same evaluation made by an other expert will probably yield another result, which can have clearly undesirable legal effects (Soeters and van Westen, 1996; van Westen, 1993).

.

2.3.2.5. Evaluation of Methods via Scale Factor

Any hazard evaluation involves a large degree of uncertainty. Prediction of natural hazards such as landslides, which are caused by interaction of factors which are not always fully understood and sometimes unknown, confronts earth scientists with especially large problems. However, the use of statistics indeed will increase the accuracy of the input data, this minor improvement will reduce the degree of uncertainty in the assessment. On the other hand the use of multivariate statistics in GIS will yield in assembling factor maps, that could not be done only on statistical packages. This assemblage is routing the earth scientists, to use GIS which confronts the user to more complex and more variable dominated platforms. More onwards, the used models have to be improved by the availability of huge amount of data and availability of adequate type and method of handling. Based on this knowledge, the available methodologies in landslide hazard zonation can be classified and rated as follows in correspondence to the analysis of scale factor (Table 2.7).

Table 2.7. Classification of Methods based on scale factor (Soeters and van Westen, 1996, van Westen, 1993).(the first number indicates the feasibility 1:Low, it would take too much time and money to gather sufficient information in relation to the expected output; 2: Moderate: a considerable investment would be needed, which only moderately justifies the output; 3: good, the necessary input data can be gathered with a reasonable investment related to the expected output. The second number indicates the usefulness 1:of no use, 2: of limited use, 3: useful).

Method	Regional Scale	Medium Scale	Large Scale	Usefulness of GIS in the analysis
Landslide Distribution Analysis	2/3	3/3	3/3	Intermediate
Landslide Density Analysis	2/3	3/2	3/1	Intermediate/high
Landslide Activity Analysis	1/3	3/3	3/3	Intermediate/high
Landslide Isopleth analysis	2/3	3/2	3/3	High
Geomorphological Landslide Hazard analysis	3/3	3/3	3/3	very low
Qualitative Landslide Hazard Analysis	3/3	3/2	3/1	high
Landslide Susceptibility Analysis	1/3	3/3	3/2	high
Information Value Method	1/1	3/3	3/2	high
Weights of Evidence Method	1/1	3/3	3/2	high
Multivariate Statistical Analysis	1/2	3/2	3/2	high
Deterministic Landslide Hazard Analysis	1/1	1/2	2/3	high
Antecedent Rainfall Analysis	2/2	3/3	3/2	Very low

CHAPTER 3

GEOLOGY OF THE ASARSUYU CATCHMENT

3.1. Geology

3.1.1. Regional Geology and Previous Works

The study area is located in northern Central Anatolia, within the North Anatolian Fault Zone (NAFZ), and it is mainly comprised of Paleozoic metamorphic/intrusive basement and its covering Mesozoic and Tertiary flyschoidal deposits (Figure 3.1) in the Pontide Tectonic unit of Ketin 1966.
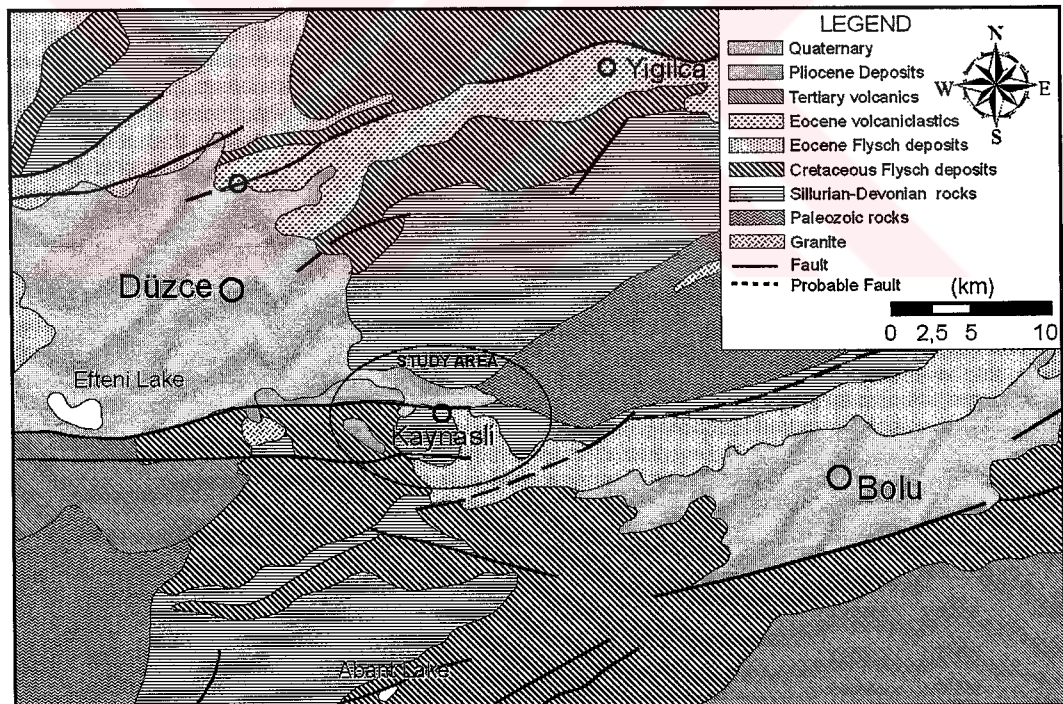
Figure 3.1. Regional geological map of the study area and its environs (modified from Tokay,1964).

The study area is hosting the main state highway connecting the capital city Ankara to the biggest metropolitan area İstanbul. The presence of the Anatolian highway and the problems associated with the construction of Bolu Tunnels attracted the attention of the scientists and hence a number of studies had been carried out concerning the geology, NAFZ and engineering geological problems of this region. Hereforth, a brief summary will be given about the previous works of the study area in table form (Table 3.1). The citations would be made in the appropriate parts in the following sections with detailed explanations when needed.

Table 3.1. Previous studies performed in the study area and its near vicinity

Research Field	Researcher	Research Area	Remarks
Geology	Blumenthal (1948)	Regional	First definitions of tectonic units
	Ketin (1955)	Akçakoca-Düzce	First sub-division of Paleozoic Massifs
	Uysallı (1959)	Bolu-Merkeşler	Coal Resources
	Abdülselamoğlu (1959)	Mudurnu-Göynük	Definition of Paleozoic and mesozoic formations
	Ketin (1967)	Bolu-Gerede-Mengen-Yığılca	Paleozoic Units
	Batum (1968)	Northern Slopes of Asarsuyu Valley	First defined the age of the conglometares that overlie Bolu massif as Sillurian
	Gözübol (1978)	Mudurnu-Dokurcun-Abant	Geology/NAFZ
	Görmüş (1980)	Yığılca	Stratigraphy/Tectonics
	Yılmaz et al. (1981)	Abant-Dokurcun	Geological evolution, relations of metamorphic rocks and the ophiolitic rocks in area.
	Görmüş (1982a)	Yığılca	Stratigraphy
	Görmüş (1982b)	Yığılca	Tectonic and evolutionary model
	Kaya (1982)	Ereğli-Yığılca-Bolu-Mengen	Stratigraphy-Tectonics
	Kaya & Dizer (1981-1982a)	Mengen	Coal Resources
	Kaya & Dizer (1981-1982b)	North Bolu	Stratigraphy of Mesozoic and Tertiary sequences
	Cerit (1983)	Mengen	Geology
	Serdar & Demir (1983)	Bolu-Mengen-Devrek	Petroleum Resources
	Öztürk et al. (1984)	Abant-Yeniçağa	Stratigraphy of north and south of NAFZ

Table 3.1. (continued)

Geology	Aydın *et al.* (1987)	Çamdağ-Sünnicedağ	Complete stratigraphic outline.
	Kaya *et al.* (1986)	Yığılca	Stratigraphy of Mesozoic and Tertiary sequences
	Cerit (1990)	Bolu Massif	Geology and tectonics
	Erendil *et al.* (1991)	Bolu Massif	Economical resources
	Yalçın & Cerit (1991)	Bolu Massif	Investigated the metamorphism by using clay mineralogy
	Koral *et al.* (1994)	Asarsuyu	Microfabric study in Paleozoic rocks
	Gedik & Alkaş (1996)	Bolu Region	Carbondioxide potential
	Sözen *et al.* (1996)	Düzce and Devrek	Geochemistry/Economical resources
	Güler (1999)	Bakacak	Geology of gypsum
	Ustaömer & Rogers (1999)	Bolu Massif	Geochemistry – Evolution model
Engineering	Orkan *et al.* (1977)	Highway Route	Elementary landslide susceptibility map
	Canik (1980)	Bolu	Hydrogeology
	Aktimur *et al.* (1983)	Bolu	Landuse pattern and NAFZ related problems
	Dalgıç (1994 a,b)	Bolu Mountain Highway Pass	Engineering geology of Highway pass and Bolu Tunnel
	Dalgıç et al(1995)	Asarsuyu	Stability and nature of the Yumrukaya landslide
	Dalgıç & Gözübol (1995)	Bolu Tunnel	Stability problems in Bolu Tunnel
	Astaldi (1995)	Asarsuyu	Enginering geology of the Asarsuyu region for Bolu Tunnel construction
	Dalgıç (1997)	Bolu Tunnel	lithology and the fracture pattern in the Bolu Tunnels
	Simşek & Dalgıç (1997)	Düzce	consolidation properties of the clays at Düzce
	Aydan & Dalgıç (1998)	Bolu Tunnel	Prediction of Deformation inside the Bolu tunnel
	Dalgıç (1998,a)	Asarsuyu	slope stability problems in the Asarsuyu valley
	Dalgıç (1998,b)	Asarsuyu	Selection of crushed Rock Quarries

Table 3.1. (continued)

<table>
<tr><td rowspan="4">Engineering</td><td>Işın (1999)</td><td>Bolu Mountain Highway Pass</td><td>Investigated two landslides in the highway by insitu instumentation</td></tr>
<tr><td>Unterberger and Brandl (2000)</td><td>Bolu Tunnel</td><td>Deformations in the Bolu Tunnels after the 12. Nov 1999 Düzce earthquake</td></tr>
<tr><td>Aydan et al. (2000)</td><td>Düzce</td><td>Interdisciplinary study concerning the geological, engineering geological, seismological and geotechnical aspects of Düzce earthquake</td></tr>
<tr><td>Sucuoğlu et al. (2000)</td><td>Regional</td><td>Engineering report regarding both the Marmara and Düzce earthquakes</td></tr>
<tr><td rowspan="15">NAFZ(Tectonics)</td><td>Ketin (1969)</td><td>Regional</td><td>First definition</td></tr>
<tr><td>Ambraseys (1970)</td><td>Regional</td><td>Characteristic features</td></tr>
<tr><td>Tokay (1973)</td><td>Gerede-Ilgaz</td><td>Characteristics</td></tr>
<tr><td>Gözübol (1978)</td><td>Mudurnu-Dokurcun-Abant</td><td>Structural properties of the NAFZ</td></tr>
<tr><td>Şengör & Canıtez (1982)</td><td>Regional</td><td>Characteristics + evolutionary model</td></tr>
<tr><td>Öztürk et al. (1984)</td><td>Abant-Yeniçağa</td><td>Paleo and Neo-tectonic structures</td></tr>
<tr><td>Nurlu(1993)</td><td>Bolu-Sapanca</td><td>RS</td></tr>
<tr><td>Neugebauer (1994)</td><td>Abant</td><td>Closing-up Structures-effect of bends in NAFZ</td></tr>
<tr><td>Astaldi (1995)</td><td>Bolu Tunnel – Asarsuyu valley</td><td>Pattern of faulting in the area, also deformation in the area gathered by GPS measurements and evaluated together with microseismic data.</td></tr>
<tr><td>Neugebauer (1995)</td><td>Adapazarı – Bolu</td><td>Kinematics of NAFZ</td></tr>
<tr><td>Şaroğlu et al. (1995)</td><td>Gerede-Eskipazar</td><td></td></tr>
<tr><td>Neugebauer et al. (1997)</td><td>Abant-Sapanca</td><td>Overstep in NAFZ</td></tr>
<tr><td>Akyüz et al. (2000)</td><td>Düzce</td><td>Slip distributions of Düzce Earthquake</td></tr>
<tr><td>Taymaz (2000)</td><td>Gölcük-Sapanca-Düzce</td><td>Seismotectonics</td></tr>
</table>

3.1.2. Stratigraphy

The basement of the Asarsuyu catchment is composed of pre-Devonien Yedigöller Formation, an assemblage of metadiorite-metagranite and amphibolite–amphibolite gneiss (Figure 3.2). Following in the sequence İkizoluk Formation and its Çatak member forms the Devonien which tectonically overlies the underlying Paleozoic

basement. Upper Cretaceous is characterized by Abant Complex and Elmalık Granite, which again have a tectonic contact with underlying sequence. Starting from Upper Cretaceous to Pliocene a carbona-clastic sequence (flysch) is seen in the area, which is represented by Atyayla, Bayramışlar, Fındıcak and Apalar Formations. Plio-Quaternary period is represented by Asarsuyu formation, which is a combination of terrace conglomerates and alluvial fill deposits in the Asarsuyu valley itself. Finally, Quaternary alluvial fill and slope debris cover unconformably the whole sequence in the Asarsuyu catchment.

3.1.2.1. Yedigöller Formation

Yedigöller Formation is the core of the Bolu massif and composed of Paleozoic magmatic rocks. The formation has widespread outcrops in the eastern and northern slopes of the Asarsuyu catchment (Figure 3.3) The formation is first named by Aydın et al. (1987).

The main lithologies of the formation are amphibolite, gneiss (metagranite), metadiorite and meta-quartzdiorite. Also very small unmappable aplite, andesite, basalt and diabase dikes are common. Some of these are mylonitized by regional paleo and neotectonic events. The unit is in general affected from the regional metamorphism, and show cataclastic deformation in the vicinity of the fault zones. This formation can be divided into two as one having greenschist metamophism and having no metamorphism. The greenschist facies of this formation generally have a medium degree of weathering resulting in a thin residual soil cover (1 - 3 meters).

The lower boundary of the formation cannot be seen in the catchment; however its upper boundary is tectonic with the overlying Devonian İkizoluk Formation, which is not observed in the study area.

This formation comprises the southwest part of the Bolu Massif. Blumenthal (1948) assigned Paleozoic age to the metamorphics in the area. Ketin (1967) indicated that the age of the assemblage of amphibolite, gneiss and basic rock in Bolu-Gerede-Mengen-Yığılca area is Cambrian. Erendil et al. (1991) assigns the name Bolu Granitoid to the formation and cites that the age should be Pre-Late Ordovician. While Canik (1980), Aydın et al. (1987) and Cerit (1990) suggest Pre-Cambrian age. Based on these literature arguments, the age of the Yedigöller Formation is assigned as Pre-Devonian.

System	Stage	Formation	Member	Lithology	Description
Cenozoic	Quaternary	Alluvium			active alluvium in the Asarsuyu valley, the terrace deposits and the slided masses
					— Unconformity —
	Pliocene	Asarsuyu			alternation of clayey silt, sandy silt and sand with cobbles
					— Unconformity —
	Eocene	Çaycuma			alternation of turbiditic sandstone and siltstone, calcareous mudstone, mudstone and marl with gypsum intercalations
	Paleocene				
Mesozoic	Cretaceous	Buldandere	Findiklidere		beige, white turbiditic limestones, greenish grey sandstone, siltstone, marl and limestone alternations, greenish purplish conglomerates, mudstones and tuffs.
					— Unconformity —
Paleozoic	M.Sillurian -M.Devonian	Aksudere			phyllites, shale, recrystallized limestone, dolomitic limestone and marls
	U.Ordovisien -L.Sillurian	Kocadere			massive to thick bedded purplish grey conglomerates and sandstones
					— Tectonic contact —
	Pre-Devonian	Yedigöller			amphibolite, gneiss (metagranite), metadiorite, meta-quartzdiorite

aplite, andesite, basalt diabase dikes |

Figure 3.2. The generalized columnar section of the study area, (not-to-scale) (modified from Erendil *et al.* ,1991).
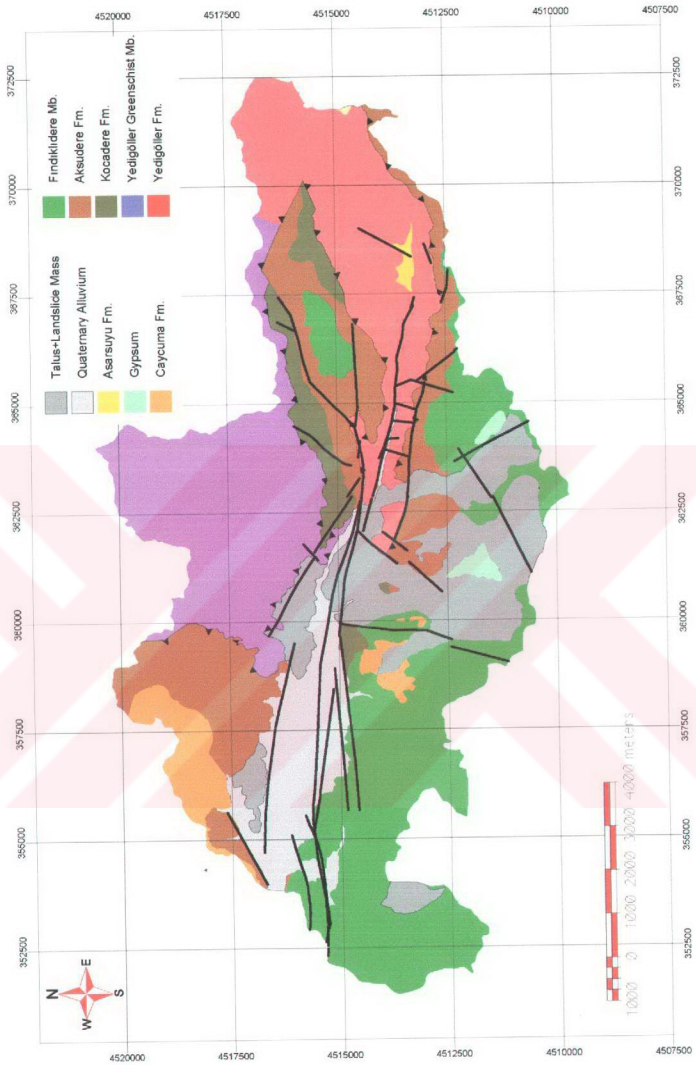
Figure 3.3. The geological map the study area (modified from Erendil *et al.* ,1991).

3.1.2.2. Kocadere Formation

This formation generally crops out as a rim to the Yedigöller Formation. The outcrops of this formation is seen only in the eastern part of the catchment at northern slopes of Asarsuyu river. The formation is defined by Erendil et al. (1991).

Kocadere Formation is composed of massive to thick bedded purplish grey polygenetic conglomerates and sandstones (greywacke). Limited mudstone and siltstone bands are also observed. Conglomerates are poorly sorted, and their components are derived from underlying metamorphics, and magmatic rocks, the matrix is composed of chlorite, quartz and serrisite (Erendil et al., 1991).

The lower boundary of the formation is tectonized and obscured. Its upper boundary is conformable with the Aksudere Formation.

The equivalent of this formation, the Kurtköy Formation of Istanbul (Kaya, 1973) is given an age of Ordovician. The other possible equivalents of this formation, "Hamzafakılı sandstones" (Tokay, 1952), Işığandere Formation (Görmüş, 1982 a, b) and "purple arkoses and conglomerates" (Batum, 1968) are aged as Late Silurian. Based on these ages, Erendil et al. (1991) assigns Late Ordovician-Early Silurian age to Kocadere Formation.

3.1.2.3. Aksudere Formation

Aksudere Formation is seen in the northern slopes of Sarıçökek village and at the eastern parts of the catchment at southern part of Asarsuyu river near Aşağıbakacak settlement and in between the two major landslides near Bakacak village. The formation is defined by Erendil et al. (1991).

Aksudere formation is represented by phyllites, shale, recrystallized limestone, dolomitic limestone and marls. The phyllites show significant schistosity and the main constituents are quartz, serrisite muscovite, chlorite, epidote and feldspars. They also host some embedded massive recrystallized limestone blocks. Phyllites and recrystallized limestones alternate with shales, siltstones, sandstones, dolomitic limestones and marls as thin to medium beds. The uppermost levels of this formation is characterized by the dominance of recrystallized limestones. The formation forms a gentle rolling topography, as it is composed of easily erodible lithologies. The main colors observed in this formation are dark grey, beige, bluish-greenish grey and brown.

The lower boundary is conformable with Kocadere Formation and its upper boundary is also conformable with Kırdoruk Formation. However, the Kırdoruk Formation does not outcropping in the Asarsuyu catchment. According to Erendil et al. (1991), the observed thickness of Aksudere Formation is about 1500-2000 meters.

Görmüş (1982, a) stated that the equivalent of lower facies of Aksudere Formation is Kocadere formation which is Lower Devonian. The Devonian aged Kartal formation of Kaya (1973) can also be correlated with Aksudere formation. Dalgıç (1994, a,b) grouped the Kocadere and Aksudere formations under İkizoluk formation and assigned an age of Devonian. Furthermore, he subdivides the uppermost section of Aksudere formation, the recrystallized limestones as Çatak Member in İkizoluk formation. Erendil et al. (1991) assigned an age of Middle Silurian to Middle Devonian.

3.1.2.4. Buldandere Formation

Erendil et al. (1991) uses this formation name to cover all of the Upper Cretaceous-Paleocene cover units, that overlie the Paleozoic Bolu Massif.

3.1.2.4.1. Fındıklıdere Member

Fındıklıdere member comprises the middle levels of Buldandere Formation outside of the Asarsuyu catchment. The member is first defined by Erendil et al. (1991). Fındıklıdere member has widespread outcrops in the southern slopes of the catchment.

The lithology of the Fındıklıdere member consists of beige-white turbiditic limestones, greenish grey sandstone, siltstone, marl and limestone alternations, greenish purplish conglomerates, mudstones and massive to medium to thick bedded gypsums. The gypsum occurrences are generally observed near the summit of Bolu Mountain and near Dipsizgöl and Yukarıaçma villages. There are also some rock quarries opened for gypsum production in these localities. Gypsums outcrops have significant karstic dissolution features, such as siphons going up to tens of meters in depth. The association of these layers resembles an Upper Cretaceous flyschoidal sequence.

The lower boundary of the member in the study area is unconformable with the Aksudere Formation, while its upper boundary is conformable with Eocene Çaycuma Formation. The thickness of the formation is estimated by Erendil et al. (1991) as 600 meters.

For the age of this member Erendil et.al. (1991) give a long list of fossils and assign Campanian - Ilerdian age. Similar formations are named and aged as: Cretaceous Flysch (Abdülselamoğlu, 1959), Gökveren Formation – Eocene (Gözübol, 1978 and Yılmaz et al., 1981), Sarıkaya Formation – Upper Cretaceous/Paleocene (Görmüş, 1980), Akveren Formation – Campanian/Lower Paleocene (Aydın et. al.,1987), Fındıcak Formation – Upper Cretaceous/Paleocene (Dalgıç, 1994, a,b).

3.1.2.4.2. Çaycuma Formation

This formation seems like the continuation of the cover units of the Bolu Massif and Fındıklıdere member. It crops out at south of the Kaynaşlı town at Akmeşe and Türbe hills, also in the north of Yörükler village. The formation is first named by Saner et al. (1979). The main lithologies of the formation is alternation of turbiditic sandstone and siltstone, calcareous mudstone, mudstone and marl with gypsum intercalations.

The lower boundary of this formation is conformable with Fındıklıdere member, while its upper boundary is not seen in the Asarsuyu catchment. The age of this formation is assigned by Erendil et al. (1991) as Eocene and is supported by both literature and fossil records. Dalgıç (1994) defined another member in the Asarsuyu catchment as Açma member, for which the age of Late Eocene was assigned and the dominant lithologies are defined as alternations of argillaceous limestones, gypsums, calcareous mudstones and clayey gypsums, which are similar to Fındıklıdere member of Erendil *et al.* (1991).

3.1.2.5. Asarsuyu Formation

Minor differences and the spatial orientation of the outcrops led Dalgıç (1994, a,b) to differentiate this formation from Quaternary alluvium. The Asarsuyu Formation crops out near Elmalık village and near the Asarsuyu portal of the Bolu Tunnels. Main lithologies of this formation include alternation of clayey silt, sandy silt and sand with cobbles.

The Lower boundary is unconformable with the older formations and its upper boundary is again unconformable with Quaternary alluvium. The age of Plio-Quaternary is based on palynological data by Astaldi (1990).

3.1.2.6. Quaternary Deposits

The Quaternary deposits are characterized by the active alluvium in the Asarsuyu valley, the terrace deposits in the northern margins of the Kaynaşlı plain and the slided mass of the to huge landslides in the southern slopes of Asarsuyu valley.

3.2. Tectonism of the Asarsuyu Catchment

The Düzce-Bolu region is one of the most tectonically active regions of Turkey. During complex tectonic evolution of the area, the catchment had been subjected to different regimes and different deformations which are reflected in the current outcrops and topography. The active faults, belonging to the North Anatolian Fault Zone, and the older thrust faults are the major large scale structural features and form the main tectonic elements of this catchment. The thrust faults are believed to act in the compressive regime after Late Cretaceous and ceased in Late Miocene (Yılmaz et al. 1981; Dalgıç, 1994) and classified as Paleotectonic features. The deformation of the region took the form of lateral strike-slip faulting after Late Miocene period in the Neoectonic regime. The deformation is still active as the North Anatolian Fault Zone and its associated faults take the act in hand and reshape the catchment (Aydan et al., 2000). During aerial photographical interpretation studies these faults are mapped and digitized. Due to the scope of this study both the thrusts and other active faults are treated together, as large discontinuities in the catchment.
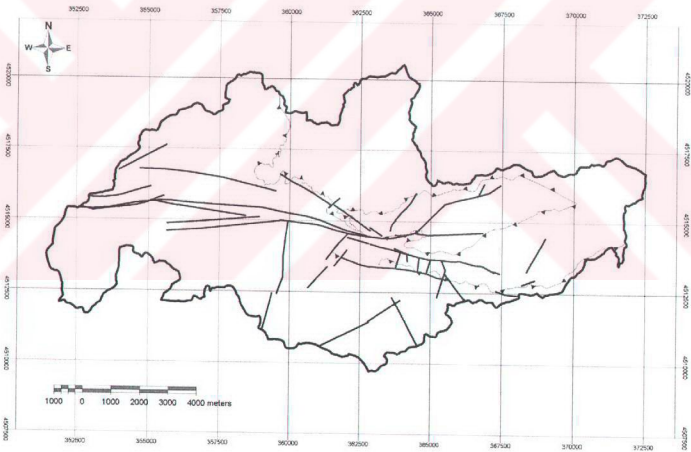
Figure 3.4. Lineament map of the Asarsuyu catchment

The earthquake activity of this region is directly controlled by the presence and activity of North Anatolian Fault zone and its associated fault segments. The North Anatolian Fault is the northern margin of Anatolian block that is escaping from the collusional zone between Arabian and Eurasian plates into the Eastern Mediterranean Sea (Şengör et al., 1985). The North Anatolian Fault is one of the most well known and most studied faults in Turkey, which has been defined some 50 years ago. The length of this fault is approximately 1200 kilometers and it is not a single fault, which in turn, worths to consider it as a fault zone (Sucuoğlu et al., 2000). The width of this zone is about 1 kilometer is widening up to 5 kilometers in the west (Demirtaş and Yılmaz, 1996). The total displacement made by this fault is reported as 20-25 kilometers by Neugebauer (1994) at Abant lake environs. Global positioning system studies (GPS) have been ongoing in the area since 1988 and the lateral slip on the Düzce fault is estimated as 7.5±1.5 mm/year (Astaldi, 1995)

The main fault strand of the North Anatolian Fault splays into two strands in the west of Bolu; the southern strand goes through the Lake Abant and then the Mudurnu valley. The northern strand steps to the north, forming the southern boundary of the Düzce basin being called the Düzce fault. In the west end of the Düzce basin, thereafter it is named the Hendek fault. Şaroglu et al. (1987, 1992) who published "Active Faults of Turkey" showed no minor connecting fault between the main trace of the North Anatolian fault and the Düzce fault. Barka and Erdik (1993) reported that there are two secondary faults, the Elmalık and Aşagı Bakacak faults that occured between the main trace and the Düzce fault as connecting faults.

3.3. Seismicity of the Asarsuyu Catchment and Environs

A total of 623 earthquakes, that are greater than 5 in instrumental magnitude have been recorded in the NAFZ in the period of January 1900 to March 2001 (Figure 3.5) having a total of 98 quakes greater than 6 in magnitude. However, the seismic evaluation of the whole fault zone is out of scope in this study, so the concentration will be turned on the study area and its vicinity. In order to compare the activity in the west and middle parts of the NAFZ and the rest of Turkey the earthquakes of the western/middle is counted and shown in Figure 3.6. 151 earthquakes out of 623 are counted, which represent nearly 25% of total earthquakes in Turkey greater than 5, in the study area. When zoomed into the study and environs, there are a total of 2596 earthquakes, of which 410 of them are greater than 3.5 in magnitude. Their spatial locations with their magnitude distributions are shown in Figure 3.7. Four earthquakes greater than 7 occurred in the area, chronologically 26-05-1957 Abant, 07-22-1967 Mudurnu, 17-08-1999 Marmara and 12-11-1999 Düzce earthquakes. The surface

rupture pattern of the 12-11-1999 Düzce earthquakes is shown in Figure 3.8. Furthermore some small landslides have occurred in the are immediately after the earthquake, some examples of these are presented in figure 3.9.

The historical records show that on the western zone of the NAFZ, 3 major earthquakes are recorded (Ambraseys and Finkel, 1995). The 25-05-1719 earthquake occurred east of Marmara sea resulted in 6000 casualties, and 4/5 of İzmit city is demolished. The second quake occurred at 2-09-1754 in the vicinity of İzmit Bay and approximately 1000 casualties was recorded. A third quake was recorded again in the east of Marmara Sea affecting the whole coast line of the Marmara region at 22-05-1766 resulting in 5000 casualties.
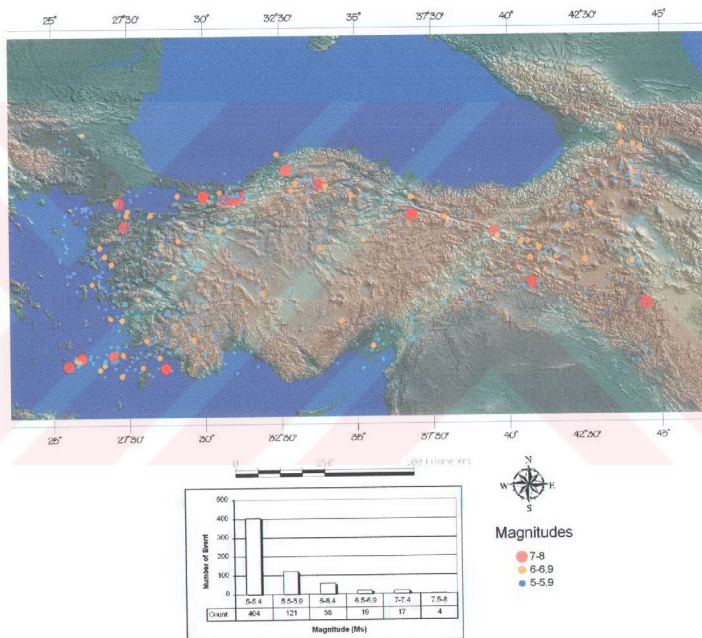
Figure 3.5. Epicenter locations of Turkey greater than 5 in magnitude (Magnitudes in Ms)

Figure 3.6. Epicenter locations of mid-west NAFZ greater than 3.5 in magnitude
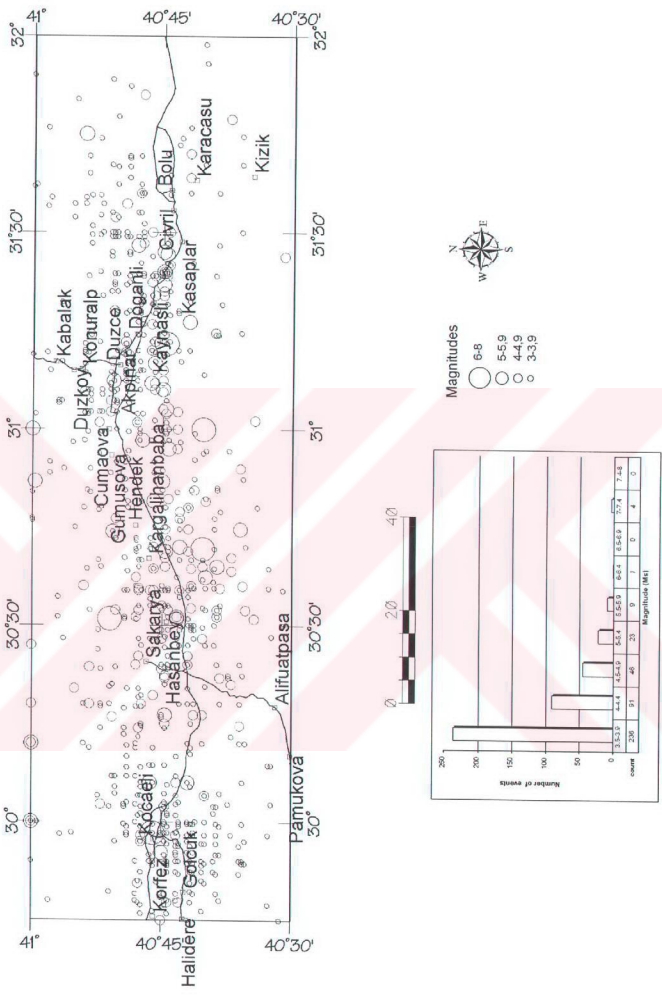
Figure 3.7. Earthquakes occurred in the nearby of the study area.
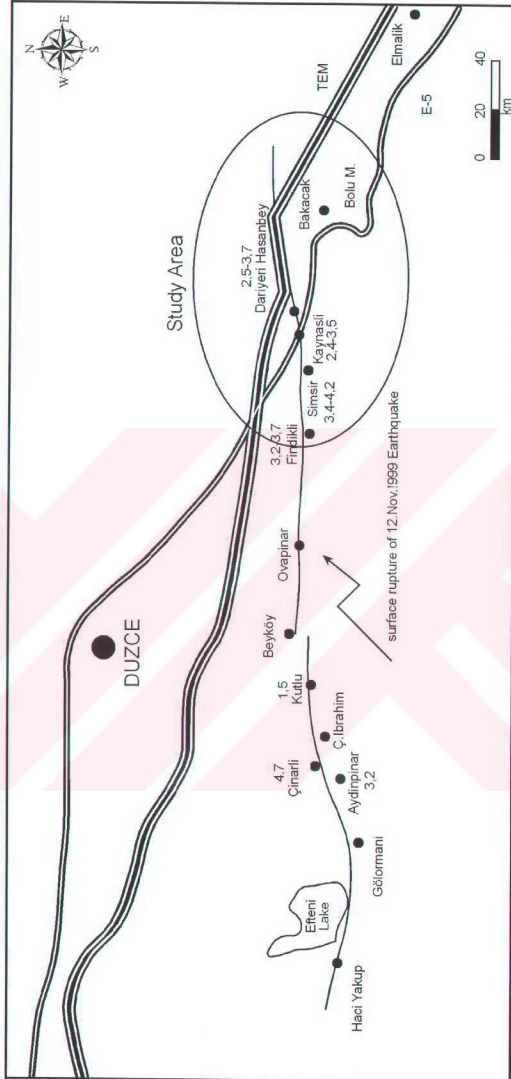
Figure 3.8. Surface ruptures of the 12 November 1999 Düzce earthquake (Aydan et al., 2000), Displacements are in meters.

Figure 3.9. Some examples of landslides occurred after 12 November 1999 Düzce earthquake.

CHAPTER 4

INPUT DATA AND DATA PRODUCTION

In this section the sources of input data and the steps in input data production are explained, as the data entry and production is the most cumbersome and time consuming steps of any kind of a Geographical Information System. The brief analysis of the distribution of the data produced are given in this chapter, however, the detailed analysis of the landslide database and the parameter maps are left to be explained in the next chapter. All of the input data and the produced data sets are valid for the common landslides except the two huge landslide bodies in the study area.

4.1. Geology

4.1.1. Data entry

The geological map of the region is compiled from the available reports, publications, dissertation thesis and an analogue map is prepared. The compiled analogue map is transformed into digital image, via digitizing and editing using TNT MIPS. A database concerning the lithology names and the ages given by the previous researchers are attached. Polygon topology is also built, validated and attached as an internal table (Figure 4.1). The resultant map is called LITHOMAP.

In the compiled geological map 11 lithologies exist. Even though, there are 7 units in the generalized columnar section in Figure 3.2, some of the lithological units show significant heterogeneity, so that they are classified as different lithological units. For example Yedigöller Formation is divided into two as one showing greenschist metamorphism and the other one not; Gypsum occurrences in the Fındıklıdere Member is separated, also the Quaternary alluvium has a distinct talus/landslide deposit unit which has to be separated from the normal valley filling alluvium. As a result, the post-classified 11 units are covering a total of nearly 1196.5 square kilometers. The largest areas are represented by Fındıklıdere member (22%), Yedigöller Fm. (16.67%), Aksudere Fm. (15.24%), Greenschist facies of Yedigöller (15.17%) of which in total

represents nearly the 70% of the area. The remaining 7 units represent only the 30% and within this 30% nearly half of it is represented by the combination of Quaternary terrace deposits and the two large landslides in the area (Figure 4.2)

Figure 4.1. LITHOMAP and the attached topological table of the study area.

Figure 4.2. The area distributions of LITHOMAP.

Following this, the faults of the area are compiled from literature In addition to this, the photo lineaments are merged to create a final fault map. For the photo lineaments Landsat TM 5 image of 1993, the stereo panchromatic aerial photographs of 1994 (1:25.000 scale), 1984 (1:15.000 scale), 1972 (1:25.000 scale) and 1952 (1:35.000 scale), and the digital elevation model of 1994 with 25-meter grid spacing are used. The snapshots of the satellite remote sensing and elevation data will be presented in following sections. The thrusts, lineaments, strike slip faults and earthquake faults/ruptures are considered as potential water sources and triggering sources so they are all together grouped as photo lineaments of the Asarsuyu catchment without any genetic discrimination. The resulting map is called the FAULTMAP including all of the mapped faults and photo-lineaments in the area (Figure 4.3). 126.284 kilometer length of fault lines are observed in the study area representing 118 fault line segments. The average length of fault-lines is 1,079.35 meters. The directional analysis is done based on "Weighted Segments" method in which the method uses the direction of each line segment and creates the Rose diagram from the total length of all segments in each direction. (Figure 4.4.a). The main directions are generally confined to in east-west trend, which is also conformable with the main trend of the North Anatolian Fault zone in the study area. The non-weighted rose diagram of the fault lines also has a major trend in E-W direction, while another small trend is seen as N-S. The N-S trending fault lines comprise shorter fault lines than those of E-W trending ones (Figure 4.4.b).
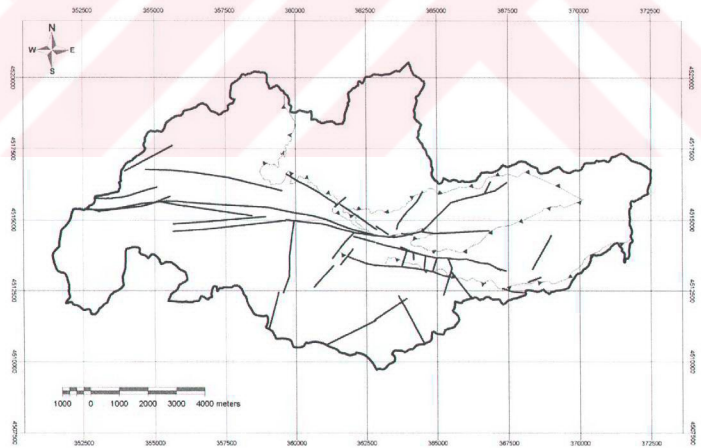
Figure 4.3. The FAULTMAP of Asarsuyu catchment.
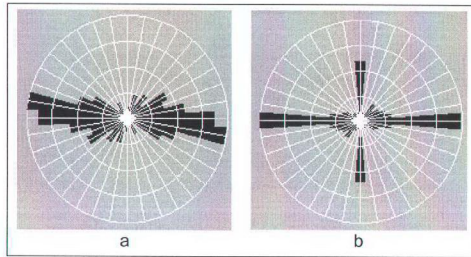
Figure 4.4. The rose diagram of FAULTMAP; a. weighted, b. non-weighted.

4.1.2. Input map generation

The LITHOMAP is used without any post-processing as it is, only a vector topology validate process is carried out to rebuild and check the polygon topology and integrity of the vector map. This is also necessary to assess the "one-to-one implied" attachment style of the database attached. However, the FAULTMAP is transferred into raster format, to a distance raster, as to represent the distance of every pixel to the nearest line in the map. This procedure calculates not the aerial orthogonal distance (oD), but the true surface distance (tD) (Figure 4.5) with the aid of a detailed Digital Elevation Model. The calculated distances are assigned as a 16-bit digital value to the every pixel in the raster image produced (Figure 4.6). The distance raster map is called DISTFAULT. The DISTFAULT to fault-lines input parameter map shows a logarithmical distribution as the frequency decreases when the distance from the fault lines increases.
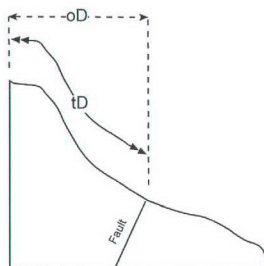
Figure 4.5. The distance calculations in distance raster map production.
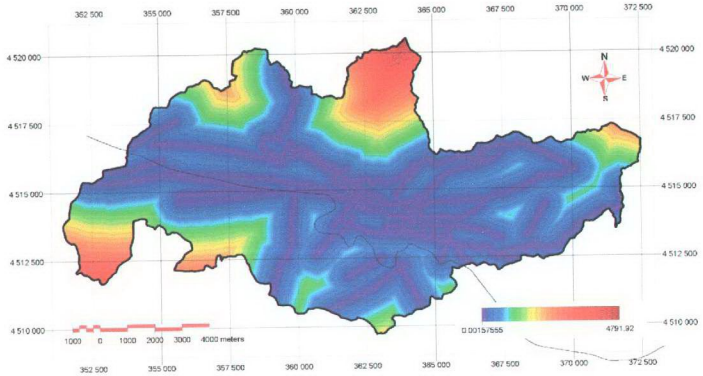
Figure 4.6. Distance raster image (DISTFAULT) produced from FAULTMAP

Another input map produced from FAULTMAP is the lineament density in 1-kilometer search radius. In order to calculate the density the number of occurrences of fault lines in a radius of 1 kilometer search distance is calculated for every pixel of 25 meters. A TNT MIPS script is written for this calculation. This density map calculated for each pixel (Figure 4.7) is called FAULTDENS. The maximum value encountered in the FAULTDENS map is 768 and the minimum is 0 meters. The distributions mean is 195.9 having a standard deviation of 154.6
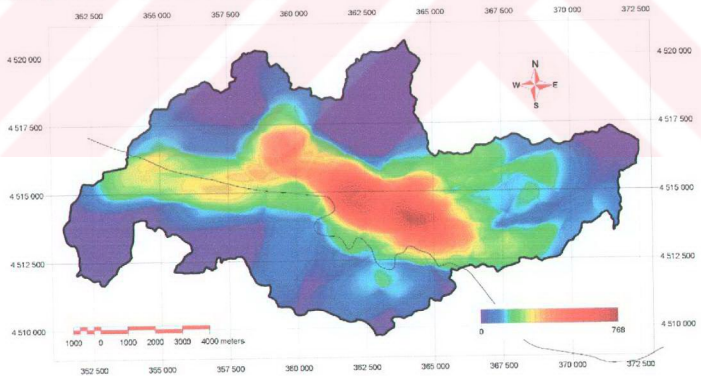
Figure 4.7. Fault density (FAULTDENS) of Asarsuyu Catchment

61

4.2. Elevation

4.2.1. Data entry

The digital elevation data of Asarsuyu catchment is gathered from General Command of Mapping Turkish Army. This data is composed of contour lines and points with elevation information in ARC/INFO E00 format (Figure 4.8), and then converted into TNT-MIPS vector format. The previously built line topology concerning the elevation information is rebuilt and validated. This map is called the CONTOURMAP.

Figure 4.8. The color-coded CONTOURMAP.

4.2.2. Input map generation

Rather than using a discrete elevation map such as the imported contour lines, it is more advantageous to work with a continuous map, regarding this advantage, the contour data is converted to a color-coded continuous map (Digital Elevation Model – DEM) by "*profiling algorithm*" in TNT-MIPS (Figure 4.9a). The Profiles method uses pairs of input elevation values on opposite sides of each output raster cell and performs a multi-directional linear interpolation procedure to create a surface raster. Edge cells are processed first; the search for cell pairs is restricted to cells parallel to the edge. For other cells the process searches in eight different directions and uses the closest pair of values (including edge cell values) to assign an interpolated value to an output cell. For all of the parameter maps the resolution of the DEM produced will create a reference layer and here forth 25 meters is taken as the working resolution of this project. Hence,

62

all of the raster inputs are fixed into 25 meters spatial resolution and the DEM is used as reference raster for these.

In order to increase visual interception of the DEM it has been converted into a relief map and presented in Figure 4.10. The produced DEM (ELEVMAP) will be used as the elevation input data for the elevation attributes of landslides.
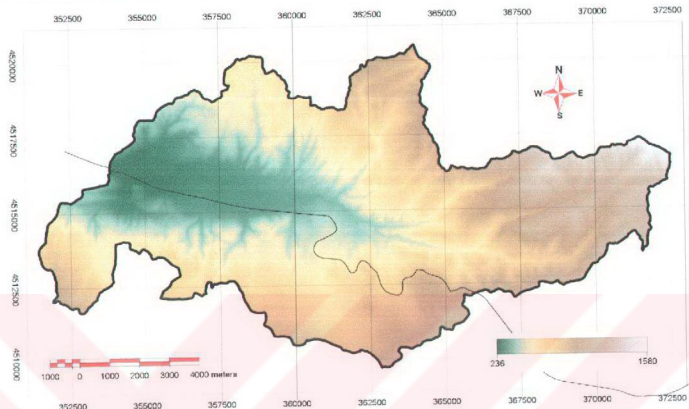
Figure 4.9. Color coded DEM of Asarsuyu catchment
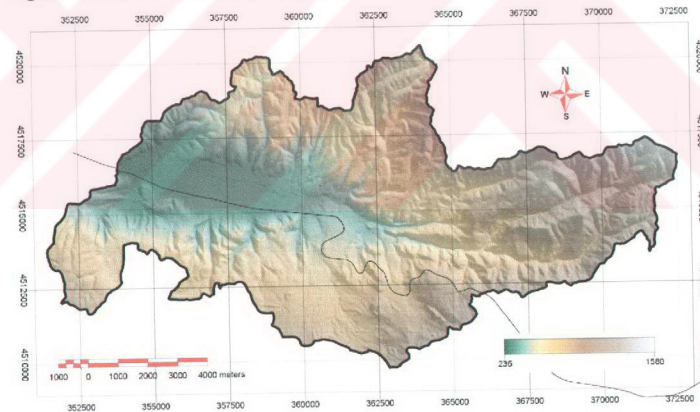
Figure 4.10. Color draped relief model of Asarsuyu catchment (illumination 045°, vertical exaggeration x3).

Following the accurate modeling of the topography in the Asarsuyu catchment, a watershed analysis is carried out to extract the microcatchments, the ridge points and the drainage-lines (Figure 4.11). The process begins by evaluating the elevation raster for depressions and constructs watershed polygons based on the depressions recognized. A vector object is generated comprising polygons that encompass the watersheds and pour point locations. The watershed process uses the Deterministic-8 (D8) algorithm (Jenson and Domingue, 1988) for flow path determination. This algorithm computes the terrain slopes between the central cell and each of its eight neighbors. Flow direction is then defined as the direction to the neighbor, either adjacent or diagonally, with steepest downward slope. The extracted drainage-lines are compared with the actual streams in the topographical sheets and corrected, if necessary. Also the Strahler order of each stream segment is assigned into a separate table manually to be used in the generation of other drainage dependent parameter maps and to assess the drainage conditions in the area.

The distances of every pixel regarding the drainage-lines are calculated (DISTDRAINMAP) and presented in Figure 4.12 The minimum distance of pixels is 1 meter and the maximum is 452.95 meters. The distribution has a mean of 97.56 with standard deviation of 73.14.
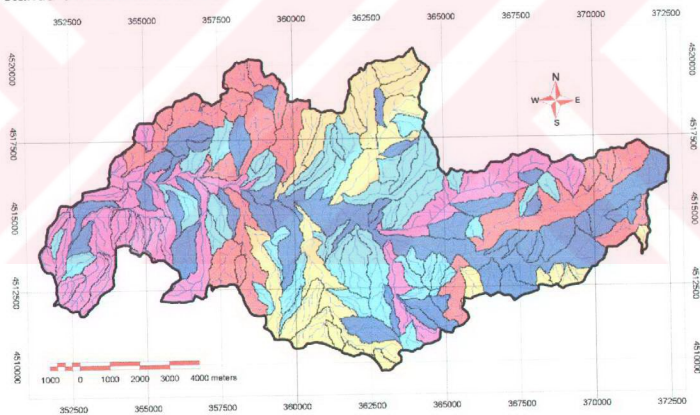
Figure 4.11. The drainage system of Asarsuyu catchment, including the microcatchments (MICROCATCHMAP), drainage-lines (DRAINAGEMAP), and ridges (RIDGEMAP). (Color coding is arbitrary).
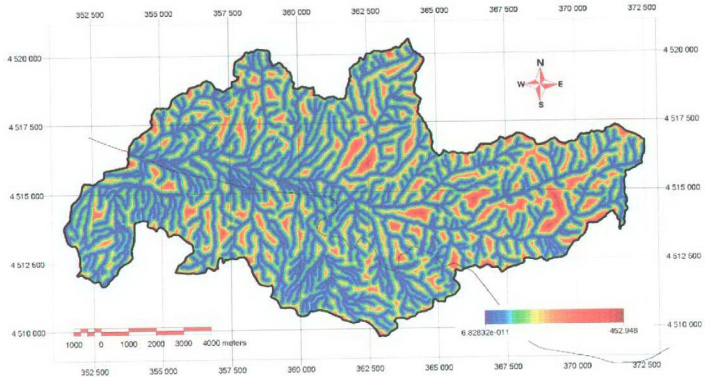
Figure 4.12. The distance raster of every pixel to the nearest drainage-line (DISTDRAINMAP).

The drainage lines are also used to calculate the kilometer square density of drainage lines in the whole study area. To maintain the 1 square kilometer search distance 564 meter search radius with 250 meter offset is used (Figure 4.13). The vector of drainage lines are converted to point data with a distance of 25 meters. The analysis show that the maximum count is achieved at 352, hence the maximum drainage length in the area is nearly 9 kilometers in 1 square kilometer area.
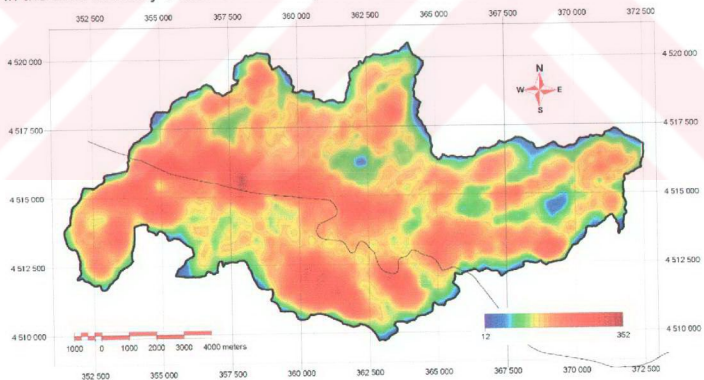
Figure 4.13. The drainage density of Asarsuyu catchment

The outputs of the watershed analysis are outputs for calculation of the nearest distances to the ridges of each pixel (Figure 4.14). The minimum value for the distance

65

to ridges is 1.62 meters and the maximum is 658.97 meters. The distribution has a mean of 138.72 meters with standard deviation of 99.86.
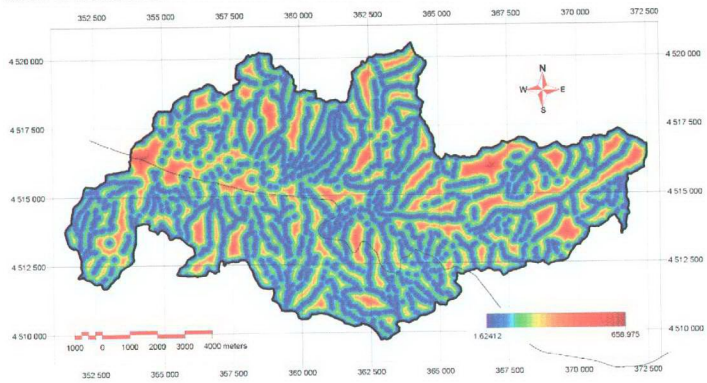
Figure 4.14. The distance raster of every pixel to the nearest ridge-line (DISTRIDGEMAP).

The topographical modeling of the Asarsuyu catchment also serves as an input to calculate the topographical derivatives such as aspect and slope maps of the catchment, which are very crucial and contain vital information for the landslide development in the area.

The aspect is a measure of slope orientation and is calculated in compass degrees as the azimuth from the North. Due to the raster data format itself and the computational limits the aspect distribution has sensitivities in principal directions and in every 22.5 degrees, which is a drawback of the algorithm used. Although the model creates reliable elevation data and slope calculations, the small size of the unique elevation facets prefers to be oriented in 16 principal directions which are 22.5 degrees apart from each other. The produced and color-coded **ASPECT** map and its frequency distribution is presented in Figure 4.15. The **ASPECT** map has a range between −1 and 359, -1 representing the flat lying areas and 0 as the north, any other value is the azimuth measurement from North. The minimum value −1 and the maximum is 359 degrees. The distribution has a mean of 182.45 with standard deviation of 110.
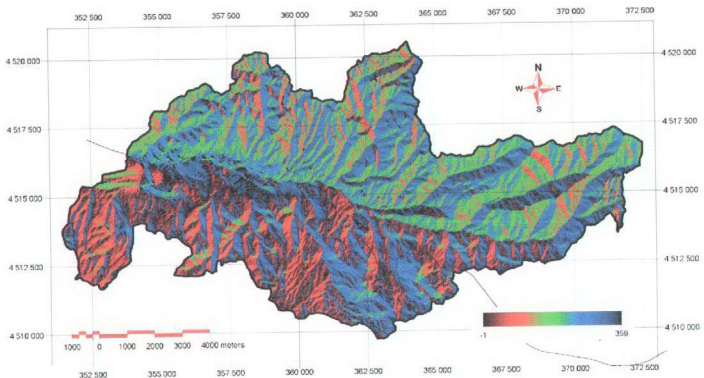
Figure 4.15. Aspect map of the Asarsuyu catchment (ASPECT).

The slope is the measure of surface steepness and is calculated in degrees. The produced and color-coded SLOPE map and its frequency distribution is presented in Figure 4.16. The SLOPE has a range between 0 and 90, 0 representing the flat lying areas and 90 as the vertical; any other value is the compass measurement from horizontal. Hence the minimum value of the data is 0 and the maximum is 56 degrees. The distribution has a mean of 16.97 degrees with standard deviation of 10.46.
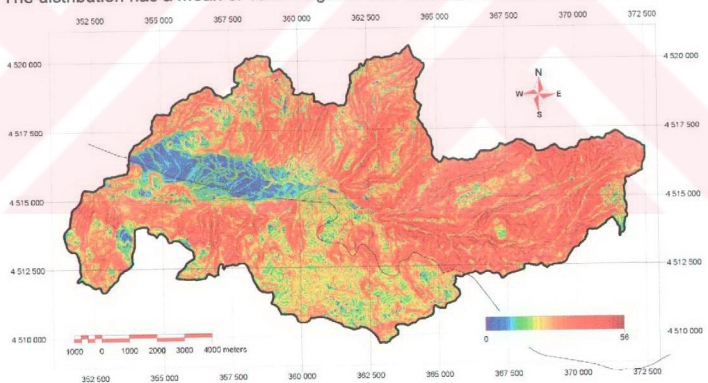
Figure 4.16. SLOPE map of Asarsuyu catchment.

67

4.3. Infrastructure

The presence of several infrastructure elements such as houses, power lines and road network might contribute to the evolution of landslides in the area. The infrastructure has a mutual relationship with the landslide hazard as either it causes the slide or it is affected from the slide. The causes can be explained as: near the houses there is a groundwater surcharge either by sanitary disposals or by small-scale backyard irrigation. For the construction of power line poles, the forest under and in the vicinity of the pylon is cut, so land cover changes. For the road building both the cut slopes, the land cover change and the economical activity near the roads, due to highway tourism, attract people. However, all of these built structures are then directly be faced with the hazard, as the elements at risk possessing vulnerability to landslide hazard.

4.3.1. Data entry

For the construction of historical infrastructure database, the necessary features are digitized from the 1:25.000 scale topographical maps of 1952, 1972 and 1994 maps (Figure 4.17). Although the map of 1972 is published in 1977, it is referred as 1972 maps, as it had been prepared from 1972 aerial photographs. No maps are produced from 1984 photographs as they were taken for forestry applications

4.3.2. Input Data Production

For the landslide database only the distance rasters of 1994 period is produced as it will yield the maximum development stage in the study area (Figure 4.18). The distance raster of settlement of 1994 has a minimum value of 0 ranging to a maximum of 6093.46 meters (Figure 4.18.a). The mean of this distribution is 1258.46 meters with a standard deviation of 1320.73. The area is divided in to two in a NW-SE trend, as all of the settlements are located south of this dividend. This could be attributed to the presence of continuously decreasing background noise in the frequency histogram of DISTSETTLEMENT. The power lines and the road network is merged as genetically they are affecting the topography in a similar way such as, land cover disturbance is needed to construct both and in order to construct the power lines a new temporary road is opened. The distance raster is then constructed from the merged vector coverage. The distance raster of merged coverage of 94 has a minimum value of 0 ranging to a maximum of 2312.68 meters (Figure 4.18.b), with a mean of 246,1 and a standard deviation of 302.099. The influence of E-5 highway is decided to be considered as a separate parameter, so a vector containing only E-5 highway is formed (Figure 4.18.c). The distance raster of E-5 highway of 94 has a minimum value of 0 ranging to a maximum of 8366.8 meters. The mean of this distribution is 2467.49 with a standard deviation of 1771.85.
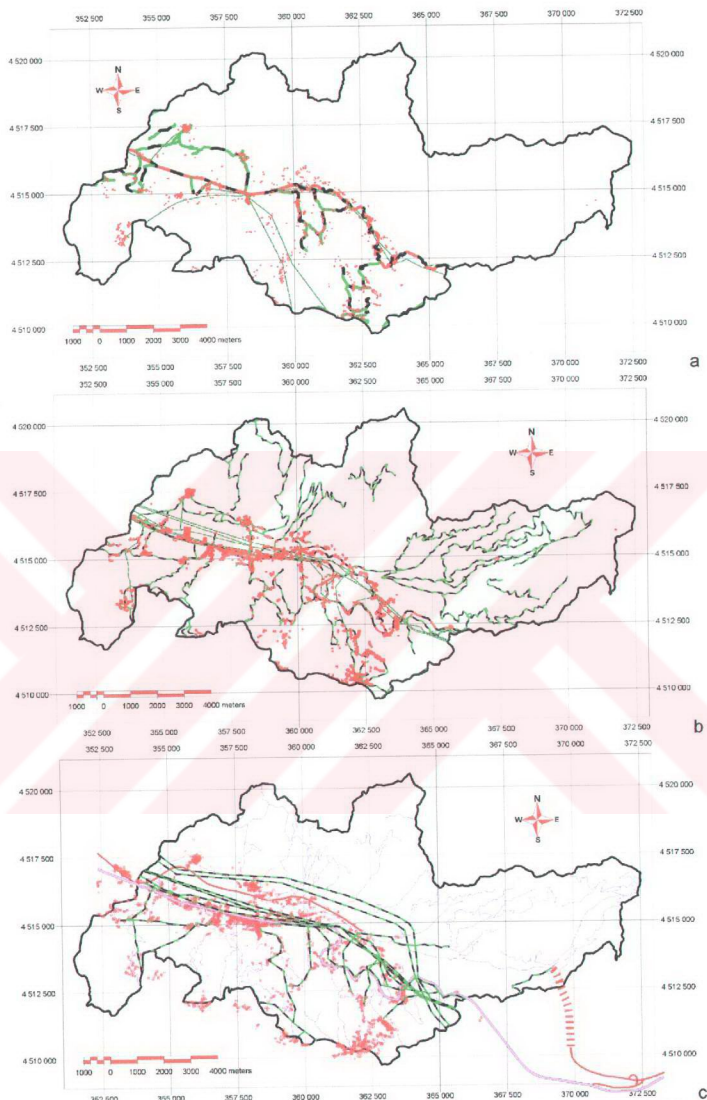
Figure 4.17. Historical infrastructure databases of Asarsuyu catchment a) 1952, b) 1972, c) 1994
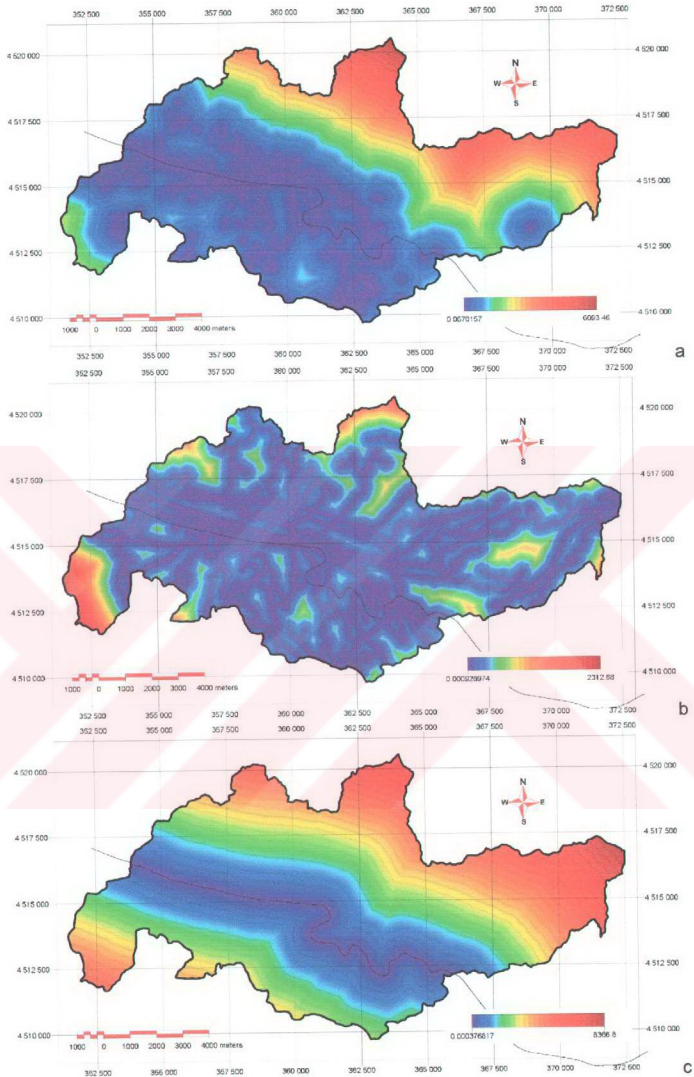
Figure 4.18. Distance rasters and frequency distributions of 1994 period a) distance raster of settlement of 1994, b) distance raster of road network and power lines of 1994, c) distance raster of E-5 road of 1994

4.4. Land cover

A landslide hazard assessment should not only depend on the production of landslide inventory map and its analysis. A complete hazard assessment system will require also the assessment of external factors leading to instability rather than the topography and material derived properties. The land cover distribution is one of the external factors that can easily be mapped and monitored in time if needed with the aid of satellite images (Soeters and van Westen, 1996). In landslide hazard assessment projects and in environmental and engineering studies, accurate and up to date information about land cover on a regional scale is often resembles vital information for help of decision rule generation. Basically a land cover map is obtained by classifying remotely sensed images. Typically this is performed by the spectral analysis of individual pixels and their association with other neighboring pixels. The results of classification depend largely on the type of area, land-cover type, and image acquisition date. However, the results of the classification are directly affected by spectral confusion of land-cover types and mixed pixels (Kam, 1995). The vegetation cover and soil moisture conditions produces distinctive spectral responses in the electromagnetic spectrum, that gives the opportunity to the classifier to classify them easily. However, landslides also produce subtle changes in the health of the vegetation, altering the natural state of the surface and underground drainage conditions, so the soil moisture. The size of the classified areas regarding the spatial resolution of the Landsat TM5 satellite is often too small to allow interpretation of individual slope instabilities. This restriction limits the use of spectral data, but incorporates in conjunction with other factor maps in which they together provide convergent information for slope instabilities.

4.4.1. Input Data

A multispectral Landsat TM5 image of Bolu, acquired on 1990, was used for this study (Figure 4.19).Other materials used to extract the land cover of Asarsuyu catchment include:

1. Aerial photographs, panchromatic black-and-white, for 1953 (1:35.000), 1972 (1:25.000), 1984 (1:15.000) and 1994 (1:25.000).
2. Topographic maps for 1994 (scale: 1:25.000) at which the infrastructure (buildings and roads) of 1994 are digitized from. The data produced are stored in a GIS database for further integration processes.
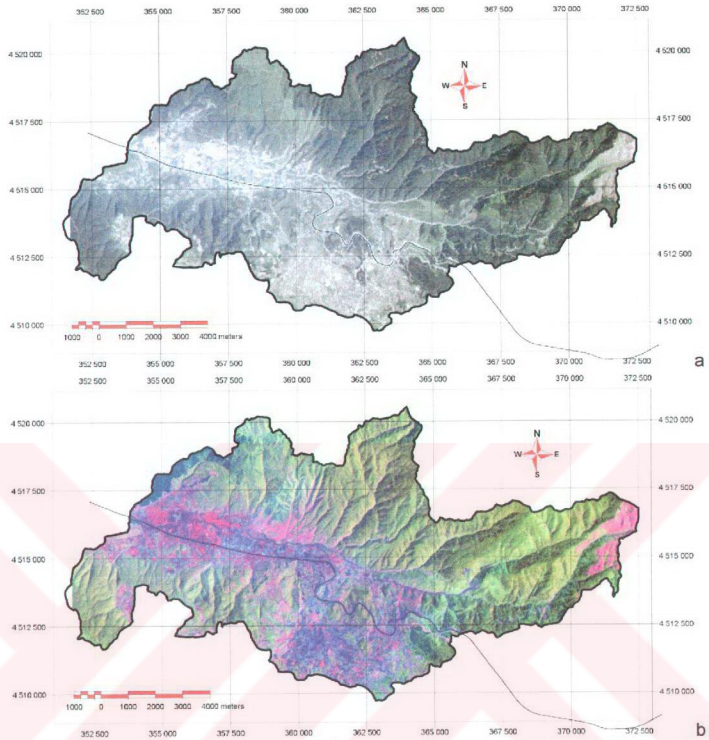
Figure 4.19. a) True color composite of Landsat TM 5 (R=3, G=2, B=1), b) false color composite of Landsat TM 5 (R=5, G=4, B=1).

4.4.2. Input Map Generation

Following several visits to the study area and detailed aerial photography interpretations, a land-cover classification scheme was developed. The classification scheme comprises six land-cover classes that display all the major land-covers encountered in this area, related to considering the landslide hazard assessment procedure. First class is the dense forest area. This is the oldest coverage detected in the study area, since it was observed on 1950's aerial photographs as a very dense forest. Second one is the young-age-forest area. In 1970's, a vast number of forest fires, removing the dense forest cover occurred in the study area, also with very intense human deforestation activities (forest industry). The reforestation studies were carried out after 1980's. Third class is the mixed group of barren areas, grasslands and

agricultural areas. Fourth one is soil moisture class, which should be on the bare land with significant amount of water in the surface. This could also be said of some form of wetland where the groundwater is very near to the surface. While the remaining two (road and settlement) come from GIS 1994 database.

The extraction of land-cover map from Landsat TM5 imagery and subsequent post classification process consisted of six major steps (Figure 4.20).
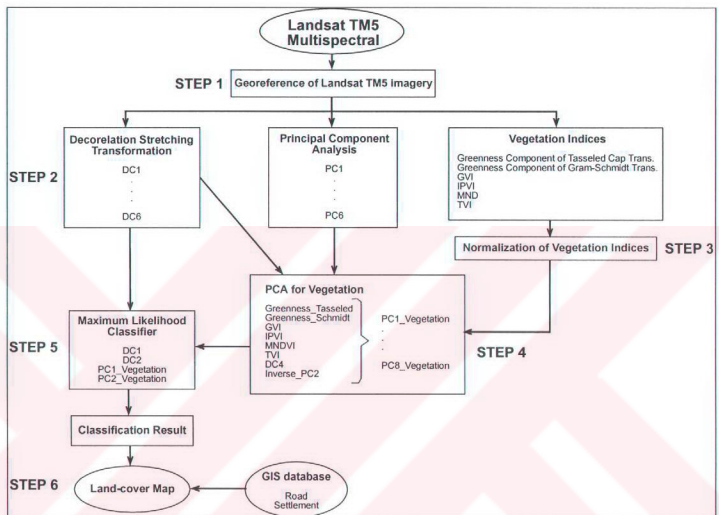
Figure 4.20. The methodological snapshot of land cover extraction scheme

4.4.2.1 Georeferencing

The procedure to rectify the Landsat digital data sets to the national coordinate system involved the following steps:

1. The determination of the ground control points (GCPs) from 1:25000 topographic maps of 1994 and from the digital image data (Table 4.1)

2. The computation of a least-squares solution for a third-order polynomial equation required to register the image data sets to the national coordinate system

3. The re-sampling of the data sets to a 30-m pixel resolution using nearest neighbor interpolation on the Landsat TM data.

73

The image was geometrically corrected based on the national grid reference. A total of twenty-six control points extracted from topographical map were employed in the georeference. The third degree polynomial equation with nearest neighbor interpolation was employed in the process.

Table 4.1. RMSE (Root Mean Square Errors) of Ground Control Points

Ground Control Points vs. RMSE (m)					
1	23.307	10	11.928	19	17.683
2	18.364	11	11.927	20	17.061
3	16.473	12	18.524	21	5.677
4	20.913	13	18.294	22	20.310
5	3.837	14	9.985	23	16.644
6	11.759	15	19.089	24	15.329
7	17.060	16	17.278	25	17.297
8	10.500	17	17.636	26	15.014
9	21.728	18	15.095	**MEAN RMSE**	**15,3**

4.4.2.2 Data Processing

4.4.2.2.1. Principal Component Analysis (PCA)

Extensive inter-band correlation is a problem frequently encountered in the analysis of multispectral image data. The Principal Components process uses the principal components statistical technique for reducing dimensionality of correlated multispectral data. The presence of correlations among the bands of a multispectral image implies that there is redundancy in the data, which means some information, is being repeated. It is the repetition of information between the bands that is reflected in their inter-correlations (Mather, 1999, pp.126-137). As all of the bands show different electromagnetic signature of the same feature and due to the orientation and range of the spectral band the information transferred is generally repetitive.

A correlation matrix showed that Band 1-Band 2, Band 1-Band 3 and Band 2-Band 3 were highly correlated (Table 4.2). This seemed to be a major deficiency of the Landsat TM5 imageries encountered elsewhere. In order to overcome this problem, a principal component analysis (PCA) was used to transform the highly correlated Landsat TM5 data into statistically independent orthogonal axes on which the original satellite data were reprojected.

The results of the PCA showed that first (52.7 %) and second (41.8 %) principal components accounted for 94.6 % of the variance within the entire Landsat TM5 data

set except the thermal 6th band. Components 3, 4, 5 and 6 respectively accounted for 4.2%, 0.75 %, 0.35 % and 0.12 % of the remaining variance, respectively. To extract vegetated areas and reduce the amount of computation, inverse of PC 2 (255 – PC 2) was used in the PCA for vegetation stage, which was obtained by looking at various combinations of color composite images. The resulting eigenvalues of the principal components are presented in Table 4.3.

4.4.2.2.2. Decorrelation Stretching

Decorrelation stretching is a color enhancement technique that is based on a principal component transformation of correlated multispectral image data which enhances the color display of highly correlated raster sets, such the first three Landsat TM bands (Table 4.2). Decorrelation stretching was performed by using reverse transformation of principal components. This enhancement exaggerates the differences in spectral properties between surface materials to a greater degree than is possible using conventional contrast enhancement of the original bands. Furthermore, to enhance the color in highly correlated images, there is a need to selectively exaggerate the least correlated portion of the spectral data, which is to decrease the correlation. Decreasing the correlation of spectral data corresponds to exaggerating the color saturation without changing the distribution of hues (or relative color composition).

The Decorrelation Stretching process involves three fundamental steps. First, a principal-component transformation is applied with the rows and columns of the eigenvector matrix transposed. Second, contrast equalization is applied by a Gaussian stretch, so that histograms of all principal components approximate a Gaussian distribution of a specified variance. Third, a coordinate transformation that is the inverse of the principal component rotation is applied so that the data are projected in their original spectral channels, using eigenvectors as weightings for each principal component. This inverse operation maximizes the spectral separability of different surface types in the restored spectral channels. The decorrelation stretched images created by this process can also be used as components for making color composites. Gillespie et al. (1987) explained in detail the Decorrelation Stretching technique and its theoretical and mathematical underpinnings.

In theory, band 1 is sensitive to water and soil moisture while band 2 and especially band 4 are sensitive to vegetation that is why related decorrelation components of these original bands represent specific features. As a result of the contrast stretching the Landsat 5 TM image of Asarsuyu catchment, subtle variations in surface materials are more easily discriminated by using decorrelated raster set. DC 1 is sensible to moisture in the area while DC 4 and DC 2 represent different non-common properties of vegetation through the whole original raster image set (Figure 4.21).

Table 4.2. Covariance, Correlation and Transformation Matrices for PCA

Variance / Covariance Matrix						
Raster	Band1	Band2	Band3	Band4	Band5	Band7
Band1	64.5906	42.3019	68.3945	-36.2414	69.6276	59.9829
Band2	42.3019	30.8834	48.569	-19.1387	53.3876	43.4149
Band3	68.3945	48.569	82.3578	-50.2956	82.4668	73.3751
Band4	-36.2414	-19.1387	-50.2956	306.3435	110.5237	-13.1521
Band5	69.6276	53.3876	82.4668	110.5237	253.4512	117.3291
Band7	59.9829	43.4149	73.3751	-13.1521	117.3291	82.456

Correlation Matrix						
Raster	Band1	Band2	Band3	Band4	Band5	Band7
Band1	1	0.9471	0.9377	-0.2576	0.5442	0.8219
Band2	0.9471	1	0.963	-0.1968	0.6034	0.8603
Band3	0.9377	0.963	1	-0.3166	0.5708	0.8904
Band4	-0.2576	-0.1968	-0.3166	1	0.3966	-0.0828
Band5	0.5442	0.6034	0.5708	0.3966	1	0.8116
Band7	0.8219	0.8603	0.8904	-0.0828	0.8116	1

Transformation Matrix						
Axis	Band1	Band2	Band3	Band4	Band5	Band7
PC 1	0.1028	0.0787	0.1174	0.1969	0.3442	0.16
PC 2	0.1497	0.0981	0.1835	-0.3974	0.0336	0.1376
PC 3	0.2276	0.1459	0.178	0.1763	-0.2548	-0.0174
PC 4	0.2783	0.0278	-0.1369	-0.0587	0.1519	-0.3465
PC 5	-0.243	0.1774	0.2511	-0.0119	0.0685	-0.2481
PC 6	-0.0707	0.4846	-0.3089	-0.0174	-0.0201	0.0983

Table 4.3. Eigenvalues and Associated Percentages

Axis	Eigenvalues	Percentages	Cumulative
PC 1	432,3325	52,7182	52,7182
PC 2	343,1456	41,8428	94,561
PC 3	34,5831	4,217	98,778
PC 4	6,1278	0,7472	99,5252
PC 5	2,8701	0,35	99,8752
PC 6	1,0233	0,1248	100
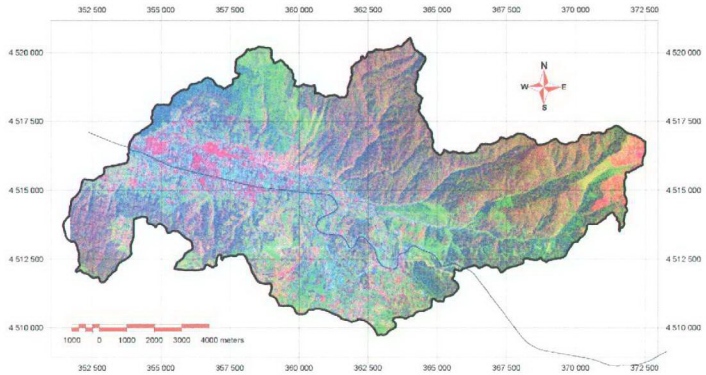
Figure 4.21. Decorrelation stretching results (R: decor_4, G: decor_3:B: decor_1).

4.4.2.2.3. Vegetation Indices

The reflectance spectrum of foliage shows a low reflectance (~0.05) in the visible part of the spectrum while solar irradiance is maximum. Light is absorbed by vegetation for photosynthesis. In the near infrared (NIR), foliage has a high reflectance (~0.5), with a very rapid transition between red and NIR regions at ~750 nm. This is completely different from the reflectance spectrum of the 'background material against which the leaves are usually observed. Soil reflectance gradually increases at higher wavelengths over the same region, though its absolute reflectance varies with soil-type and moisture content (wet soil being darker than dry soil). So the ratio or difference between two spectral bands on either side of 750 nm will give a measure of the quantity of foliage present. Those bands are usually chosen centered in the red part of the spectrum at 660 nm and in the near infrared at 870 nm. Under the light of this fact, remotely sensed spectral bands can tell user something useful about vegetation by calculating various indices. Under the light of this fact, remotely sensed multispectral bands can tell user something useful about vegetation by calculating various indices. A vegetation index is a number that is generated by some combination of remote sensing bands and may have some relationship to the amount of vegetation in a given image pixel. Therefore, six different vegetation indices were calculated to extract different types of vegetation cover located in the area. These are Greenness component of Tasseled Cap transformation, Greenness component of Gram Schmidt, GVI, IPVI, MNDVI and TVI.

4.4.2.2.3.1. The Tasseled Cap

This index is first defined by Kauth and Thomas (1976) in which it rotates the MSS data such that the majority of information is contained in two components or features that are directly related to physical scene characteristics (Lillesand and Kiefer, 1994, pp. 577-579). The index is then reapplied to TM data by Crist and Cicone (1984). The transformation for the six nonthermal Landsat TM bands (1-5 and 7) computes three index values: First one is greenness, which is strongly related with the amount of green vegetation, the second, brightness which is the weighted sum of the all input bands and defined in the principal variation in the soil reflectance and thirdly wetness is related to canopy and soil moisture. Most of the variability in soil and vegetation conditions contained in the six TM bands is expressed in these three dimensions. Each of these indices is computed cell by cell as a weighted sum (linear combination) of the input band values. The computation has the form:

Greenness =-0.24717 * TM1 - 0.16263 * TM2 - 0.40639* TM3 + 0.85468 * TM4 + 0.05493 * TM5 -0.11749 * TM7
Brightness =0.33183 * TM1 + 0.33121 * TM2 + 0.55177* TM3 + 0.42514 * TM4 + 0.48087 * TM5 + 0.25252 * TM7
Wetness =0.13929 * TM1 + 0.22490 * TM2 + 0.40359 * TM3 + 0.25178 * TM4 - 0.70133 * TM5 - 0.45732 * TM7

4.4.2.2.3.2. The Gram-Schmidt

The Gram-Schmidt transformation computes the Gram-Schmidt orthogonalization indices to delineate bare soil from vegetation in TM imagery. Input consists of a pair of raster objects; the spectral band of one object should be red and the other photo-infrared. Output also consists of a pair of raster objects: Greenness and Soil Brightness. The Gram-Schmidt transformation computes Bright Soil coefficients for both red and infrared rasters based on the parameter values entered for Dark Soil and Bright Soil (Figure 4.22). The operation computes Green Vegetation coefficients for the red and infrared rasters from the parameter values entered for Dark Soil and Green Vegetation (Table 4.4). These coefficients are used in the following formulas to produce two output rasters, one that displays the amount of green vegetation and another that displays the amount of soil brightness.

Greenness = 0.5 + (red coeff1 * red cell value) + (NIR coeff1 * NIR cell value)
Soil Brightness = 0.5 + (red coeff2 * red cell value) + (NIR coeff2 * NIR cell value)

Table 4.4. Parameters used in the Gram-Schmidt Transformation

	Cell value of Red Band	Cell value of NIR Band
Dark Soil	19	14
Bright Soil	43	22
Greenness	10	100
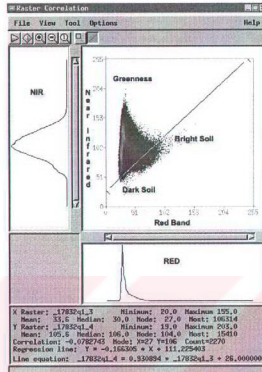
Figure 4.22. Near Infra Red versus Red band Raster Correlation Graph

4.4.2.2.3.2. Global Vegetation Index (GVI)

GVI stands for Green Vegetation Index (Terrill, 1994). There are several GVI's. The basic way these are devised by using two or more soil points to define a soil line. The distance of the pixel spectrum in band space from the soil line along the "greenness" axis is the value of vegetation index.

GVI = -0.2848*TM1-0.2435*TM2-.5436*TM3+0.7243*TM4+0.0840*TM5-0.1800*TM7

4.4.2.2.3.3. Infrared Percentage Vegetation Index (IPVI)

IPVI is the Infrared Percentage Vegetation Index first defined by Crippen (1990). IPVI is functionally equivalent to NDVI and RVI, but it only ranges in value from 0 to1. It also eliminates one mathematical operation per image pixel which is important for the rapid processing of large amounts of data (Terrill, 1994).

IPVI=(NIR)/(NIR+red)= 0.5*(NDVI+1)

4.4.2.2.3.4. Modified Normalized Difference Index (MNDI)

The Modified Normalized Difference operation creates an output raster object known as the ND or green biomass raster object. The operation shows a measure of the difference between the values of two input bands RED and NIR. The normalized

difference is modified by the values of Path Radiance for Band A, Path Radiance for Band B, and Minimum Euclidean Distance fields. The expression for the operation is as follows:

*Modified ND = Scale * (NIR - b) * 254 / (NIR + RED - a - b) + 1*
(a and b are path radiances for RED and NIR, respectively).

4.4.2.2.3.5. Transformed Vegetation (TVI)

The TVI operation computes the Transformed Vegetation Index from a pair of input raster objects. The spectral band of one object should be RED and the other NIR. The expression for the operation is as follows (Lillesand and Kiefer, 1996):

*TVI = 100 * (sqrt[(NIR-Red) / (NIR+Red)] + 0.5)*

4.4.2.4. Normalization

The values of pixels in all vegetation indices fall within a range of −1 to +1. Therefore, they were all transformed into 0-255 range to be put into PCA. For normalization process, the following formula was applied:

Normalized DN = [(Input DN – Min DN) / (Max DN – Min DN)] * Scale Factor
Scale Factor = 255

4.4.2.5. PCA for Vegetation

The land cover class vegetation can be made up of several land use classes. It can be natural vegetation, forestry or agricultural vegetation. Within these classes vegetation can be separated based on species, biomass, diseases and other. To obtain different species distribution and health conditions, there is a need to obtain components, which represent different vegetation types and their properties. However, having 8 different factors resulted as most of the data as being redundant, conserving the species information and removing the redundant information PCA was applied to normalized vegetation indices, DC 4 and inverse of PC 2 (Table 4.5).

The results of the PCA for vegetation showed that the first principal component accounted for 75.6 % of the variance within the entire data set. Components 2, 3, 4, 5, 6, 7 and 8 respectively accounted for 20.6 %, 2.09 %, 0.73 %, 0.41 %, 0.28 %, 0.13 % and 0.06 % of the remaining variance (Table 4.6.). Thus 96.3 % of the total variance the eight components were explained by the first two principal components.

Table 4.5. Covariance, Correlation and Transformation Matrices of PCA for Vegetation

Variance / Covariance Matrix								
Raster	DC 4	Greenness_T	Greenness_G	GVI	IPVI	MODIFIED_ND	PC 2 INVERSE	TVI
DC 4	1622.3652	587.3567	650.7342	419.4895	616.597	638.939	403.7972	395.832
Greenness_T	587.3567	297.3846	309.1295	237.7942	481.9862	516.4237	231.7356	331.5238
Greenness_G	650.7342	309.1295	337.117	242.5585	456.0567	487.8237	232.8455	311.6548
GVI	419.4895	237.7942	242.5585	238.0185	510.2078	486.4798	225.6202	307.4057
IPVI	616.597	481.9862	456.0567	510.2078	1372.325	1326.2998	499.0305	836.5095
MODIFIED_ND	638.939	516.4237	487.8237	486.4798	1326.2998	1460.6291	491.7419	921.0824
PC 2 INVERSE	403.7972	231.7356	232.8455	225.6202	499.0305	491.7419	237.6426	311.0036
TVI	395.832	331.5238	311.6548	307.4057	836.5095	921.0824	311.0036	607.6737

Correlation Matrix								
Raster	DC 4	Greenness_T	Greenness_G	GVI	IPVI	MODIFIED_ND	PC 2 INVERSE	TVI
DC 4	1	0.8456	0.8799	0.6751	0.4132	0.4151	0.6503	0.3987
Greenness_T	0.8456	1	0.9763	0.8938	0.7545	0.7836	0.8717	0.7799
Greenness_G	0.8799	0.9763	1	0.8563	0.6705	0.6952	0.8227	0.6886
GVI	0.6751	0.8938	0.8563	1	0.8927	0.8251	0.9487	0.8083
IPVI	0.4132	0.7545	0.6705	0.8927	1	0.9368	0.8738	0.916
MODIFIED_ND	0.4151	0.7836	0.6952	0.8251	0.9368	1	0.8347	0.9777
PC 2 INVERSE	0.6503	0.8717	0.8227	0.9487	0.8738	0.8347	1	0.8184
TVI	0.3987	0.7799	0.6886	0.8083	0.916	0.9777	0.8184	1

Transformation Matrix								
Axis	DC 4	Greenness_T	Greenness_G	GVI	IPVI	MODIFIED_ND	PC 2 INVERSE	TVI
PC 1_Veg	0.1532	0.0889	0.089	0.0801	0.1888	0.1964	0.0791	0.1245
PC 2_Veg	0.3895	0.0673	0.0981	0.0101	-0.1571	-0.1652	0.0046	-0.108
PC 3_Veg	-0.0082	-0.0509	-0.0469	0.1409	0.2723	-0.2066	0.1077	-0.1664
PC 4_Veg	-0.1458	0.134	0.2029	0.1399	-0.111	-0.0752	0.153	0.0383
PC 5_Veg	-0.0493	0.1118	0.2109	-0.0354	0.1411	-0.0987	-0.3414	0.0114
PC 6_Veg	-0.0492	0.0071	0.1321	0.025	-0.0335	0.2873	-0.0233	-0.4425
PC 7_Veg	-0.0235	0.0658	0.1165	-0.407	0.1026	-0.0525	0.1955	-0.0367
PC 8_Veg	0.0307	-0.522	0.3237	0.0099	-0.0033	0.0008	0.0248	0.0849

Table 4.6. Eigenvalues and Associated Percentages

Axis	Eigenvalues	Percentages	Cumulative %
PC 1_Veg	4670,2763	75,6546	75,6546
PC 2_Veg	1274,0529	20,6386	96,2932
PC 3_Veg	129,2741	2,0941	98,3873
PC 4_Veg	45,2019	0,7322	99,1196
PC 5_Veg	25,0813	0,4063	99,5259
PC 6_Veg	17,5736	0,2847	99,8105
PC 7_Veg	7,7659	0,1258	99,9363
PC 8_Veg	3,9298	0,0637	100

4.4.3. Maximum Likelihood Classification

Supervised classification began with the selection of training areas for each of the land-cover classes. Spectral signatures were generated from the training-area. The signature files were employed in a maximum-likelihood classifier to lace all pixels in one of four land-cover classes. In the classification four different data were used. These are:

   i.   Two decorrelation components (DC 1 and DC 2)

   ii.   Two PCA-vegetation components (PC 1-vegetation and PC 2-vegetation)

These four data sets represent different unique spectral signatures of features in four-dimensional spectral space. DC 1 was selected for extracting soil moisture. DC 2 was selected to differentiate soil and vegetation types, which have spectrally same response. Because study area is dominantly covered by vegetation. PC 1-Vegetation and PC 2-Vegetation are also added, which are used to indicate different spectral responses of vegetation, into supervised classification (Figure 4.23).
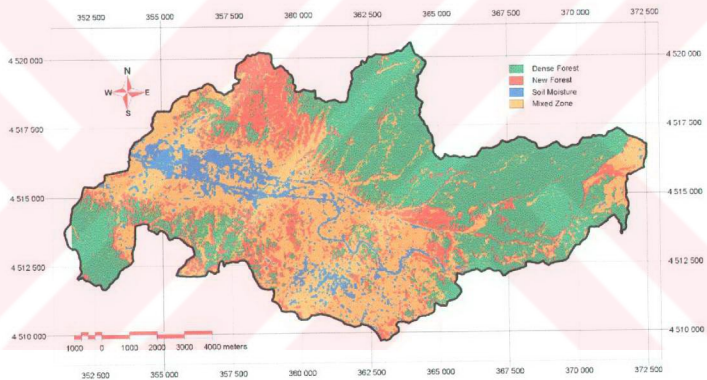
Figure 4.23. Product of Maximum Likelihood Classification

4.4.3.1. Accuracy Assessment

Classification accuracy was determined by comparing a sample of classified pixels with ground information derived from aerial photographs and field data (Congalton, 1991). To determine the accuracy of classification, approximately 29 reference areas (226 pixels) were selected as reference data for the comparison of ground information with the classification result. These pixels had to be pure rather than mixed pixels to ensure that the correct land cover was identified for each pixel. As with the training pixel, they were chosen with the aid of 1:15000, 1:25000 and 1:35000 scale

82

aerial photographs of the study area. For each pixel, the ground information determined from the aerial photographs (field checking when necessary) was compared with the classification results by means of confusion matrices. The total number of correct pixels in a category is divided by the total number of pixels of that category as derived from ground information (i.e., the column total). This accuracy measure indicates the probability of a reference pixel being correctly classified. This accuracy measure is often called "producer's accuracy", because the producer of the classification is interested in how well a certain area can be classified. On the other hand, if the total number of correct pixels in a category is divided by the total number of pixels that were classified in that category, then this result is a measure of commission error. This measure, called "user's accuracy" or reliability, is indicative of the probability that a pixel classified on the map or image actually represents that category on the ground (Lunetta et al., 1991).

User's accuracy of dense forest is 100 per cent, which is perfect. Although user's accuracy of both young-age forest and soil moisture looks low, there are several factors, which creates complexity to differentiate these classes. For example, even by looking at aerial photographs, it is also difficult to differentiate the young-age forest from old and dense forest at 1990's. For soil moisture 7 pixels out of 56 is classified as mixed class, which is not very abnormal since they are associated with each other (Table 4.7. In the mixed group 6 out of 83 pixels were misclassified as forest as the mixed group itself by definition also contains some scattered very small patches of dense trees. However, the overall accuracy (Kappa Coefficient) of the classification is 92.48 %, which indicates that the classification is within the desired confidence level.

Table 4.7 Error Matrix of the Classification

	Name	Forest	Y_Forest	Moisture	Mixed	Total	Accuracy
C	Forest	62	0	0	0	62	100.00%
l	Y_Forest	2	22	0	1	25	88.00%
a	Moisture	0	0	49	7	56	87.50%
s	Mixed	6	0	1	76	83	91.57%
s	Total	70	22	50	84	226	
i	Accuracy	88.57%	100.00%	98.00%	90.48%		
f	Overall Accuracy = 92.48% Khat Statistic = 89.45%						

(Table header spanning: "Ground Truth Data" over Forest, Y_Forest, Moisture, Mixed)

4.4.4 Integration of RS, GIS at Database Level

In the database level of integration, analysis are performed with the integrated vector features, database and raster products. The integration of remote sensing information into a GIS occurs naturally in a Raster format (or Raster GIS) because both data structures are approximately the same. Integration into a vector system requires

somewhat more effort, updating vector information by using an image as a backdrop for vector editing (Faust et al., 1991).

Road network and building layers were first converted into raster format. Result of maximum likelihood classification was used as a reference data in this conversion. In raster map of building, DN of 100 represents buildings while DN of 250 indicates other features located in the area. In raster map of road, DN of 200 represents road while DN of 250 indicates other features. DN of Land cover obtained from supervised classification are indicated as follows:

1 : Forest Area

2 : Young-age Forest Area

3 : Soil Moisture

4 : Mixed Group

Logical and Relational Raster Overlay Analysis:

```
if ( Building_Layer = = 100 )
{
Landcover = 5   ( DN of 5 shows the Buildings in the product of this process)
}
else if ( Road_Layer = = 200 )
{
Landcover = 6 ( DN of 6 shows the Road in the product of this process)
}
else
{
Landcover = Classification_Result
}
```

Finally, the product of this process includes the additional two classes (Buildings and Road) (Figure 4.24). This is also called integration results to GIS / Attribute Update (Archibald, 1987). As a result the majority of the area (44%) is covered by dense forest, %33 is mixed class with agriculture grassland and 16% is covered by young forest (Figure 4.24.b)

It is noticeable that the classification accuracy is significantly higher for vegetated areas than for other classes. For example, forest area and mixed group were most accurately classified at levels of 100 per cent and 91.57 percent, respectively.

Although soil moisture and young-aged forest areas classes showed more errors of commission, forest and mixed group represented more errors of omission. The errors of commission were largely caused by confusion between young-age forest and old - dense forest, soil moisture and mixed group, while the errors of omission resulted from confusion between old-dense forest and mixed group, mixed group and soil moisture Indeed, confusion between mixed group and soil moisture has the highest percentage in the total classification error. This is because within the mixed classes there may be green vegetation and/or agricultural fields, which reflect same spectral signature with soil moisture.
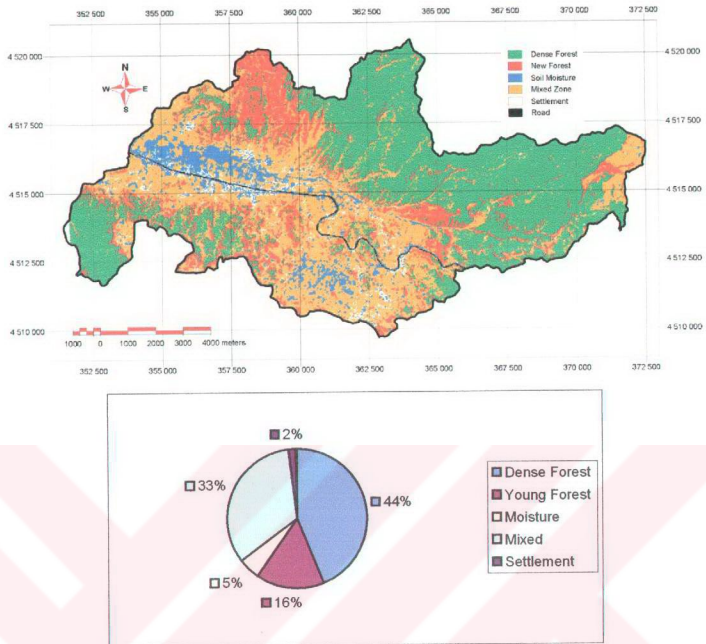
Figure 4.24.a) Land-cover map of the study area, b) areal distributions of land cover.

4.5. Landslide Inventory

In this section the creation of landslide attribute databases are explained. Four landslide attribute databases are created using the photo interpretation of the stereo panchromatic aerial photographs.

4.5.1. Input Data

The input data for landslide attribute databases and the inventories of four different time periods consist of 4 sets of analogue stereo panchromatic aerial photographs. The time periods and scales of the photographs are as follows: 1952 (1:35.000), 1972 (1:25.000), 1984 (1:15.000) and 1994 (1:25.000).

4.5.2. Data Production

The landslide occurrences of the Asarsuyu catchment are mapped in four different time periods. The geomorphological and morphometrical attributes of the landslides are mapped and they are transferred into each years corresponding topographical map in order to minimize the georeference residual errors in a considerable amount. Furthermore, a relational database is constructed and attached into the landslide map (Figure 4.25). The attribute database of the each landslide consists of 7 attributes of which they are named as *Massinfo, Type, Style, Depth, Landcover, and Distribution of activity and Possible Cause*.

"Massinfo" attribute is the morphology of the slope instability seen in the photograph; it has two terms that can be used *"Scarp&Path"* and *"Scarp&Body"*. *"Scarp&Path"* is used to define the landslides where scarp is clearly visible and the direction of movement is easily estimated although there are no signs of body. On the other hand, the latter *"Scarp&Body"* is assigned to the landslides where it's scarp and body is clearly seen.

"Type" attribute is the class of the landslide according to the Varnes (1978) classification. The available items that can be used in the database are confined only to *"slide", "flow"* and *"flow/slide (complex)"* due to the occurrence in the study area. The *"slide"* and *"flow"* is straightforward and self-explanatory, however the *"flow/slide (complex)"* is used for the slope instabilities that are started as a slide and then the displaced mass is flowed out remaining only a visible scarp.

"Style" attribute is defined by "the way in which different movements contribute to the landslide" (Varnes, 1978) and the definitions used in this study is based accordingly on the report of Unesco Working Party on World Landslide Inventory, 1993. The terms that can be used for the study area from the available terms are "successive", "single" and "multiple". The "single" slides consist of a single movement of displaced material; the "multiple" movements are landslides with repeated development of the same type of movement. The definition of "successive" is "A successive movement is identical in type to an earlier movement but in contrast to a multiple movement does not share displaced material or a rupture surface with it".

"Depth" attribute is assigned as the relative measure obtained from the interpretation of the aerial photographs, which is significantly dependent on the experience of the interpreter.

"Distribution of activity" attribute is self explanatory as is assesses the activity distribution of a landslide. The available terms are "advancing", "retrogressive", "diminishing" and "moving" and the definitions used in this study is based accordingly on the report of Unesco Working Party on World Landslide Inventory, 1993. The "retrogressive" is defined as the movement is continuing by the extension of the rupture surface in the direction opposite

to the movement of displaced material". If the rupture surface is extending in the direction of movement it is said to be "advancing". The term "Diminishing" is used for a landslide whose displacing material is decreasing in volume. Landslides whose displaced material continue to move but rupture surfaces show no visible changes can be simply described as "moving".

The "land cover" attribute is directly taken from the aerial photographical interpretation, dependent on the study area characteristics, and five classes have been defined. These are "bare land", "grass land", agricultural fields", "dense forest" and "young forest".

Figure 4.25. The landslide attribute database

The produced historical landslide inventory maps are presented in Figure 4.26. For every landslide inventory map a polygon topology is built and validated, in order to yield area (Polystats) and shape dependent (Fuzzystats) attribute databases. The photo interpretation database, Polystats and Fuzzystats databases are explored in detail for every period in the next chapter.

After compiling the four years inventory databases, all of them are merged together to form a single training resource for statistical analyses in order to figure out the parameters controlling the instability. If a slide occurs in any one of the periods it is included in the merged landslide map. Following the merge process a buffer of 100 meters is calculated around the landslide polygon and manually edited just to be buffered in the crown and flank areas of the polygons. The buffer is calculated in order to get the pre-failure surface attributes of the slides. The decision rule for the boundaries of the buffer line is as follows: If the distance between the slide boundary and microcatchment divide line is smaller than 100 meters then use the microcatchment divide line, if the distance is larger than 100 meters then use the 100 meter buffer line for the seed zone generation. After finishing the merge and buffering processes, a polygon grid of 25 meter pixel size is overlaid over the landslide polygons. The midpoints of this polygon grid are calculated and the attributes of landslides and the previously produced input maps in this chapter are transferred as separate attribute tables to these points (Figure 4.27). These separate attribute tables are then merged to construct a huge relational database concerning all of the parameters, which are going to be

used for later statistical model creation. Each parameter map will now be treated as a new variable in this database.
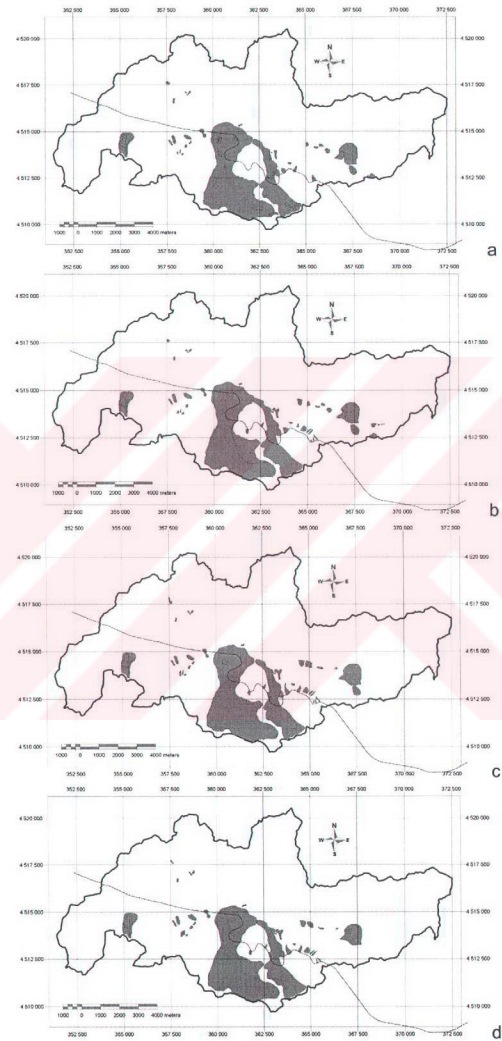
Figure 4.26. Landslide inventories of the four time periods, a. 1952, b. 1972, c. 1984, d. 1994
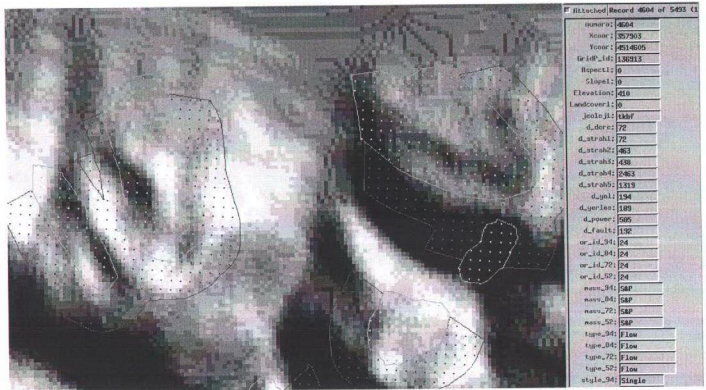
Figure 4.27. Polygon grid midpoints and transferred attributes (white points are merged polygons, black ones are buffered zone)

CHAPTER 5

LANDSLIDE DATABASES

This chapter will deal with the characteristic features and the descriptive statistics of the landslides mapped. The aim of this data exploration is to explore the training data, which will be used in the further analyses and to characterize the landslides and their evolutions in all of the four periods starting from 1952 to 1994.

5.1. Topological, Morphometrical database

In this section a brief description of the morphometrical characteristics of the landslides will be given. In order to do this, the vector topology dependent statistical attribute table/database (Polystats) and the fuzzy properties attribute table/database (Fuzzystats) will be used. The topological database is consisted of 11 variables originally, however 4 of them are valid if the polygon is hosting an island polygon, 2 of them are related with the spatial position of the polygon, which will be stressed in the following chapters, and 1 of them is a derived variable from Fuzzystats database, therefore, only 4 variables from the Polystats database will be used in order to trim the redundancy. The Polystats variables and their definitions are presented in Table 5.1.

Table 5.1. The names and definitions of variables used from Polystats database.

NAME	DEFINITION
Area	The area of a polygon.
BoundLen	The total length of a polygon's boundary.
MaxDim	The length of the longest diagonal from and to all polygon angles.
Roughness	Roughness is a measure of the irregularity of a polygon's perimeter. It is calculated as: *(MaxDim * BoundLen) / Area*.

In the Fuzzystats table, 12 parameters are available; the names definitions and formulas are given in Table 5.2. These fuzzy properties are reflecting the measures of shape, that in the scope of investigation. A shape is a difficult property to measure or define precisely and mathematically. It is difficult, if not impossible, to construct a

measure which is unique to a single shape. There have been attempts to characterize a variety of shapes including simple forms, like sand grain shapes, to very complex forms as are indicative of fossilized organisms. From a larger or more regional perspective the evaluation of shapes can be applied to drainage basins, coral atolls, salt diapirs, oil fields, or structural traps. Based on these, if the similarly shaped objects are present in a vector object, characteristic shape measure values associated with a known shape should be defined. All of the available variables are given in Table 5.2, but "thinness ratio", "Normalized Dispersion", "Simplicity", Shore line Development" will not be implemented for analyses as they are directly dependent of the digitizing process and even some of them undefined for the landslide polygon vector object, such as there are no sliver polygons, or the distances in between boundary vertices are equal. Also the "correlation" and "orientation" variables are not implemented as "orientation" yields a bi-directional value which is not applicable to landslides, as it is not concerning the orientation of the landslide but the orientation of the polygon without any thematic and genetic information is concerned. This means that the value is only dependent on the orientation of a physical long axis, which may not be in the same direction with the landslide or the downslope gradient. Moreover, the "correlation" is a dummy variable that is used for the calculation of "orientation" variable, yet it is also skipped. Only 6 of the Fuzzystats variables out of 12 are explored to find and evaluate shape dependent differences or similarities in the area from 1952 to 1994.

5.1.1. Polystats Database

5.1.1.1. 1952 Period

33 landslides have been mapped from 1:35.000 scale panchromatic stereo aerial photographs. The minimum area recorded is 3175,15 square meter and the maximum is 817501,36 square meters, having a mean of 79687,86 and the standard deviation of the distribution is 153440,75 (Table 5.3). As for the sake of reader's agility, the general definitions and rules of thumbs of skewness and kurtosis parameters will be introduced briefly below.

Skewness is "A measure of the asymmetry of a distribution". The normal distribution is symmetric, and has a skewness value of zero. A distribution with a significant positive skewness has a long right tail. A distribution with a significant negative skewness has a long left tail. As a rough guide, skewness values more than twice its standard error is taken to indicate a departure from symmetry.

Kurtosis is "A measure of the extent to which observations cluster around a central point". For a normal distribution, the value of the kurtosis statistic is 0. Positive

kurtosis indicates that the observations cluster more and have longer tails than those in the normal distribution and negative kurtosis indicates the observations cluster less and have shorter tails.

Table 5.2. The names, definitions and formulas of fuzzy properties

Name	Definition	Formula
Form Ratio	Measures shape with a maximum value approximating one for squares	$FR = Area / (Long\ Axis^2)$
Grain Shape Index	Measures shape with a minimum value approximating two for long skinny polygons, pi for circles, four for squares, and may be very large value for fractals	$GSI = Perimeter / Long\ Axis$
Compactness	Measures shape with a maximum value approximating one for circles.	$C = 2 * sqrt\ (pi * Area) / Perimeter$
Thinness Ratio	Measures the shape of sliver polygons.	$TR = 4 * pi * Area / (Perimeter^2)$
Circularity 1	Measures shape, reflecting the element's similarity to a circle, with a maximum value approximating one for circles.	$C1 = sqrt\ (Area / (pi * MaxRadius^2)),$
Circularity 2	Measures shape, reflecting the element's similarity to a circle, with a maximum value approximating 1.0 for circles.	$C2 = sqrt\ (MinRadius / MaxRadius)$
Normalized Dispersion	Measures shape presenting a value approximating one for circles.	$ND = pi * WeightedRadiusSum / Area)$
Simplicity	Measures a shape's simplicity, used. in separation of polygons with the same shape but with a different number of boundary vertices	$S = MeanSegmentLength^2 / Area$
Shore Line Development	Measures polygon generalization, which is used in cartography for studying the relationship between measured distance and map scale.	$SLD = Perimeter / (2 * sqrt\ (pi * Area))$
Correlation	Measures the complexity and integrity of the polygon boundary shape.	Cor = the correlation coefficient between the X and Y coordinates of the boundary
Orientation	Measures a polygon's orientation.	$O = arccos\ (Cor)$
Elongation	Measures polygon proportions.	$E = ShortAxis / LongAxis,$

Based on the definitions, above the area variable of 1952 has a long right tail, and it clusters around smaller values with longer tails than a normal distribution (Table 5.3 and Figure 5.1.a). The distribution type is not crucial here and will not be implemented as for the original landslide polygons the population size is quite small such as N=33 in this case. However, just to test the validity of skewness and kurtosis rules, a Q-Q plot of normality test is applied and seen that the distribution is far away from a normal distribution (Figure 5.1.b).

Table 5.3 The descriptive statistics of Polystats 1952.

Descriptive Statistics of Variables of Polystats Database (1952)

		AREA	BOUNDLEN	MAXDIM	Roughness
N	Valid	33	33	33	33
	Missing	0	0	0	0
Mean		79687,8645907315	1081,3042428379	398,8217684336	7,8582125715
Std. Error of Mean		26710,6068250663	183,9006305842	53,4735059243	,7545948866
Median		29514,2572021500	798,1903720200	293,1430513500	6,8409510700
Mode		3175,15747070[a]	220,57400666[a]	84,03677303[a]	4,86070260[a]
Std. Deviation		153440,7542336397	1056,4286931289	307,1819047120	4,3348175988
Variance		23544065059,78822000	1116041,5836660140	94360,7225824886	18,7906436150
Skewness		4,044	2,924	1,875	4,247
Std. Error of Skewness		,409	,409	,409	,409
Kurtosis		17,847	9,625	3,334	21,019
Std. Error of Kurtosis		,798	,798	,798	,798
Range		814326,21105957	5210,71663773	1247,00648479	24,81103996
Minimum		3175,15747070	220,57400666	84,03677303	4,86070260
Maximum		817501,36853027	5431,29064439	1331,04325782	29,67174256
Sum		2629699,53149414	35683,04001365	13161,11835831	259,32101486

a. Multiple modes exist. The smallest value is shown
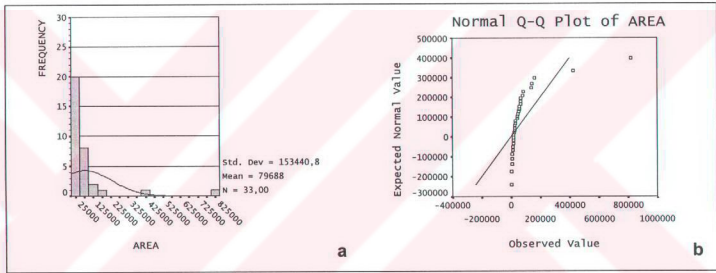
Figure 5.1. The distribution plots of area of 1952: a) Frequency distribution, b) Normal Q-Q Plot.

The boundary lengths are ranging from 220.57 meters to a maximum of 5431.29 meters. The mean is 1081.30 meters and the standard deviation is 1056.42. The skewness and kurtosis suggest that the distribution is not normally distributed and having long right tails yielding in an asymmetrical histogram with a clustering behavior (Figure 5.2.a). The MaxDim is ranging from 84.03 meters up to 1331.04 meters with a mean of 398.82 meters and a standard deviation of 307.18. It has also a long right tail with values clustering around mean values (Figure 5.2.b). The roughness has a minimum of 4,86 and a maximum of 29.67 with a mean of 7.85 and standard deviation of 4.33. It has also long right tail and an asymmetrical distribution (Figure 5.2.c).
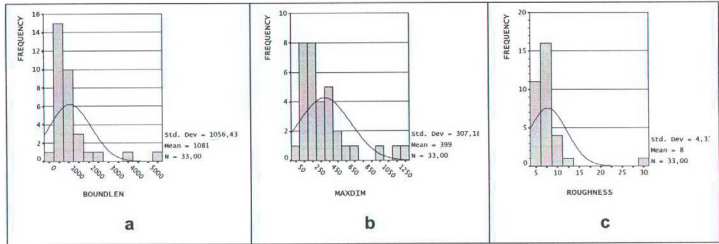
Figure 5.2. Histograms of Polystats database, a) "BoundLen", b) "MaxDim", c) "Roughness".

The correlations of the variables in the Polystats table are calculated and found that the "MaxDim" and "BoundLen" within themselves has the best correlation coefficient (0.956), also the "area" has good correlations with "BoundLen" and "MaxDim", 0.841 and 0.817, respectively. However, none of the variables are correlatible with "roughness", as the highest correlation coefficient of "roughness" is 0.423 with "MaxDim" (Table 5.4). Their scatterplots are presented in Figure 5.3.

Table 5.4. Correlations of Polystats variables 1952.

		AREA	BOUNDLEN	MAXDIM	ROUGHNESS
AREA	Pearson Correlation	1,000	,841**	,817**	-,012
	Sig. (2-tailed)		,000	,000	,947
	Sum of Squares and Cross-products	753410081913,223	4363532624,387	1231861404,435	-255540,974
	Covariance	23544065059,788	136360394,512	38495668,889	-7985,655
	N	33	33	33	33
BOUNDLEN	Pearson Correlation	,841**	1,000	,956**	,243
	Sig. (2-tailed)	,000		,000	,173
	Sum of Squares and Cross-products	4363532624,387	35713330,677	9931185,608	35649,475
	Covariance	136360394,512	1116041,584	310349,550	1114,046
	N	33	33	33	33
MAXDIM	Pearson Correlation	,817**	,956**	1,000	,423*
	Sig. (2-tailed)	,000	,000		,014
	Sum of Squares and Cross-products	1231861404,435	9931185,608	3019543,123	18023,079
	Covariance	38495668,889	310349,550	94360,723	563,221
	N	33	33	33	33
ROUGHNESS	Pearson Correlation	-,012	,243	,423*	1,000
	Sig. (2-tailed)	,947	,173	,014	
	Sum of Squares and Cross-products	-255540,974	35649,475	18023,079	601,301
	Covariance	-7985,655	1114,046	563,221	18,791
	N	33	33	33	33

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 5.3. Scatter plots of highly correlated variable pairs of Polystats database, a) "area-BoundLen", b) "area-MaxDim", c) "BoundLen MaxDim".

The correlations are quite logical as the "area" is closely related with the perimeter ("BoundLen") of a polygon, so with the maximum diagonal ("MaxDim"). As the "area" increases the two other parameters should have to be increased also. On the other hand, "roughness" is a derived component from the above three variables, but it has very low correlation, which implies that the irregularity of the polygons are not affected with how long or how large the polygon is.

5.1.1.2. 1972 Period

45 landslides have been mapped from 1:25:000 scale panchromatic stereo aerial photographs. The previous period's vector data set was used as a precursor for the interpretation, every landslide in the 1952 database checked in the 1972 period. A significant land-use change is observed in the study area; quite large areas are deforested either by human abuse of the forests as an economical source or by extensive forest fires.

The "area" variable has a minimum value of 2027.52 square meters, and the maximum is 786179.14 square meters (Table 5.5.). The distribution has a mean of 64906.05 and a standard deviation of 129606.99. Similar to 1952 period the "area" variable is asymmetrical and has long right tail. Although only 5 out of 45 case create this long right tail, similar results with less positive skewed distribution is obtained when the last 5 is omitted in the histograms (Figure 5.4).

The "BoundLen" variable is ranging from 176.39 to 5502.14 meters with a mean of 964.23 and a standard deviation of 950.44 (Table 5.5). The kurtosis and skewness values suggest a long right tail with clustering around smaller values. The "MaxDim" variable has a minimum of 65.51 meters and the maximum of this variable is 1328.03 meters. The variables distribution suggests a less positively skewed and less clustered

distribution rather than "area" and "BoundLen" variables with a mean of 357.15 and a standard deviation of 277.49 (Figure 5.4). For the "roughness" variable the minimum maximum range is defined as 4.9 and 30.68 with a mean of 8.11 and standard deviation of 4.24. The distribution has again a right long tail and clustering around small values.

Table 5.5. The descriptive statistics of Polystats 1972.

Descriptive Statistics of Variables of Polystats Database (1972)

		AREA	BOUNDLEN	MAXDIM	ROUGHNESS
N	Valid	45	45	45	45
	Missing	0	0	0	0
Mean		64906.0579833980	964.2372695904	357.1536664922	8.1114281029
Std. Error of Mean		19320.6706300324	141.6841362389	41.3659686927	.6332609391
Median		21517.4001464800	684.0998115000	264.8689724500	6.6746764100
Mode		2027.52832031[a]	176.39683459[a]	65.51217920[a]	4.90223315[a]
Std. Deviation		129606.9986989081	950.4460798904	277.4913538558	4.2480435223
Variance		16797974111.73878000	903347.7507790310	77001.4514647170	18.0458737673
Skewness		4.520	3.268	2.102	3.792
Std. Error of Skewness		.354	.354	.354	.354
Kurtosis		23.029	12.706	4.614	18.215
Std. Error of Kurtosis		.695	.695	.695	.695
Range		784151.61450196	5325.74997547	1262.52586739	25.77959581
Minimum		2027.52832031	176.39683459	65.51217920	4.90223315
Maximum		786179.14282227	5502.14681006	1328.03804659	30.68182896
Sum		2920772.60925291	43390.67713157	16071.91499215	365.01426463

a. Multiple modes exist. The smallest value is shown

Figure 5.4. The frequency distributions of variables of 1972 Polystats database, a) area, b) BoundLen, c) MaxDim, d) roughness.

The correlations are also similar with the period of 1952 as the maximum correlation is in between "MaxDim" and "BoundLen" (0.957) and the other two best correlations are in between pairs "area"-"BoundLen" (0.837) and "area"-"MaxDim" (0.815). Likewise, the roughness has the minimum correlation with the other variables (Table 5.6 and Figure 5.5).

Table 5.6. Correlations of Polystats variables 1972, the best correlations are shown in bold.

		AREA	BOUNDLEN	MAXDIM	ROUGHNESS
AREA	Pearson Correlation	1.000	.837**	.815**	-.046
	Sig. (2-tailed)		.000	.000	.763
	Sum of Squares and Cross-products	739110860916.506	4536214108.6	1290093036.0	-1117680.682
	Covariance	16797974111.739	103095775.19	29320296.273	-25401.834
	N	45	45	45	45
BOUNDLEN	Pearson Correlation	.837**	1.000	.957**	.188
	Sig. (2-tailed)	.000		.000	.215
	Sum of Squares and Cross-products	4536214108.564	39747301.034	11102333.528	33464.256
	Covariance	103095775.195	903347.751	252325.762	760.551
	N	45	45	45	45
MAXDIM	Pearson Correlation	.815**	.957**	1.000	.349*
	Sig. (2-tailed)	.000	.000		.019
	Sum of Squares and Cross-products	1290093036.019	11102333.528	3388063.864	18084.486
	Covariance	29320296.273	252325.762	77001.451	411.011
	N	45	45	45	45
ROUGHNESS	Pearson Correlation	-.046	.188	.349*	1.000
	Sig. (2-tailed)	.763	.215	.019	
	Sum of Squares and Cross-products	-1117680.682	33464.256	18084.486	794.018
	Covariance	-25401.834	760.551	411.011	18.046
	N	45	45	45	45

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Figure 5.5. Scatter plots of highly correlated variable pairs of Polystats database (1972), a) "area"-"BoundLen", b) "area"-"MaxDim", c) "BoundLen"-"MaxDim".

5.1.1.3. 1984 Period

Only 39 landslide bodies are mapped from 1:15.000 scale panchromatic stereo aerial photographs of 1984 period. Although the scale of the photograph is much more adequate for landslide monitoring, the number of landslides mapped is decreased probably due to intensive re-forestation studies in the study area.

The "area" variable ranges in between 4815.04 and 956223.4 square meters (Table 5.7). The distribution is again having a long right tail and a mean 74395.84 and a standard deviation of 161246.42 (Figure 5.6). The "BoundLen" variable has a minimum of 278.82 and a maximum of 5412,4 meters. It is again positively skewed and the kurtosis value indicates a clustering in smaller values. The mean and the standard deviation is 1047.86 and 984.93, respectively. The "MaxDim" of landslides are ranging between 113.96 and 1307.1 meters with a long right tail and clustering in the smaller values. The mean is found to be 393.5 meters and the standard deviation is 285.31.

Table 5.7. The descriptive statistics of Polystats 1984

Descriptive Statistics of Variables of Polystats Database (1984)

		AREA	BOUNDLEN	MAXDIM	ROUGHNESS
N	Valid	39	39	39	39
	Missing	0	0	0	0
Mean		74395.8434682997	1047.8647222372	393.5067514408	8.8682533931
Std. Error of Mean		25820.0918495585	157.7151385261	45.6874778206	.6695358135
Median		24466.1716308600	739.9429591700	327.1816415600	7.8408232500
Mode		4815.04296875[a]	278.82556722[a]	113.96302751[a]	5.14394527[a]
Std. Deviation		161246.4219189557	984.9307244129	285.3182075415	4.1812498149
Variance		26000408581.6658900000	970088.5318925580	81406.4795546676	17.4828500148
Skewness		4.755	3.096	2.049	3.245
Std. Error of Skewness		.378	.378	.378	.378
Kurtosis		24.886	11.056	4.288	14.631
Std. Error of Kurtosis		.741	.741	.741	.741
Range		951408.35791016	5133.58286545	1193.14480709	24.13985703
Minimum		4815.04296875	278.82556722	113.96302751	5.14394527
Maximum		956223.40087891	5412.40843267	1307.10783460	29.28380230
Sum		2901437.89526369	40866.72416725	15346.76330619	345.86188233

a. Multiple modes exist. The smallest value is shown

The correlations of the variables of 1984 database are similar with the period of 1952 and 1972. However, except "BoundLen-MaxDim" pair they exhibit slightly lower correlation coefficients, as the maximum correlation is again in between "MaxDim" and "BoundLen" (0.955) and the other two best correlations are in between pairs "area"-"BoundLen" (0.799) and "area"-"MaxDim" (0.792). On the other hand, the "roughness" variable is showing the minimum correlation with the other variables, and in this period the correlation coefficients are even lower than that of the previous two periods (Table 5.8 and Figure 5.7).
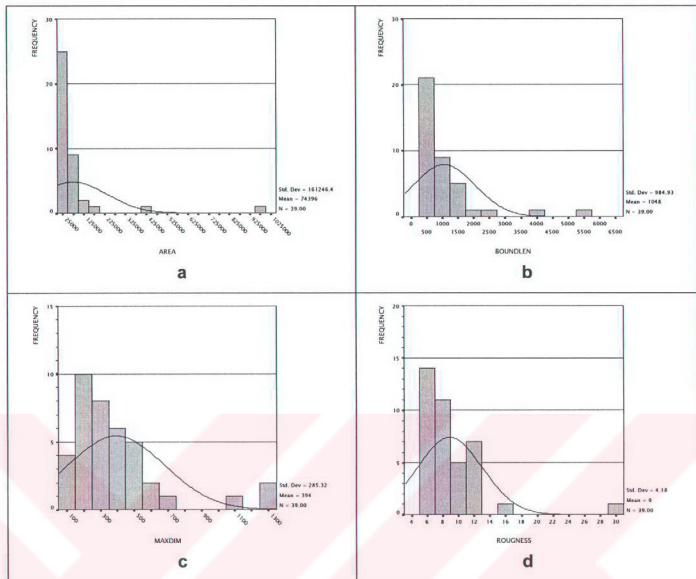
Figure 5.6. The frequency distributions of variables of 1972 Polystats database, a) area, b) BoundLen, c) MaxDim, d) roughness.

Table 5.8. Correlations of Polystats variables 1984, the best correlations are shown in bold.

		AREA	BOUNDLEN	MAXDIM	ROUGHNESS
AREA	Pearson Correlation	1.000	.799**	.792**	-.129
	Sig. (2-tailed)		.000	.000	.435
	Sum of Squares and Cross-products	988015526103.304	4822526244.705	1384862450.239	-3299801.888
	Covariance	26000408581.666	126908585.387	36443748.691	-86836.892
	N	39	39	39	39
BOUNDLEN	Pearson Correlation	.799**	1.000	.955**	.128
	Sig. (2-tailed)	.000		.000	.438
	Sum of Squares and Cross-products	4822526244.705	36863364.212	10201484.903	20001.389
	Covariance	126908585.387	970088.532	268460.129	526.352
	N	39	39	39	39
MAXDIM	Pearson Correlation	.792**	.955**	1.000	.300
	Sig. (2-tailed)	.000	.000		.064
	Sum of Squares and Cross-products	1384862450.239	10201484.903	3093446.223	13585.801
	Covariance	36443748.691	268460.129	81406.480	357.521
	N	39	39	39	39
ROUGHNESS	Pearson Correlation	-.129	.128	.300	1.000
	Sig. (2-tailed)	.435	.438	.064	
	Sum of Squares and Cross-products	-3299801.888	20001.389	13585.801	664.348
	Covariance	-86836.892	526.352	357.521	17.483
	N	39	39	39	39

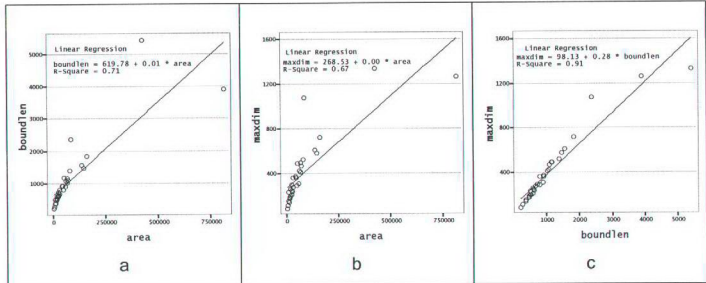** Correlation is significant at the 0.01 level (2-tailed).
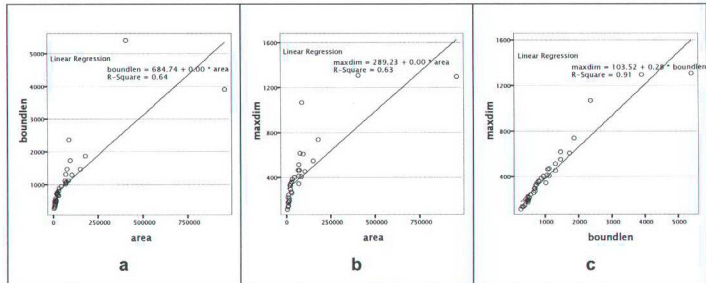
Figure 5.7. Scatter plots of highly correlated variable pairs of Polystats database 1984, a) "area"-"BoundLen", b) "area"-"MaxDim", c) "BoundLen"-"MaxDim".

5.1.1.4. 1994 Period

From the 1:25.000 scale panchromatic stereo aerial photographs of 1994 period 37 landslide bodies are mapped. The re-forestation studies seem to work well, however, the road-cuts seem to cause and/or trigger the landslides in the 1994 period, which is recorded in the aerial-photographs.

The range of the "area" variable is in between 5098.61 and 921243.32 square meters respectively. The data is again clustered in smaller values and have a long right tail with positive skewness and kurtosis values (Table 5.9. and Figure 5.8). The distribution has a mean of 73854.92 and 159508.47 as standard deviation. The "BoundLen" variable has a minimum of 289.25 and maximum of 5556.74 meters with 1043.56 as mean and 1012.49 as the standard deviation. The distribution is again right long tailed and clustered in smaller values. The "MaxDim" variable is found to have a minimum of 118.26 and 1325.77 meters as maximum value. It has again a long right tail and clustering with a mean of 390.85 and a standard deviation of 286.96. The "roughness" variable has a minimum of 4,95 and a maximum of 22.4 with a mean of 8.49 and a standard deviation of 3.24. The distribution is again long right tailed and clustered in small values.

The correlations of the variables of Polystats database of 1994 period is similar to that of 1952, 1972 and 1984 as the highest correlatible variables are "boundlen"-"maxdim", "area-MaxDim" and "area"-"BoundLen" pairs in decreasing correlation coefficient order. Conformably the "roughness" variable has the least correlation with the other remaining three variables (Table 5.10 and Figure 5.9).

100

Table 5.9. The descriptive statistics of Polystats 1994

Descriptive Statistics of Variables of Polystats Database (1994)

		AREA	BOUNDLEN	MAXDIM	ROUGHNESS
N	Valid	37	37	37	37
	Missing	0	0	0	0
Mean		73854,9253061662	1043,5660012157	390.8584145365	8,4932245689
Std. Error of Mean		26223,0310271408	166,4534916783	47.1766174767	.5342227747
Median		23967,4675293000	722,0782377300	327.5310183200	7,9804270400
Mode		5098,61462402[a]	289,25587197[a]	118,26133385[a]	4.95063299[a]
Std. Deviation		159508,4705627398	1012,4970622184	286.9641610937	3,2495502770
Variance		25442952181,2644300000	1025150,3010008170	82348.4297521847	10,5595770025
Skewness		4.640	3.222	2.113	2.480
Std. Error of Skewness		.388	.388	.388	.388
Kurtosis		23.500	11.823	4.520	8.704
Std. Error of Kurtosis		.759	.759	.759	.759
Range		916144,71289063	5267,48713772	1207,50785113	17,48556595
Minimum		5098,61462402	289,25587197	118,26133385	4.95063299
Maximum		921243,32751465	5556,74300969	1325,76918498	22,43619894
Sum		2732632,23632815	38611,94204498	14461,76133785	314,24930905

a. Multiple modes exist. The smallest value is shown
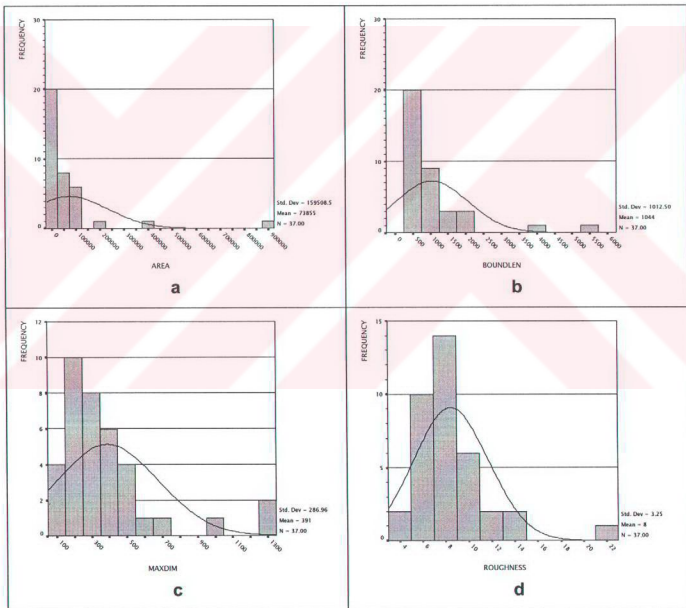
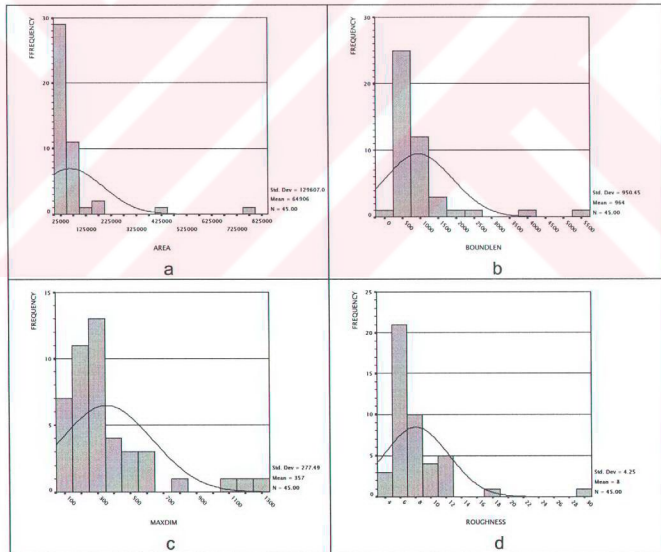**Figure 5.8.** The frequency distributions of variables of 1994 Polystats database, a) area, b) BoundLen, c) MaxDim, d) roughness.

Table 5.10. Correlations of Polystats variables 1994, the best correlations are shown in bold.

		AREA	BOUNDLEN	MAXDIM	ROUGHNESS
AREA	Pearson Correlation	1.000	.803**	.815**	-.113
	Sig. (2-tailed)		.000	.000	.504
	Sum of Squares and Cross-products	915946278525.519	4669481948.475	1343207587.405	-2115075.769
	Covariance	25442952181.264	129707831.902	37311321.872	-58752.105
	N	37	37	37	37
BOUNDLEN	Pearson Correlation	.803**	1.000	.955**	.120
	Sig. (2-tailed)	.000		.000	.481
	Sum of Squares and Cross-products	4669481948.475	36905410.836	9990958.352	14163.130
	Covariance	129707831.902	1025150.301	277526.621	393.420
	N	37	37	37	37
MAXDIM	Pearson Correlation	.815**	.955**	1.000	.269
	Sig. (2-tailed)	.000	.000		.107
	Sum of Squares and Cross-products	1343207587.405	9990958.352	2964543.471	9043.903
	Covariance	37311321.872	277526.621	82348.430	251.220
	N	37	37	37	37
ROUGHNESS	Pearson Correlation	-.113	.120	.269	1.000
	Sig. (2-tailed)	.504	.481	.107	
	Sum of Squares and Cross-products	-2115075.769	14163.130	9043.903	380.145
	Covariance	-58752.105	393.420	251.220	10.560
	N	37	37	37	37

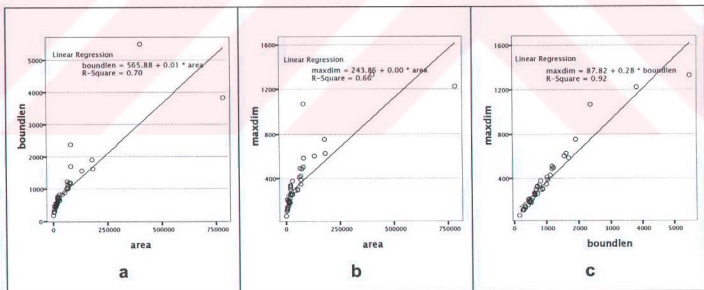** Correlation is significant at the 0.01 level (2-tailed).
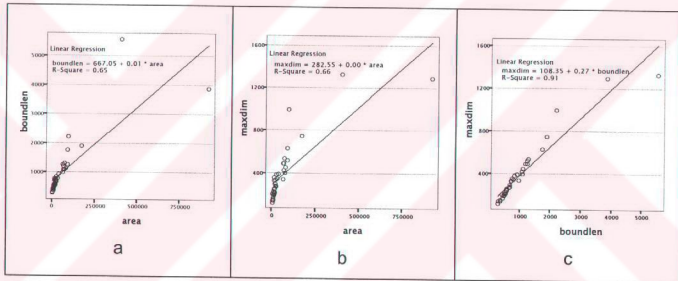
Figure 5.9. Scatter plots of highly correlated variable pairs of Polystats database (1994), a) "area-BoundLen", b) "area-MaxDim", c) "BoundLen MaxDim".

5.1.1.5. The Comparison of Four Periods

Three major events have occurred through period 1952 and 1994 in the area, which are revealed from aerial photographical interpretation. First one is the sudden de-forestation in the area in between the 1952-1972 period. This might either be due to an extensive forest fire or human abuse of the forest resources for economical raw material needs. The change in land cover should affect the states of landslides in the area and this is supported by the increase of number of landslides in the area from 33 to 45. On the other hand, not only the generation of new landslides but, it is also expected that present landslides could also re-activate, yielding to an increase in the "area", "MaxDim" and "BoundLen" variables. Although there are slight variations in the histograms (Figure

5.10), the variables are not showing a significant response to this assumption and the statistical analyses of the 3 variables could not say that these variables in these periods differ significantly in between 1952 and 1972 periods (Table 5.11). The Paired Samples t-Test of the all-available variables in the Polystats database fails to reject the null hypothesis of *"The two population means are equal"*. Also the intensive reforestation which is observed in 1984 period decreases the mappable landslide number from 45 to 39. However, the variables in the database are still ignorant to this. On contrary the present route of the E-5 highway is constructed at these years and there exist some slope instabilities related to this road, which might be on the charge to balance the effect of reforestation, yet no net change could be observed in the histograms. Furthermore, more intensive reforestation and road cut of E-5 in the 1994 period is recorded, stabilization in the forest areas and reactivation or generation in the road cuts had created a balance giving out a number of 37 slides in 1994 period. Furthermore, some statistical tests are used to consider whether the differences are significant or not, such as One-way ANOVA. The analyses show that the differences are not significant (Table 5.12). The landslides are open to external factors such as land cover change or geomorphological regime changes and they sometimes give strong responses. Here in this case, the landslides give a statistically insignificant response. The number of landslides change, the minimum-maximum, and the mean of the variables change slightly. A classification based on these attributes will be quite fictitious or it will force the nature to fit into a generated model.

Table 5.11. The results of Paired Samples t test.

Paired Samples Test

| | | Paired Differences | | | | | | | |
| | | | | | 95% Confidence Interval of the Difference | | | | |
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	A_52 - A_72	3133.2352943428	225138.3699281057	39799.2170203623	-78037.8029747261	84304.2735634117	.079	31	.938
Pair 2	A_72 - A_84	-2274.5032497102	226608.6488145063	37254.2323797402	-77829.5884501831	73280.5819507627	-.061	36	.952
Pair 3	A_84 - A_94	-22118.4051513676	176248.2498681428	28975.0337926642	-80882.4973644682	36645.6870617330	-.763	36	.450
Pair 4	B_52 - B_72	6.8854877403	1551.3268994336	274.2384426067	-552.4275034747	566.1984789553	.025	31	.980
Pair 5	B_72 - B_84	-11.4809617414	1476.2485762631	242.6937709486	-503.6867426775	480.7248191948	-.047	36	.963
Pair 6	B_84 - B_94	-69.1765357922	1375.9556543506	226.2057161523	-527.9429917081	389.5899201238	-.306	36	.762
Pair 7	M_52 - M_72	9.1444219503	444.0748629193	78.5020867312	-150.9616395081	169.2504834087	.116	31	.908
Pair 8	M_72 - M_84	-7.4818299246	414.7648201123	68.1869163963	-145.7713060134	130.8076461642	-.110	36	.913
Pair 9	M_84 - M_94	-21.4812258095	378.1292478052	62.1640654097	-147.5557939434	104.5933423245	-.346	36	.732
Pair 10	R_52 - R_72	-2.8773702812E-02	5.0865587919	.8991850537	-1.8626737106	1.8051263049	-.032	31	.975
Pair 11	R_72 - R_84	-.2011592211	6.0301049122	.9913431409	-2.2116962981	1.8093778559	-.203	36	.840
Pair 12	R_84 - R_94	.4110326595	5.5999202025	.9206212103	-1.4560736943	2.7781390132	.446	36	.658
Pair 13	A_52 - A_84	27069.3210870145	187088.1695223885	34741.4009083855	-44095.2126496878	98233.8548237167	.779	28	.442
Pair 14	A_52 - A_94	-8360.2636152007	261756.4594128563	49467.3211193707	-109858.8226564464	93138.2954260449	-.169	27	.867
Pair 15	A_72 - A_94	1592.8329171311	225210.0548926682	38067.4472200363	-75769.5276795204	78955.1935137826	.042	34	.967
Pair 16	B_52 - B_84	76.4396940662	1552.4303504508	288.2790789233	-514.0732300303	666.9526181627	.265	28	.793
Pair 17	B_52 - B_94	-78.3791612904	1749.3247757647	330.5913084969	-756.6964965475	599.9381739667	-.237	27	.814
Pair 18	B_72 - B_94	29.5674699980	1454.6627507336	245.8828825955	-470.1266680919	529.2616080879	.120	34	.905
Pair 19	M_52 - M_84	19.4612184045	434.1497285018	80.6195806536	-145.6805063749	184.6029431839	.241	28	.811
Pair 20	M_52 - M_94	-33.3368587636	494.8430475471	93.5165458442	-225.2169613225	158.5432437954	-.356	27	.724
Pair 21	M_72 - M_94	7.5023579734	412.3526156054	69.7003135055	-134.1457214458	149.1504373926	.108	34	.915
Pair 22	R_52 - B_84	-.8524458559	6.7143750055	1.2468281373	-3.4064575169	1.7015658052	-.684	28	.500
Pair 23	R_52 - R_94	-.7490438543	5.6189166838	1.0618754416	-2.9278322901	1.4297445816	-.705	27	.487
Pair 24	R_72 - R_94	.2064057986	5.9090675511	.9988147165	-1.8234299248	2.2362415219	.207	34	.838
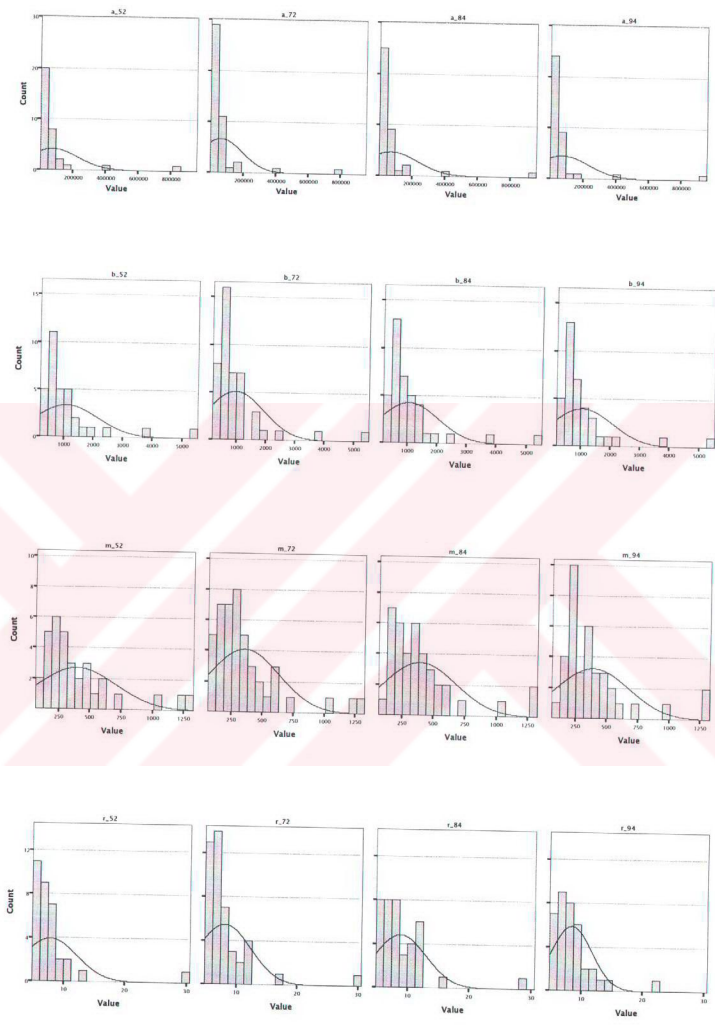
Figure 5.10. The distributions of Polystats variables in 4 periods. (A_52: "Area" in 52 period, b_72 "BoundLen" in 1972, m_84: "MaxDim" in 1984, r_94: "roughness" in 1994), each column represents single period and each row represents single parameter.

Table 5.12. The ANOVA table of Polystats variable.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
AREA	Between Groups	4505636342.854	3	1501878780.951	.066	.978
	Within Groups	3396482747458.552	150	22643218316.390		
	Total	3400988383801.406	153			
BOUNDLEN	Between Groups	300699.730	3	100233.243	.101	.959
	Within Groups	149229406.760	150	994862.712		
	Total	149530106.489	153			
MAXDIM	Between Groups	44889.490	3	14963.163	.180	.910
	Within Groups	12465596.681	150	83103.978		
	Total	12510486.171	153			
ROUGHNESS	Between Groups	21.763	3	7.254	.446	.720
	Within Groups	2439.812	150	16.265		
	Total	2461.575	153			

5.1.2. The Fuzzystats Database

As defined in section 5.1. of this chapter these parameters are used to see the changes in the shapes of the landslides. The previous database was directly concerned with the spatial distribution, orientation and topological (area related) properties of the landslide; this database is generated from the Polystats database regarding the topology of the vector object. In order to assess the shape, 6 variables are extracted from Fuzzystats database. In this section a variable based data exploration will be carried out rather than period basis like in the previous section. Before starting to explore the variables separately, the descriptive statistics (Table 5.13) and the results of one-sample Kolmogorov-Smirnov Normality test for the cumulative 4 periods (Table 5.14) is given. These state that the normality theory and normality assumptions are valid for the parameters which are going to be used.

5.1.2.1. Form Ratio

The form ratio theoretically approximates to 1 when the shape is a perfect square, to Π/4 for perfect circles and to lower values near 0 for long skinny polygons. The distribution of "Form Ratio" suggests that the landslides generally have a long but narrow outline as expected; since majority of them have flow component in the slide style. The minimum and the maximum of the variable is 0,082 and 0,74; very few of them resemble a circle as the value approximates to 0.7. The distribution mean is 0,40 and standard deviation is 0,14 (Table 5.13) yielding in a conclusion that landslides are not circles and are not skinny long polygons but their shape is in between them, namely they are ellipsoidal in shape. The positive near 0 skewness value and negative 0 near kurtosis value also suggest no departure from normal distribution with shorter tails but a slightly longer right tail and clustering around a fixed value (mean) (Figure 5.11). The past periods distributions are similar to each other, and their mean are ranging between

0,38 to 0,42 (Figure 5.12) yielding that the land cover change in the time period and its effect is not adequate to change the shape of the landslides, if a reactivation is present or new generation of landslides are seen the shape is still ellipsoidal.

Table 5.13. The Descriptive Statistics of Fuzzystats database cumulative periods.

		FormRatio	GSI	Compactness	CIRCUL1	CIRCUL2	Elongation
N	Valid	154	154	154	154	154	154
	Missing	0	0	0	0	0	0
Mean		,4020	2,7111	,8101	,6070	,5971	,6811
Std. Error of Mean		,0117	,0216	,0085	,0093	,0114	,0150
Median		,3878	2,6829	,8295	,6178	,5998	,6838
Mode		,0822[a]	2,1626[a]	,4279[a]	,2943[a]	,1258[a]	,3333[a]
Std. Deviation		,1450	,2684	,1052	,1158	,1412	,1856
Variance		,0210	,0721	,0111	,0134	,0199	,0344
Skewness		,2576	,3895	-1,2066	-,4422	-,4961	-,1608
Std. Error of Skewness		,1955	,1955	,1955	,1955	,1955	,1955
Kurtosis		-,4440	-,2282	1,9392	-,1388	,3179	-,9013
Std. Error of Kurtosis		,3886	,3886	,3886	,3886	,3886	,3886
Range		,6622	1,3215	,5319	,5382	,7344	,7396
Minimum		,0822	2,1626	,4279	,2943	,1258	,2604
Maximum		,7444	3,4841	,9597	,8325	,8602	1,0000
Sum		61,9041	417,5088	124,7518	93,4758	91,9520	104,8850

a. Multiple modes exist. The smallest value is shown

Table 5.14. The result of One-Sample Komogorov-Smirnov Test for Fuzzystats database.

		FormRatio	GSI	Compactness	CIRCUL1	CIRCUL2	Elongation
N		154	154	154	154	154	154
Normal Parameters[a,b]	Mean	,4020	2,7111	,8101	,6070	,5971	,6811
	Std. Deviation	,1450	,2684	,1052	,1158	,1412	,1856
Most Extreme Differences	Absolute	,0752	,0698	,1063	,0887	,0631	,0689
	Positive	,0752	,0698	,0779	,0362	,0408	,0444
	Negative	-,0633	-,0474	-,1063	-,0887	-,0631	-,0689
Kolmogorov-Smirnov Z		,9329	,8656	1,3186	1,1012	,7826	,8548
Asymp. Sig. (2-tailed)		,3489	,4419	,0618	,1768	,5726	,4580

a. Test distribution is Normal.
b. Calculated from data.

Std. Dev = ,15
Mean = ,4
N = 154,00
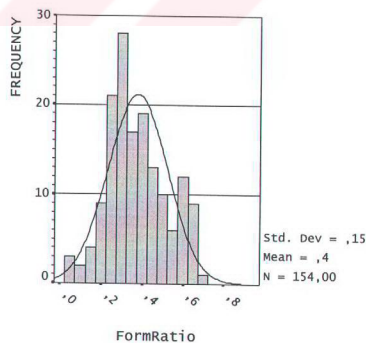
FormRatio

Figure 5.11. The frequency distribution of cumulative "Form ratio".
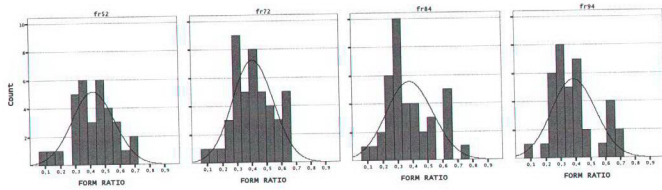
Figure 5.12. The frequency distributions of "form ratio" in all periods.

5.1.2.2. Grain Shape Index

In theory, the "Grain Shape Index" (GSI) variable approaches to Π for circles, 4 for squares and 2 for long skinny polygons. Larger values might be expected when the polygon is of fractal nature; however, there are no fractals in the study area so the maximum is 3,344 and the minimum is 2,178 (Table 5.13). This range states that the shape of the landslides in the region range from long skinny polygons to circles. The mean of the distribution is 2,71 and the standard deviation is 0,27. The distribution is slightly positive skewed and the kurtosis value is near to 0 yielding in there is no strong tendency to cluster around some values. The cumulative frequency distribution of "Grain Shape Index" is shown in Figure 5.13. When each period is explored individually it is seen that means are converging around 2,7. Slight changes occur through time and no net trend changes could be seen in "Grain Shape Index" (Figure 5.14).
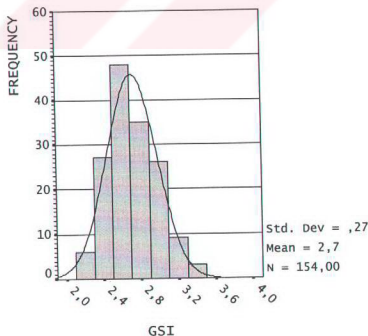
Figure 5.13. The frequency distribution of cumulative "Grain Shape Index".
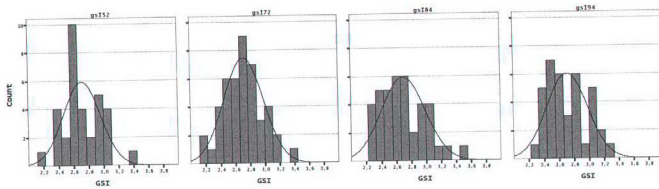
Figure 5.14. The frequency distributions of "form ratio" in all periods.

5.1.2.3. Compactness

The "compactness" variable measures the shape of the polygon with a maximum value approximating one for circles. The mean of the cumulative distribution of all periods is 0,8 and the standard deviation is 0,1. The negative skewness value and the positive higher kurtosis value suggest a distribution to have a long left tail and more clustering than a normal distribution (Figure 5.15). However, the non-parametric tests are just on the limit to consider it as a normal distribution (Table 5.14). Although the distributions mean itself has a converging value at 0.8, namely more circular than elongated, it conflicts with the form ratio when investigated in periodical basis. An increase in the long skinny polygons in 1972 period in the "form ratio" histogram is observed (Figure 5.12), in the same period in compactness histogram (Figure 5.16) it is observed that an increase in the near circular region.
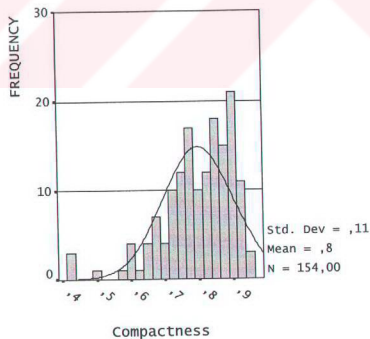
Figure 5.15. The frequency distribution of cumulative "Compactness".
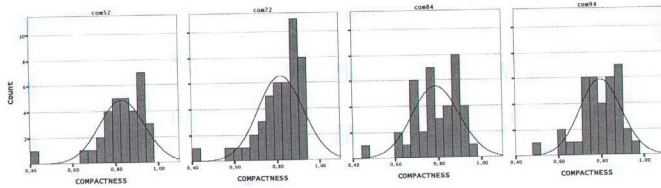
Figure 5.16. The frequency distributions of "compactness" in all periods.

5.1.2.4 Circularity 1

The "Circularity1" variable measures shape, reflecting the element's similarity to a circle, with a maximum value approximating to 1 for circles. The minimum is represented with 0,3 and the maximum is 0.83. This variable has a mean of 0,6 and a standard deviation of 0,11 (Figure 5.17). The negative skewness value implies a slightly longer left tail and the small negative kurtosis value implies less clustering. In the period analysis, it seen that the mean values decrease from 0,63 to 0,58, departuring from circularity field (Figure 5.18).
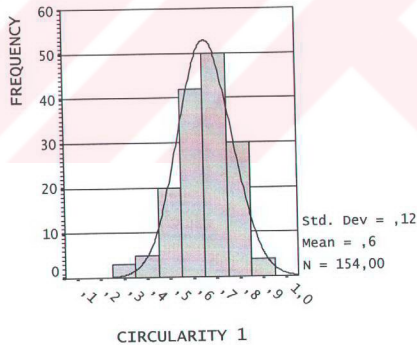
CIRCULARITY 1

Figure 5.17. The frequency distribution of cumulative "Circularity 1".
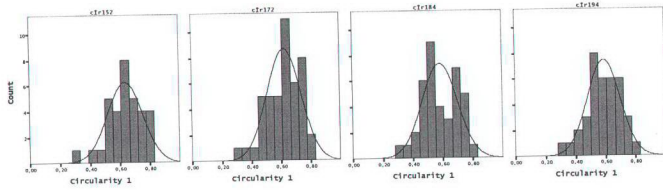
Figure 5.18. The frequency distributions of "Circularity 1" in all periods.

5.1.2.5. Circularity 2

Likewise "Circularity1" variable this variable also measures the circularity of the element with a maximum value approximating to 1. In the calculation of "Circularity 1" only area and maximum radius of the polygon is implemented, however in the calculation of "Circularity 2" both the minimum radius and the maximum radius is used rather than the area (Table 5.2). Although the calculation scheme is different than "Circularity 1", the mean and standard deviation of "Circularity 2" are similar as respectively 0.6 and 0.14 (Figure 5.19). However, the decrease in circular elements is more pronounced in 1972 period (Figure 5.20).
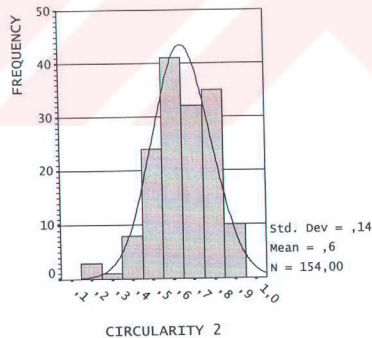
Figure 5.19. The frequency distribution of cumulative "Circularity 2".
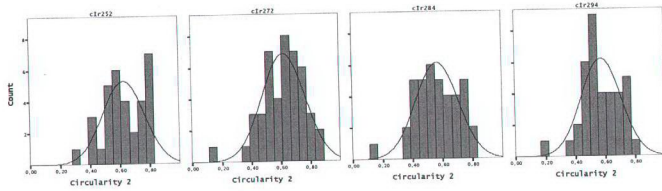
Figure 5.20. The frequency distributions of "Circularity 2" in all periods.

5.1.2.6. Elongation

The "Elongation" variable measures the ratio of the polygons short axis to long axis. The variable theoretically has a maximum value of 1 and a minimum as a convergent value to 0. The cumulative distribution has a minimum of 0,26 and a maximum approaching to theoretical maximum. The mean of the cumulative distribution is 0,68 and the standard deviation is 0,18. The negative skewness and kurtosis values imply that the distribution has a low tendency to cluster and has a slightly longer left tail (Figure 5.21). The effect of de-forestation is clearly seen on the 1972 histogram as a new peak around 0,6 is formed when compared into 1952 period (Figure 5.22). In the 1984 period the elongation decreases probably due to intensive re-forestation studies and in 1994 period a very minor increase can be attributed to the new road-cuts.
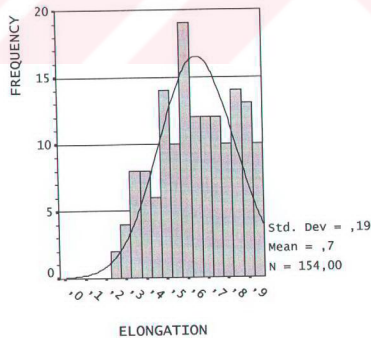
Figure 5.21. The frequency distribution of cumulative "Elongation".
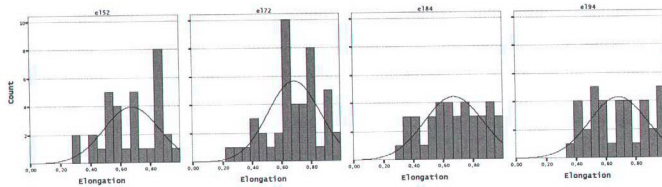
Figure 5.22. The frequency distributions of "Elongation" in all periods.

5.1.2.7. The significant changes and evaluation of fuzzy parameters with relation to Polystats database.

The effect of de-forestation, re-forestation and road construction is sometimes well represented in the frequency histograms of the fuzzy variables and sometimes due to the formulation they are very ignorant to the real fact. Although statistically the Polystats show no significant change in any of the periods there are some minor changes, which could lead some clues for the exploration of Fuzzy parameters. A brief table showing the changes in the Polystats parameters is given in Table 5.15. Although the 1952-1972 period in the table states that all the parameters are decreasing, the photo interpretation yields that this is because of newly generated landslides are relatively very small in area, in maxdim and in boundlen. Furthermore, in 1972-1984 periods, it is seen that the newly generated landslides are either disappearing or getting stable; nearly all the variables are converging to the values in 1952. It is obvious that when the effects of smaller landslides are taken off the distribution all of the variables should increase as seen in the table also. In 1984-1994 periods, the larger landslides remain relatively stable and some larger slides (larger than that of de-forestation stage) are developed in relation with the increased activity in the highway. Another striking feature is seen in the Table 5.15. As the mean values of all three variables show the same response in each period, such as decrease, increase and again decrease. This could be only attributed to generation of smaller slides (due to de-forestation) so the mean value is pulled down, vanishing of smaller landslides (reforestation) so the mean values are increased and again the mean values are pulled down (road activity) as newer slides are generated or the older small ones are reactivated.

112

Table 5.15. The changes through time in Polystats database

Parameter / Period	1952-1972 Intensive de-forestation Minimal road activity	1972-1984 Re-forestation and Minor road activity	1984-1994 Re-forestation and Intensive road activity
Number of landslide	▲	▼	▼
Area			
Min	▼	▲	▲
Max	▼	▲	▼
Mean	▼	▲	▼
Boundlen			
Min	▼	▲	▲
Max	▲	▼	▲
Mean	▼	▲	▼
Maxdim			
Min	▼	▲	▲
Max	▼	▼	▲
Mean	▼	▲	▼

The effects of deforestation are clearly seen in the "Form Ratio" variable as it has an increase in 1972 period in near 0.3-0.4, giving rise to an increase in log skinny polygons. However, the compactness variable, which is sensitive to circularity of the polygon, shows an increase that can be attributed to a larger increase in area than the perimeter. For the direct circularity measuring variables ("Circularity1" and "Circularity2"), a pronounced decrease is observed in the circular side rather than an increase in the lower (noncircular) values. However, an abrupt increase in elongation is observed in 1972 period. It is hard to say only looking to elongation graph of 1972 that the landslides are getting longer, but it could be concluded that the ratio of short axis to long axis is getting clustered around the mean, though the "Grain Shape Index" is having an increase in the values around 2.5 and 2.8. Based on all of the evidences reflected in Polystats and Fuzzystats variables the effects of de-forestation could be said to generate or re-activate flow type slope instabilities, of which the last word is left to be justified after the evaluation of photo characteristics database variables as they would sit on the true fact observations.

The effect of re-forestation activities is seen in 1984 and 1994 periods cumulatively. In the images of 1984 reforestation is started few years ago and yet the land cover is still not mature, in 1994 the land cover is very similar to that of 1952, mainly covered with dense forests. These are reflected as a decrease from both sides of the "Form Ratio" graph, clustering into mean values from 1984 to 1994, giving rise to more ellipsoidal shape. Namely the increase in elongation in 1984 is balanced with the

road cuts and dense forests in 1994. The "Grain Shape Index" responds to these changes as cutting off the extreme values and resulting in more homogenous distribution. This is probably due to the decrease in long axis and so the perimeter. Although the reforestation seems to be efficient, some outlier peaks are observed in the 2.4-2.5 field of "Grain Shape Index" 1994, which could be attributed to the landslides occurring by the road cuts. Also for the compactness variable the increase in 1972 period is trimmed in both 1984 and 1994 periods giving rise to a decrease in perimeter and area. Both of the circularity variables show slight increments in the circular side and decreases in the elongated side. A very sharp increase in the "Circularity 2" 1994 could easily be attributed to newly generated landslides caused by road cuts. The "elongation" graph of 1972 is trimmed down to mean values in the 1984 period especially in the 0,6-0,8 range and a very slight increment in 1994 period in 0,5 group. As a result in 1984 and 1994 periods, the newly generated flow type instabilities are either vanished or stabilized but other new or regenerated larger slides (larger than that of de-forestation) occur due to the road activities in the area.

5.2. Photo-characteristics Database

The photo characteristics of the landslides have been mapped and recorded in each year's relational database, which was also linked to the polygon landslide map. Each of the mapped seven attributes is investigated in this section in period basis. This exploration in the data probably will not yield in finding the direct causes, but it will be helpful for delineating the general scheme and general trends of the evolutions of slope instabilities in the Asarsuyu catchment and also will be used in conjunction with the Polystats and Fuzzystats databases. The definitions of the variables used in photocharacteristics database and the available items are presented in the previous chapter.

5.2.1. Massinfo

The "Massinfo" variable as described in the previous chapter in section 4.5.2 is the morphology seen in the photograph. The available items for this variable is Scarp&Path and Scarp&Body. The frequencies of the above mentioned items are shown with the data table in Figure 5.23. It is seen that there is a sharp increase in the 1972 period, following a continuous decreasing trend in both the number of landslides and in the S&B item.

	1952	1972	1984	1994
s&b	22	34	27	25
s&p	11	11	12	12
# of slide	33	45	39	37
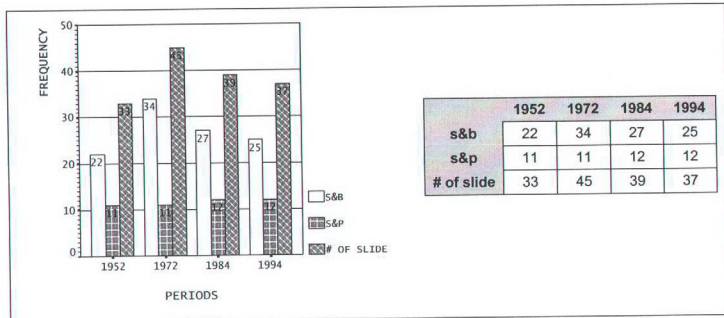
Figure 5.23. The frequencies of the mass info variable through time and data table

5.2.2. Type

The "Type" variable has significant drawbacks when a time dependent study is carried on, such as if the slide has started as a slide in the photo year; continued its activity as a flow in the period between two photo periods and it is seen in the next photo set as diminished, only the pathway of the flow and the scarp is seen. What code will be assigned to this slide is a big dilemma between the landslide investigators. A significant increase is seen at the 1972 period both in the slides and in the flows. After 1972 period although the number of slides decreases from 19 to 13 the percentage decreases only 9 % from 42% to 33% (relative to the number of total landslides) as the number of landslides decreases (Figure 5.24). Most of the landslides present in 1952 and 1972 period are diminished in 1984 and 1994 periods. The further analysis of the types will be handled in detail in further sections in the analysis chapter.

	1952	1972	1984	1994
Flow	17	25	26	25
Slide	15	19	13	12
Flow/slide	1	1	0	0
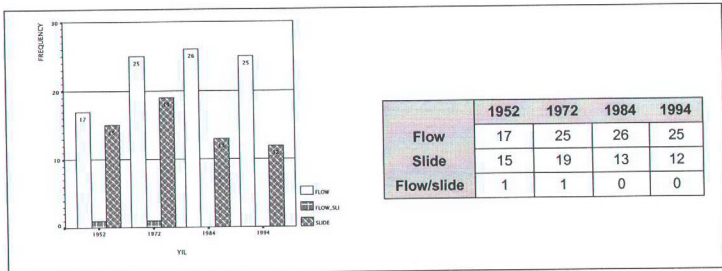
Figure 5.24. The frequencies of the type variable through time and data table

5.2.3. Style

The "Style" variable is in fact a similar variable to distribution of activity variable, like the very generalized form of it. It deals with the mass of the landslide if it is coherent and displays the same activity through its life span or is it sporadic, as it slides and stabilizes. This stabilization does not require the stabilization of the slided mass indeed the scarp of the first instability stabilizes. This small difference should have to be stressed as it the definition stabilizes the whole landslide then a variable called "distribution of activity" could only be attributed to multiple slides (Figure 5.25). Generally the landslides in the area are characterized by single landslides, however the number of multiple slides have a very slight increasing trend in the newer periods, with probable contribution other factors.

	1952	1972	1984	1994
single	27	38	30	29
multiple	6	7	9	8
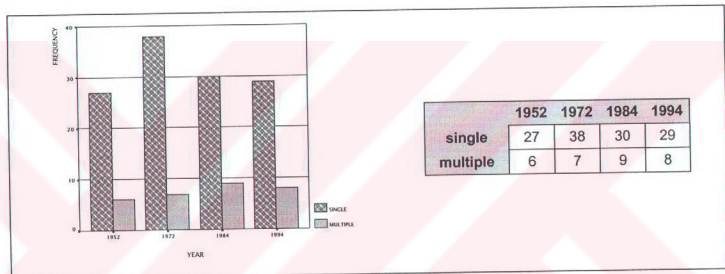
Figure 5.25. The frequencies of the style variable through time and data table

5.2.4. Depth

The "Depth" variable is totally dependent to the experience of the interpreter and the position of the parallax of the stereo-pair. A unified approach have been tried to fill the database column, as after the interpretation of the photographs, a second pass only for depth variable is conducted. The degree of depth is relatively chosen, dependent of the photo characteristics and the characteristics of the displaced mass. The landslides in the area are characterized dominantly by shallow slides (Figure 5.26). The quantity of deeper landslides decreases through time, where the shallow ones remain approximately same.

	1952	1972	1984	1994
deep	13	13	9	8
shallow	20	32	30	29
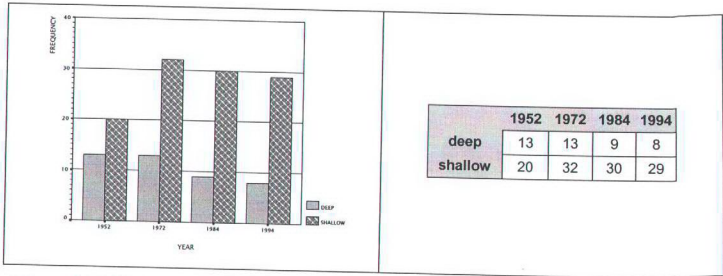
Figure 5.26. The frequencies of the depth variable through time and data table

5.2.5. Distribution of Activity

"Distribution of activity" is probably the most important parameter in evaluating the history of change in the landslides through time. Also the previously explained parameters such as "type" and "style" could be well understood with the help of this activity parameter. However, during photo interpretation there are some trivia situations such as the slide is diminishing via moving its slided mass or the slide is both enlarging and widening. Based on these special cases, the distribution of activity is re-grouped as activity originated in or by the scarp area yielding further development of the original slide, activity originated and confined in the slided mass and stabilization of the slide (Table 5.16). In the photo interpretation if no sign of activity is observed the slides gain an attribute as dormant.

Table 5.16. The re-grouping of distribution of activity variable in the photo database

Activity confined in scarp area	
	Advancing
	Enlarging
	Widening
	Combinations of the above
Activity confined in slided mass	
	Moving
	Diminishing
	Combinations of these
Stabilized slides	
	Dormant/inactive

As this analysis deals with the two periods of time the number of the slides used in this analysis is taken as the lowest amount of slide of the two years. Furthermore if the slide does not have an attribute as either scarp related or mass related the stable attribute is given. The striking fact in the graph presented in Figure 5.27 is the increase

in scarp in 1972-1984 periods that might be attributed to the new re-routing of E-5 highway. The mass related variable has an increasing trend with a very small slope through time and this might again be attributed to the construction of highway in the study area. When stayed out of the between variable changes, the starting period has a significant discrepancy within the variables themselves as the difference of mass related activity is started as 18, which means there are 18 cases which were not active in the 1952 period and activated in 1972. This is one of the clear evidence of de-forestation.

scarp	mass	Stabile	valid count	year
6	18	13	37	1994-1984
9	16	14	39	1984-1972
4	14	15	33	1972-1952
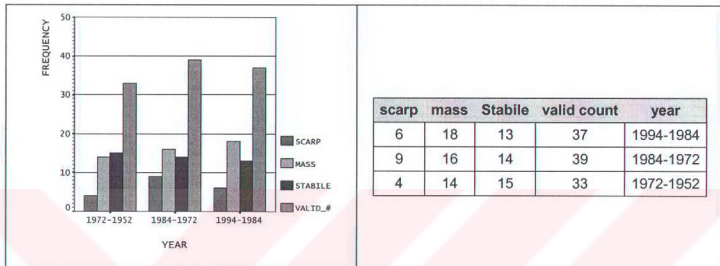
Figure 5.27. The frequencies of the distribution of activity variable through time and data table

5.2.6. Land cover

A "Land cover" attribute scheme is constructed with the help of field and remote sensing studies, and aerial photographical interpretation. The major elements for the scheme are dense forest, bare land, agricultural area, grassland and forest. There are also some associations present within the landslide polygon so possible combinations of the above major elements are also accepted. Although the landslide polygons do not reflect all of the area, they could be used as random sampled information source, so the below conclusions could have been derived.

The striking fact in the graph is the abrupt decrease of dense forest from 1952 to 1984 period and flawless increase in 1994 period. On the other hand, the increase in forest in 1984 period is the reason for increase of dense forest in 1994 period and shows the efficiency of re-forestation studies started at 80's. The only change in the land cover variable, except the former items, is the decrease in bare land amount, which might be an indicative of increased human activity (Figure 5.28).

denseforest	bare	agri	DF/Agri	GR/Agri	Grass	Bare/GR	DF/GR	Forest	Year
22	5	3	2	2	8	1	2	2	1952
8	20	3	2	2	8	1	1	2	1972
5	15	3	2	2	9	1	1	9	1984
22	5	3	2	2	8	1	2	2	1994
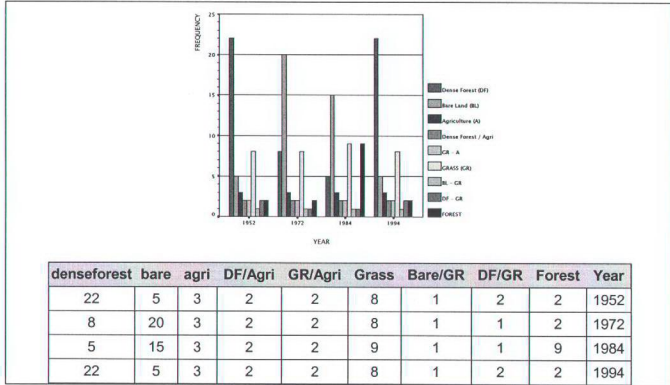
Figure 5.28. The frequencies of the land cover variable through time and data table

5.3. Landslide Attribute Databases

All of the parameter maps produced are crossed with the landslide inventory map and the landslide buffer map so that the attributes in each parameter map could be transferred to the inventory maps. The inventory map will have 4 attribute databases, of which the three of them are investigated in the previous sections of this chapter. The last attribute database, but the most important one is the transferred attribute database. It consists of (13) previously produced variables; the nature and the ranges are given in Table 5.17.

Table 5.17. The nature and ranges of transferred attribute database

Variable	Definition	Nature	Range (Min-Max) (unit)
LITHOMAP	Outcropping material	Nominal	-
DISTFAULT	Distance to Fault line	Ratio	$0 - 4791$ (m)
FAULTDENS	Density of Fault line in km^2	Ratio	$0 - 178$ (# / km^2)
ELEVMAP	Elevation above Mean Sea Level	Ratio	$220 - 1580$ (m)
DISTDRAIN	Distance to drainage lines	Ratio	$0 - 452$ (m)
DRAINAGE DENS	Drainage Density	Ratio	$12-352$ (#/km^2)
DISTRIDGEMAP	Distance to ridges	Ratio	$0 - 658$ (m)
ASPECTMAP	Aspect of the slopes	Ratio	$-1 - 359$ (degrees)
SLOPEMAP	Amount of Slope	Ratio	$0 - 56$ (degrees)
DISTSETTLEMENTMAP	Distance to settlement	Ratio	$0 - 6093$ (m)
DISTPOWER+ROAD	Distance to powerlines and roadnetwork	Ratio	$0 - 2312$ (m)
DISTE-5MAP	Distance to E-5 highway	Ratio	$0 - 8366$ (m)
LANDCOVERMAP	Type of the land cover	Nominal	-

All of the attributes are transferred to the central grid nodes of 25 meters by 25 meters. The resulting data set is consisted of 5493 rows of data. The aim of this section is to characterize the landslides by their attributes, to characterize the seed attributes and to compare the original available topography and other attributes with the slided masses attributes where needed. However a note should be added as the area of the landslides occur here as a natural weighting parameter, the larger the landslides the larger number of grid nodes they will have, this argument can be carried without any tension unless the number 8 landslide is omitted. The number 8 has no activity through all of the periods. This landslide is known as Kom Landslide and believed to be a huge dormant slide, creating its own stable parameters (probably being the oldest landslide in the dormant slide in the area), which should not be mixed with the rest. Also this landslide is represented with 1346 nodes out of 5493 nodes, nearly 24% percent of the total nodes. In order to remove the effect of this dormant landslide the nodes of his slide are not used so the remaining node population is decreased down to 4147 nodes.

For all of the parameters a comparison will be given in the proceeding sections with the seed data distributions. The seed data is extracted using a 100-meter buffer and the microcatchment boundaries. The decision rule for defining the seed zone is: If the distance between the slide boundary and microcatchment divide line is smaller than 100 meters then use the microcatchment divide line, if the distance is larger than 100 meters then use the 100 meter buffer line for the seed zone generation. Following the seed node generation, the same attributes as the slided mass is also transferred, and the number of the seed nodes is 4430. The seed cells of the Kom landslide is included, because it is coherent with the similar landslides in the valley and due to the seed cell selection decision rule the number of the seed cells have already been decreased significantly. This inclusion stands for the sake of the factual data preservation.

5.3.1. LITHOMAP

The "LITHOMAP" of the slided mass gives out some preferred conditions like the sum of the first three most preferred lithologies are approximately 92.9% which is quite homogeneous and remarkable (Figure 5.29). Landslides exist in 8 lithologies, but the area has 11 lithologies. The landslide missing lithologies are green schist facies of Yedigöller Formation and Asarsuyu formation. There are no landslides seen in the gypsum, this is due to the fact that the huge landslides are not taken into account in this analysis.

In the LITHOMAP of seed cells, the situation slightly changes, though the preferred conditions are still the same. The cumulative total of the preferred first three lithologies are 89,5 % which is 3,4% lower than that of slided mass. Moreover the

distributions in the preferred lithologies are also changing. While percentage of Aksudere formation is increasing the percentage of Yedigöller (21,4 % to 18,8 %) and Fındıklıdere (43.4 % to 32,6 %) formations are decreasing. Also in the remaining 10% of the data some slight changes have been observed as the percentage of Quaternary is decreased as expected and Çaycuma formation is increased.

Figure 5.29. The preferred lithologies and their percentages: a) slided Mass, b) the seed cells.

Although the preferred lithologies represent 92.9% in slided mass and 89.5 % in seed cells they can only range up to 54.08 % in the study area (Figure 4.2). The remaining approximately 45 % of the available lithologies in the study area is only reflected to slided mass and seed cells as approximately 10%, which are direct indications that they are not preferred by landslides.

5.3.2. DISTFAULT and FAULTDENS

The distance to fault and the density of fault variables are two highly negative correlated variables (Table 5.18). In theory, the distance to fault should be decreasing when the density of the fault lines are increasing. The negative correlation is much better represented in the seed cells rather than the slided mass, as indicating the seed cells are much more sensitive to faults. The comparison of minimum, maximum and the mean values of both the slided mass, seed cells and the whole data is given in Table 5.19.

It is seen that the maximum value of the landslide data can go up to 1441 and the seed cells have a maximum of 1517 that is nearly one third of the maximum of the whole data (Figure 4.5), which means there is redundant information in the whole data, not relevant to landslides. On the other hand the maximum of the FAULTDENS are nearly the same both in observed and in whole data (Figure 5.30). This is because of

the spatial orientation of both the landslides and the fault lines. A polynomial distribution is seen in the frequency histogram of FAULTDENS as having peaks around 50, 225 and 600. These could be attributed to landslides very distant to faults, landslides near to one set of faults and landslides near to two or more sets of faults. This could easily be seen when the thematic maps of the fault density is investigated (Figure 4.6). Hence, it can be said that landslides are not much far away than 1500 meters to the fault lines and the preference order related to density is: near to one set of fault, very far away from faults and the least preferred as crisscrossing of two or more sets of fault lines, which is valid also for the seed cells.

Table 5.18. The correlation state of DISTFAULT and FAULTDENS variables: a) slided mass, b) seed cells.

Correlations (a)

		D_FAULT	DENSFAY
D_FAULT	Pearson Correlation	1,000	-,648**
	Sig. (2-tailed)	.	,000
	N	4147	4147
DENSFAY	Pearson Correlation	-,648**	1,000
	Sig. (2-tailed)	,000	.
	N	4147	4147

** Correlation is significant at the 0.01 level (2-tailed).

Correlations (b)

		D_FAULT	DENSFAULT
D_FAULT	Pearson Correlation	1,000	-,703**
	Sig. (2-tailed)	.	,000
	N	4430	4430
DENSFAULT	Pearson Correlation	-,703**	1,000
	Sig. (2-tailed)	,000	.
	N	4430	4430

** Correlation is significant at the 0.01 level (2-tailed).

Table 5.19. The comparison of whole data and landslide data

		Slided mass data	Seed cells	Whole Data
DISTANCE TO FAULT	Min	1	0	0
	Max	1411	1517	4791
	Mean	361,47	408,52	725,887
	St.Dev	286,44	339,66	827,725
FAULT DENSITY	Min	0	0	0
	Max	671	705	768
	Mean	229,1	235,56	195,193
	St.Dev	163,58	161,03	154,61
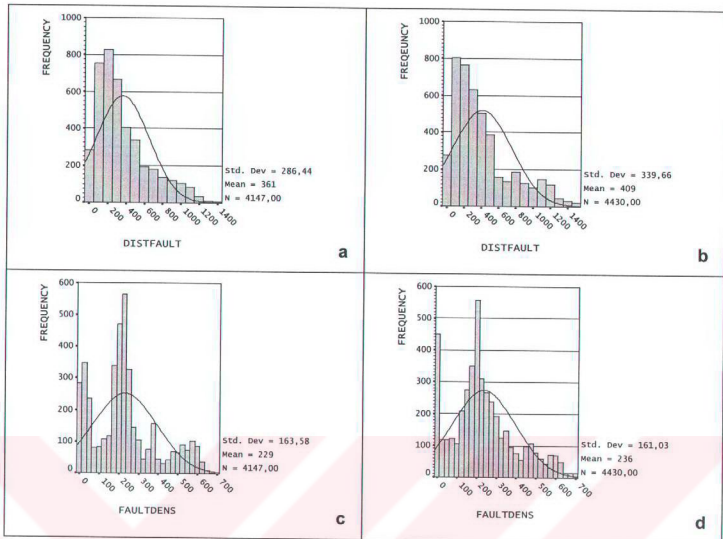
Figure 5.30. The frequency distributions of DISTFAULT (a. slided mass, b. seed cells) and FAULTDENS (c. slided mass, b. seed cells).

5.3.3. ELEVMAP

Although the available elevation data in the study area ranges between 220 and 1580 meters, the landslides are observed only between 295 and 1095 meters and their seed cell values ranges between 300 and 1130 meters. This is probably at higher elevations more sound and intact rocks are present. The descriptive statistics of the three groups are given in Table 5.20. The majority of the data is distributed in the range of 350 to 800 meters. An imperfect double peak is observed in the distribution, resulting that some means of preference is present and concentrated around 700 and 400 meters, but either near/equal preferences exist or the preferences are indistinguishable.

Table 5.20. The descriptive statistics of the ELEVMAP.

		Slided mass data	Seed cells	Whole Data
ELEVATION	Min	295	300	220
	Max	1050	1130	1580
	Mean	594,62	661,84	679,96
	St.Dev	177,191	179,21	265,06
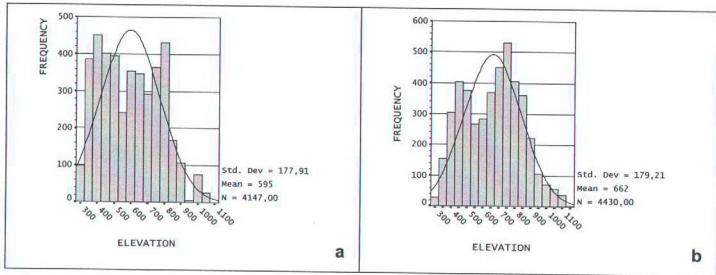
Figure 5.31. The frequency distributions of ELEVMAP, a. Slided mass, b. Seed cells

5.3.4. Distance to drainage

The distance to drainage variable is the distance to all drainage lines without any ordering in between the drainage lines. The descriptive statistics of the drainage related parameter maps are presented in Table 5.21.

Table 5.21. The descriptive stats of distance to drainage variable group

		Landslide data	Seed Cells	Whole Data
DISTANCE TO DRAINAGE	Min	0	0	0
	Max	342	397	452,948
	Mean	87,16	118,3	97,56
	St.Dev	66,29	76,99	73,144

The frequency histograms of the drainage related parameters of the landslide data are given in Figure 5.32. The shapes of the cumulative histogram of both the landslide data and the whole data (Figure 4.13.a, b) and the first order strahler order streams are very similar to each other, due to the abundance of first order streams.
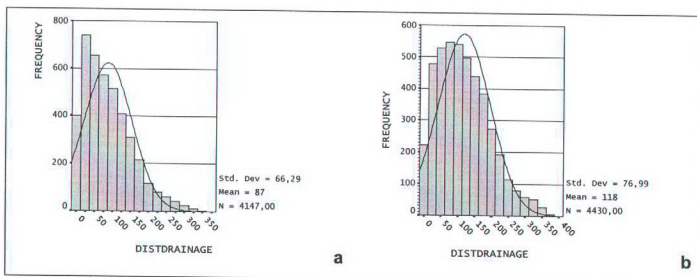
Figure 5.32. The frequency distributions of Distance to Stream map, a. Slided Mass, b. Seed Cells.

5.3.5. Drainage Density

The drainage density data in the area is ranging from 12 to 352 drainage lines in one square kilometer. The seed cells of the landslides are characterized by a range of 55 to 287, no significant preference is observed as the mean and standard deviation of the whole data and seed cells are similar. The difference in range does not change the distribution parameters, only a fraction of the whole data is represented with same weightings. The drainage density attribute frequencies of slided mass and the seed cell is presented in Table 5.22.

Table 5.22. The descriptive statistics of the DRAINAGE DENSITY.

		Slided mass data	Seed cells	Whole Data
DRAINAGE DENSITY	Min	58	55	12
	Max	292	287	352
	Mean	176.33	160	160.12
	St.Dev	42.53	37	50.22
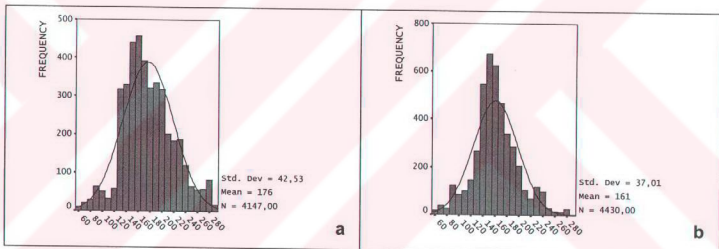
Figure 5.33. The frequency distributions of Drainage Density, a. Slided mass, b. Seed cells.

5.3.6 DISTRIDGE

All of the distance to ridge parameter distributions are similar to each other (whole data & landslide related data), which is an indication of no significant distribution free preference is present. However, a genetic preference is seen in the distributions as both the whole and landslide data exhibit clustering around smaller values. This situation can be summarized as both the area and the landslides do tend to prefer nearer distances to ridges, namely to the microcatchment divide lines. They tend to be located in the upper parts of the slopes. Although the distributions are similar, here are

slight changes like, the Distance to Ridge of seed cells seems to have a more sharp peak around 25-100 meters of distance than that of the slided mass data as expected (Figure 5.34). The descriptive statistics and the comparison of the data are shown in Table 5.23.
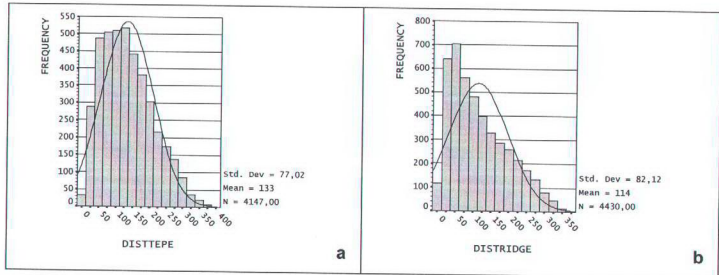
Figure 5.34. The frequency distributions of Distridge, a. Slided mass, b. Seed cells.

Table 5.23. The descriptive stats of distridge

		Slided mass data	Seed cells	Whole Data
DISTANCE TO RIDGE				
	Min	4	4	1
	Max	377	364	658,97
	Mean	132,58	113,88	138,172
	St.Dev	77,02	82,12	99,86

5.3.7. Aspect

Although the aspect distribution of the whole data have systematical errors arising from the gridding model, no significant direction is seen. However, in the slided data and in the seed cells, the north facing slopes exhibit a very distinctive preference (Figure 5 35). The descriptive statistics and the comparison of the data is shown in table 5.24.

Table 5.24. The descriptive stats of aspect

		Slided mass data	Seed cells	Whole Data
ASPECT				
	Min	-1	-1	-1
	Max	358	358	359
	Mean	156,36	168,37	182,45
	St.Dev	137,07	126,25	110,08
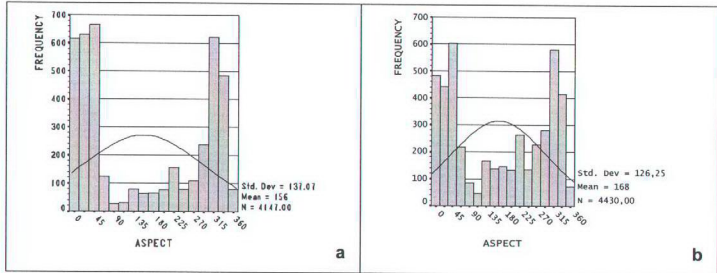
Figure 5.35. The frequency distributions of Aspect, a. Slided mass, b. Seed cells.

5.3.8.Slope

The slope distribution in the study area has a broader distribution than the landslide data (Figure 4.16), also this is seen in the descriptive statistics of the slope of landslide related data as the means are similar to each other (Table 5.25). The landslide related data (Figure 5.36) is much more clustered than the original available data, due to the fact that a very small portion of the landslide related data are associated with alluvial slopes that are gentler than the rest of the data.

Table 5.25. The descriptive stats of Slope

		Slided mass data	Seed cells	Whole Data
SLOPE	Min	0	1	0
	Max	52	51	56
	Mean	21,44	22,41	16,97
	St.Dev	8,88	9,12	10,46
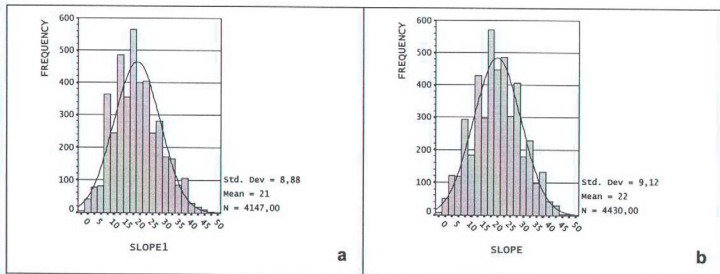
Figure 5.36. The frequency distributions of Slope: a. Slided mass, b. Seed cells.

5.3.9. Distsettlement

The distances to settlement parameter histogram has a significant background noise, reflecting the irrelevant data present in the whole study area (Figure 4.18.b). This irrelevance is also seen in the compared descriptive statistics, as the maximum values of the landslide related data is quite less than that of the whole data (Table 5.26). The magnitude of irrelevance is much better seen when the slided data and the seed cell histograms are explored (Figure 5.37). On the other hand both of the landslide related histograms show a bimodal distribution having one peak around 150-200 meters and the other around 1800 meters apart from the individual settlements. The first one could easily be attributed to the effect of households and the other should be attributed to external other factors.

Table 5.26. The descriptive stats of Distsettlement

		Slided mass data	Seed cells	Whole Data
DISTANCE TO SETTLEMENT	Min	2	4	0
	Max	2218	2629	6093,46
	Mean	700,04	699,45	1258,46
	St.Dev	639,12	651,25	1370,20
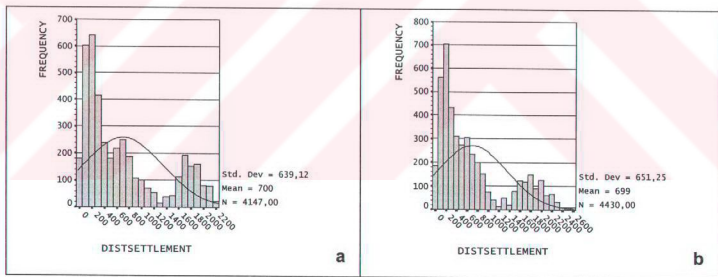
Figure 5.37. The frequency distributions of Distsettlement, a. Slided mass, b. Seed cells.

5.3.10. Distpower & Dist-Roadnetwork

The distance to power lines and road network has also significant low frequency noise in the histogram shown in Figure 4.18.h. The means also suggest a preference in smaller values, which are nearer to power lines and road network. Furthermore, the difference in maximums are quite striking as the whole data range up to 5566 meters

but the slided data and seed cells range only up to 700 meters approximately (Table 5.27). The nearly 5000 meters of difference is the source of background noise in the whole data histogram. On the other hand no significant preference is seen in the frequency histograms of the slide related data, only a general clustering is present in the smaller values, which are nearer to power lines and road network. This could be attributed to the fact that the land cover of the region is disturbed during the construction of power lines and the road network. The descriptive stats of the distance to power lines and road network parameter map are presented in (Table 5.27). The frequencies of merged vector of power lines and road network are presented in Figure .5.38.

Table 5.27. The descriptive stats of Distpower, Dist road & Distroad+Distpower

			Slided mass data	Seed cells	Whole Data
DISTANCE TO POWER & ROAD NETWORK		Min	0	0	0
		Max	547	683	2312,68
		Mean	184,53	183,78	246,121
		St.Dev	119,31	136,88	302,099
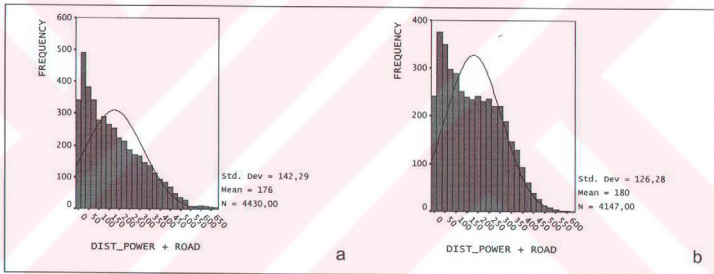
Figure 5.38. The frequency distributions of Dist-power+road a. Slided mass, b seed cells.

5.3.11 Dist E-5 Highway

The E-5 highway is probably the most important and active highway among the Turkey's national highways, which connects İstanbul and Ankara. The high traffic activity of this road creates more vibration and local load than the other roads in the area, based on this the road itself is extracted from the cumulative road network in the area as a separate parameter. Both of the slided mass and seed cell frequency histogram of the distance to E-5 highway reveals important information. The bimodal distribution of these histograms suggests that the landslides prefer two conditions. One group is caused by the presence of E-5 highway as the frequency of seed cells show a large peak in the

values smaller than 400 meters. The other group is very distant to the highway, in the order of kilometers, at which the effect of vehicle vibration would be minimal, resulting in the conclusion that the landslides should be caused by other parameters.

Table 5.28. The descriptive stats of Dist_E-5 Road

		Slided mass data	Seed cells	Whole Data
DISTANCE TO E-5	Min	0	0	0
	Max	3101	3112	8366,8
	Mean	1254,62	1217,83	2467,49
	St.Dev	737,07	899,59	1771,85
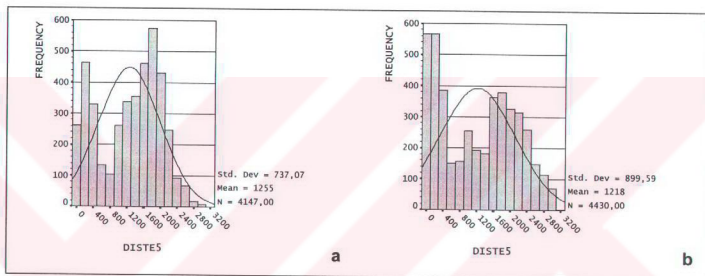
Figure 5.39. The frequency distributions of Distance to E-5 highway, a. Distroad, b. Diste-5

5.3.12. Land cover

The 93% percent of the land cover units in the study area is represented with three units (Dense Forest 44%, Mixed zone 33%, Young forest 16%) after the maximum likelihood classification of Landsat TM5 (Figure 4.23). Although the cumulative percentage of the same land cover units does not change in the slided mass and seed cells, the proportions of the Landcover units significantly change (Figure 5.40). For example the three units are represented as 93.9 % in the slided mass data and 95.1% in the seed cells. The main differences in the whole data and landslide data is observed as an increase in mixed zone and forest unit (Table 5.29), on the other hand there is a sharp decrease in the dense forest as expected. In other words, landslides are less associated with dense forest unit.

Table 5.29. The % change of Landcover units

	Whole data	Slided mass	Seed cells
Mixed	33%	42,5%	38,2%
Dense forest	44%	26%	28,1%
Forest	16%	25,4%	26,8%
Total	93%	93,9%	95,1%
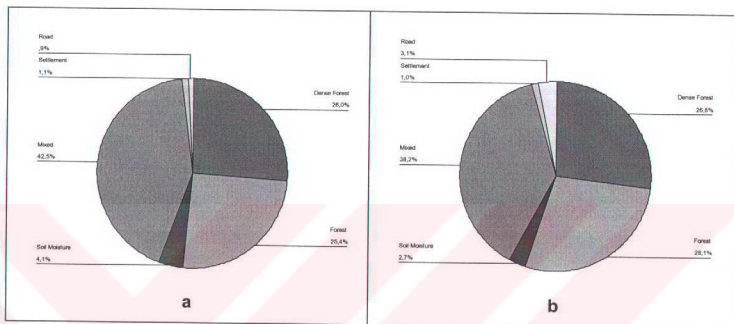
Figure 5.40. The percentage distribution of Landcover units, a.) slided mass, b) seed cells.

HAZARD ASSESSMENT

In this chapter various methods are considered to analyze the presented data in Chapters 4 and 5, in order to assess the landslide hazard using Geographical Information Systems.

6.1. Thematic Landslide Attribute spatial distribution analysis (TLASDA)

This is the simplest type of analysis that only reflects the current situation as where the landslides have occurred in a selected period. Such a map only shows the outlines of the individual landslides not the general scheme of landslide hazard in the area. However, such thematic maps are useful for gathering the information about the frequency, type, and depth. No direct information is present for the possible causes of the landslides, as the map does not contain parameter information. Nevertheless, this type of analysis provides a visual input showing the severity of the hazard with regard to the attribute of the landslides.

19 combinations of attribute maps are available for the cumulative of 4 periods, however only the maps of the last period (1994) are presented here in this chapter. All of the maps used in the analysis are converted to binary vector maps using their attributes in order to give more visual perception. The "Massinfo" attribute has two available items as "scarp & body" and "scarp & path". Although the map shows no clear preferences, the map reveals that the larger the landslides, they possess "scarp & body" attribute (Figure 6.1).

The "Type" attribute has two available items as "flow" and "slide". The flow type shows close relation with the E-5 highway as most of the flow dominated landslides occur in the Bolu mountain pass of the E-5 highway (Figure 6.2).

The "Depth" and "Style" attributes show similar spatial preferences. The "Style" attribute has two available items as "single" and "multiple" and the "Depth" attribute has two available items as "shallow" and "deep". The single type landslides are generally smaller landslides (Figure 6.3 and Figure 6.4).
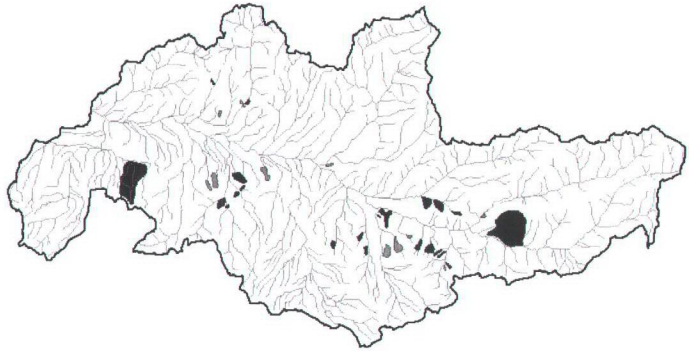
Figure 6.1. The Thematic Landslide Attribute spatial distribution of massinfo attribute of 1994 photo characteristics database, where black areas show Scarp & Body and grey areas show Scarp & Path attributes.
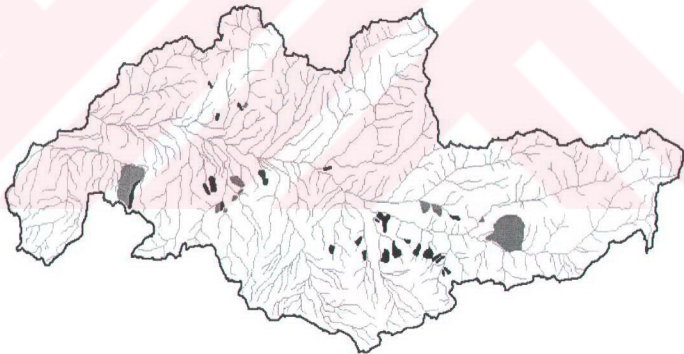
Figure 6.2. The Thematic Landslide Attribute spatial distribution of type attribute of 1994 photo characteristics database, where black areas show Flow and grey areas show Slide attributes.
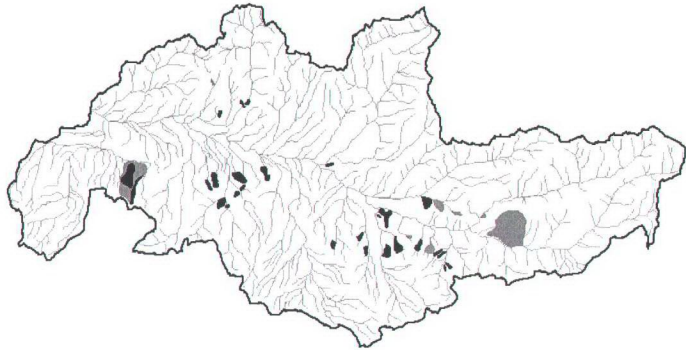
Figure 6.3. The Thematic Landslide Attribute spatial distribution of style attribute of 1994 photo characteristics database, where black areas show Single and grey areas show Multiple attributes.
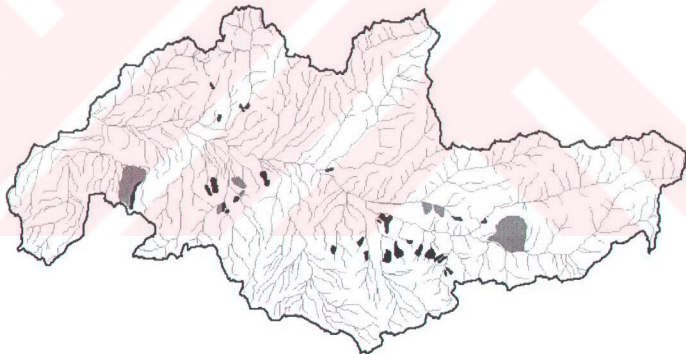
Figure 6.4. The Thematic Landslide Attribute spatial distribution of depth attribute of 1994 photo characteristics database, where black areas show Shallow and grey areas show Deep attributes.

The "Distribution of Activity" attribute has two available items as "Scarp related activities" and "mass related activities". Although not significant, the landslides near to E-5 highway resembles more activity than the rest of the landslides (Figure 6.5).
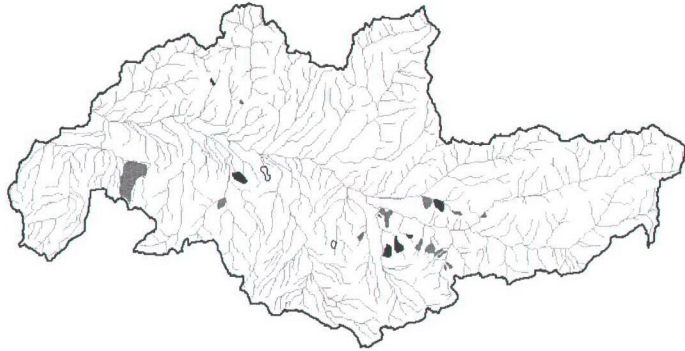
Figure 6.5. The Thematic Landslide Attribute spatial distribution of activity attribute of 1994 photo characteristics database, where black areas show Scarp, grey areas show Mass attributes, white polygons with black border show stable attribute. Polygons which do not have any attribute in this field are not shown.

Due to the low level of this analysis, only some general statements regarding the whole area could be generated. This analysis was first developed for regional analysis, even for national scale analyses. Based on the very coarse resolution of the analysis, the statements are very general and their validity is suspicious. After the investigation of the above maps the severity of the landslide hazard is found to be dominant in the southern slopes of the Asarsuyu catchment, in the northern slopes hardly any landslide occurs. The eastern part which constitutes the higher parts of the catchment is also landslide free. Landslides are generally occurring near to the E-5 highway and near to the greater landslides. Although these distribution maps could be overlaid with other available maps, this was not done, as parameter versus attribute information will be used in other analyses.

6.2. Landslide activity analysis (LACTA)

The TLASDA and the analyzed data in the previous chapter do not yield any information about the trend of activity changes. They do not provide any information regarding the individual landslides. This analysis is consisted of a two-dimensional matrix resulted from non-graphical querying of the constructed GIS database. The "Distribution of Activity" column of the photocharacteristics database is used in constructing this matrix. First the value fields in the "Distribution of Activity" column is reclassified in to 5 new value items, in coherence with section 5.2.5. Two new value

items are introduced to the matrix as "present" and "absent", in order to consider the total number of the landslides. The details and probable meaning of the crossed items are presented in Table 6.1.

Table 6.1. The two dimensional matrix of LACTA

		Activity in older photo set				
		Scarp	Body	Present	Absent	Dormant
Activity in younger photo set	Scarp	No change, still activity confined in scarp area	The activity migrates from body to scarp area	Reactivation in scarp area	New landslide	New landslide
	Body	The activity migrates from scarp to body area	No change, still activity confined in body area	Reactivation in body area	New landslide, development stage is not recorded	New landslide, development stage is not recorded
	Present	No activity can be classified but still landslide is present	No activity can be classified but still landslide is present	No change	New landslide, development stage is not recorded	New landslide, development stage is not recorded
	Absent	Landslide becomes invisible	Landslide becomes invisible	Landslide becomes invisible	No change	Landslide becomes invisible
	Dormant	Stabilization	Stabilization	Stabilization	New landslide, development stage is not recorded	No change

Although the row named "absent" in the "activity in younger photo set" could be attributed to some errors of photo interpretation. Once there occurs a slide there should be its relics. It should be noted that either the land cover or some human activities had viped out the relict features of the landslide.

In the matrix of 1952 and 1972 (Table 6.2) all of the value items are clustered around "present" and "absent" as 1952 period is the first photo set, hence no activity information before 1952 could be found. The increase in number of landslides is reflected in the matrix in "absent – present" pair, as 13 new landslides. 14 slides are reactivated showing activation in the body area, 14 landslides are still present but no signs of activity could be seen. 4 landslides possess activity in the scarp area, and only one slide is vanished.

In 1972-1984 matrix the striking result is the "body-body" pair, it constitutes 10 out of 47 slides (Table 6.3). This could be attributed to the reactivation in the slided mass area of older slides in the study area. Two new slides are recorded in 1984 period and 8 landslides are vanished.

Table 6.2. Two-dimensional matrix of 1952 and 1972 periods.

		Activity in 1952 photo set					
		Scarp	Body	Present	Absent	Dormant	Σ
Activity in 1972 photo set	Scarp	-	-	4 (8.7%)	-	-	4
	Body	-	-	14 (30.43%)	-	-	14
	Present	-	-	14 (30.43%)	13 (28.26%)	-	27
	Absent	-	-	1 (2.17%)	1	-	2
	Dormant	-	-	-	-	-	-
	Σ	-	-	33	14	-	

Table 6.3. Two-dimensional matrix of 1972 and 1984 periods.

		Activity in 1972 photo set					
		Scarp	Body	Present	Absent	Dormant	Σ
Activity in 1984 photo set	Scarp	3 (6.38%)	1 (2.13%)	4 (8.51%)	1 (2.13%)	-	9
	Body	1 (2.13%)	10 (21.28%)	5 (10.64%)	-	-	16
	Present	-	-	12 (25.53%)	1 (2.13%)	-	13
	Absent	-	3 (6.38%)	5 (10.64%)	-	-	8
	Dormant	-	-	1 (2.13%)	-	-	1
	Σ	4	14	27	2	-	

In 1994 period two more landslides are vanished, The main activity is continuing in body areas ("body-body" pair) (Table 6.4). On the other hand, 10 landslides are observed as present reflecting that no significant activity could be recorded.

Table 6.4. Two-dimensional matrix of 1984 and 1994 periods.

		Activity in 1984 photo set					
		Scarp	Body	Present	Absent	Dormant	Σ
Activity in 1994 photo set	Scarp	5 (12.82%)	1 (2.56%)	2 (5.13%)	-	-	7
	Body	3 (7.69%)	14 (35.9%)	-	-	-	17
	Present	-	-	10 (25.64%)	-	-	10
	Absent	1 (2.56%)	1 (2.56%)	-	8	-	10
	Dormant	-	-	1 (2.56%)	-	1 (2.56%)	2
	Σ	9	16	13	8	1	

When all of the matrices are explored together, it is evident that the main activity is confined to body related activities, probably some forms of flow type. However, it could not be denied that the scarp activities are also in an up trend, probably notifying that the area is becoming geomorphologically mature, so the landslides. As a de-forestation re-forestation cycle has been observed in the area, probably the area is reshaping itself due to the changing geomorphological constraints. In the deforested period new landslides and intense surface processes (soil creep, erosion, etc.) took place changing the kinematical dynamics, yielding in the creation of new down gradient forces of nature. Although the number of landslides decreases through 1972 to 1994, the activity states even become more instable, not generating new slides but activating the older ones.

6.3. Landslide Isopleth analysis (LIA)

This analysis is the most basic quantitative analysis that could be applied to landslide hazard evaluation, and is extensively used in any kind of spatial and attribute dependent data. The backbone of this analysis is that it consists of a counting circle. The circle is moved through the map with a constant offset and at each location the points falling into the counting circle is counted and recorded in the center of the counting circle. Generally a counting circle of 1 km^2 area is used as a convention to call the density at each point as per km^2. The offset amount is generally selected as the half radius of the counting circle. In this case to achieve the 1 km^2 area a radius of 564 meter is used and the offset is taken as 250 meters (Figure 6.6).

The resulting text file is then linearly interpolated and a continuous density surface is formed using ordinary kriging. This raster file is then converted into vector via contouring using cubical convolution algorithm. The isopleth intervals in contouring are selected as 20 percent (Figure 6.7).

The quality and resolution of the isopleth map is strongly dependent on the circle size and the offset distance. The larger the size of the search circles the lower the resolution is and the greater the generalization. However, small circle sizes are creating redundant information as when size gets smaller and smaller the resultant map gets similar to the original landslide inventory map. On the other hand, the offset size of the circles also affects the final product. When offset distance exceeds the radius of the counting circle the reason for doing an isopleth map vanishes as the data turns into discrete rather than continuous data.
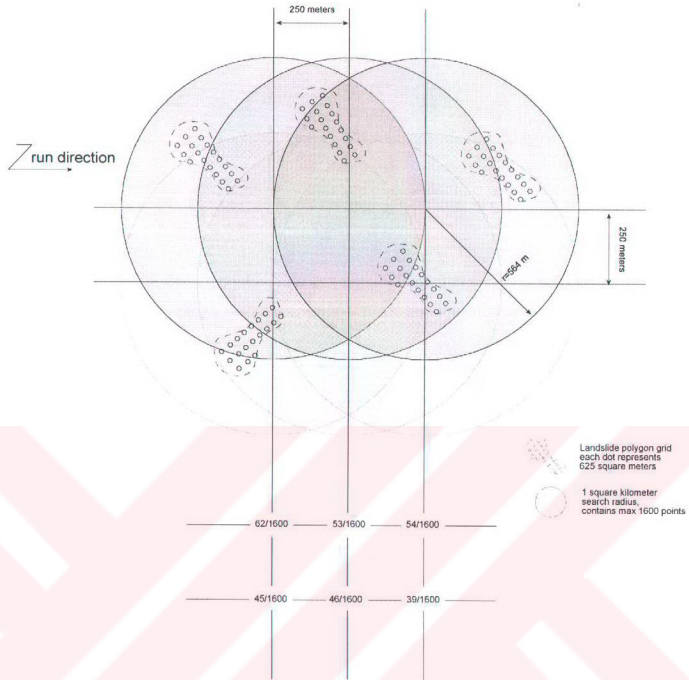
Figure 6.6. The mechanism of the isopleth analysis

The resultant isopleth map of Asarsuyu catchment with 564 meter radius and 250 meter offset reveals information quite conformable with the non-spatial results of the previous chapter. In order not to create duplication the overlay of every parameter map is not shown here, only significant results are listed below.

1. The E-5 road has cut through the maximum density areas of the isopleth

2. Few settlements are located in the 80-100% interval of the isopleth, but other intervals have a dispersed layout.

3. The geological units show significant preferences, as the greenschist facies of Yedigöller formation has no landslides, and the flysch sequence has the most landslides.

4. Strong relationship with the fault density is seen especially in the lower density intervals
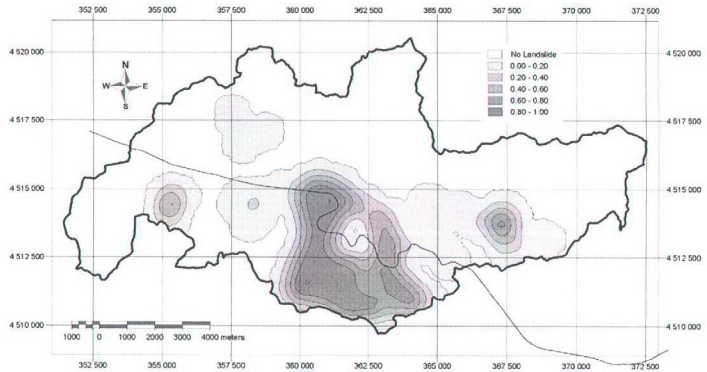
Figure 6.7. The isopleth map of the Asarsuyu catchment, each contour level represents 20 % landslide density.

6.4. Statistical analyses

Two types of statistical analyses are planned to be carried out in the Asarsuyu catchment for the estimation of future landslide hazards. The aim to use statistical methods is to increase the objectivity of the assessment, and to let the data derive its own decision rules.

To achieve this goal, as explained in earlier chapters, the nodes of seed cells are used as decision rule generator or training samples. 4430 seed cells are introduced to a database containing all of the available variables in the system.

However, a major impediment is compromised in the nature of the data as all of the bi-variate methods used are designated for some form of categorical data not for continuous data sets. This arouses from the fact that the all-available bi-variate methods base themselves to the landslide density or abundance in certain parameter classes. If the continuous data is used as it is, the densities will be calculated for the whole map and not even a single natural preference in the area will be utilized for hazard assessment. Consequently, a continuous to discrete categorical conversion seems to be indispensable. Some efforts have been carried out to categorize some continuous data in the literature in the last century, unfortunately any single example in landslide hazard assessment or in geosciences are affirmed. Some authors of mathematical, medical and statistical experience have proposed methods of conversion. Generally these methods depend on the optimum bin width classification of the histograms of various parameters, and further they do not have a spatial dimension. The earliest published rule for selecting the bin width appears to be that of Sturges (1926). This proposal is more of a

number-of-bins rule rather than a bin width rule itself, but essentially amounts to choosing the bin width.

$$\hat{h} = \frac{\text{Range of Data}}{1 + \log_2 n}$$

where n is the sample size

However, Scott (1979, 1992) showed that this bin width leads to an over smoothed histogram, especially for large samples and proposed an unbiased estimation of a probability density function, which is achieved when:

$$W = 3.49 \, \sigma N^{-1/3}$$

Where W is the width of the histogram bin, σ is the standard deviation of the distribution and N is the available samples. This estimator worked well for Gaussian distributions, where it led to overlay large bin widths and hence over smoothing. Friedman and Diaconis (1981) suggested a more simple method:

$$W = 2(\text{IQR})N^{-1/3}$$

Where W is the width of the histogram bin, IQR is the inter quartile range (the 75th percentile minus the 25th percentile) and N is the sample size. Numerical comparisons by Emerson and Hoaglin (1983) of the Scott and Freedman-Diaconis (FD) rules showed the FD rule led to narrower bin widths, although in practical applications the two rules will often lead to the same choice of interval width (Izenman, 1991).

All of the methods cited above has a number of disadvantages which are listed below:

- Over smoothed class divisions
- Distribution dependent
- Valid for only one parameter map
- Cannot be applied to multi-modal distributions (multi-modality overrides the assumptions)

Due to the cited disadvantages of former methods a new method is proposed to classify the continuous data sets into categories, which could be called as "percentile method" (Figure 6.8). The core of this method is dependent on equal frequencies. The frequency domain is quite free of distribution parameters and could easily be implemented in bimodal or multi-modal distributions. The stages of this method is as follows. First the percentiles of each variable of seed cells are found and recorded; secondly the whole
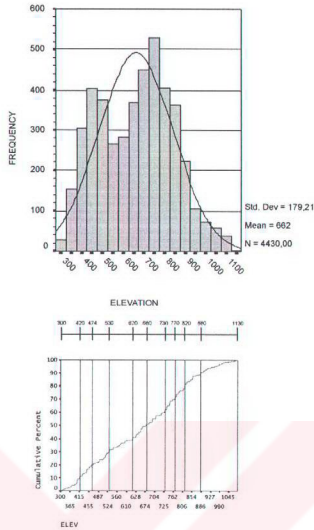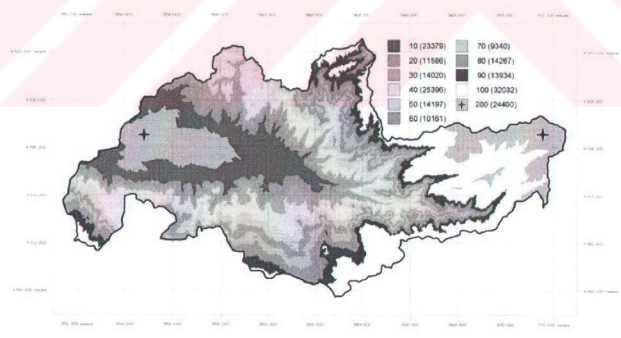
Figure 6.8. The snapshot of methodology of percentile method and reclassified parameter map production. The areas with a star are the nature's own decision rule and not taken into calculations as there is no landslide in these area.

parameter map is classified accordingly to the seed cells percentile limits (Figure 6.8.). The resultant classes of the parameter map have same landslide density but the areas of the percentile classes on the whole parameter map are not equal to each other. This constitutes the natural weighting of each parameter class. Furthermore, the minimum and maximum values of the parameter map are dependent of the seed cells data base, such as if the whole area has elevations of the magnitude as 200 meters as the minimum and the minimum of the seed cells are 300 meter, the area lying in between 200-300 meters are not taken into consideration as the nature itself creates her first decision rule as not having any landslides in that range.

All of the 13 parameters are analyzed and out of 13 excluding the 2 categorical ones (geological map and the land cover map) the remaining 11 parameters are re-classified according to the seed cell percentile values. All of the percentile maps are shown in Figures 6.9, 6.10 and in 6.11 with their frequency and cumulative histograms overlaid with percentile ranges.

Table 6.5. The percentiles of seed cells within each variable.

		ASPECT	SLOPE	ELEV	D_DRAIN	D_E5	D_SETTLE	DENS_FAULT	D_FAY	D_RIDGE	D_PRO	DENSDR
N	Valid	4430	4430	4430	4430	4430	4430	4430	4430	4430	4430	4430
	Missing	0	0	0	0	0	0	0	0	0	0	0
Mean		168,37	22,41	661,84	118,30	1217,83	699,45	235,56	408,52	113,88	176,22	160,68
Median		180,00	22,00	680,00	108,00	1206,00	459,00	209,00	306,00	96,00	145,00	157,00
Mode		45	22	750	25	1	153	0	114	13	9	150
Std. Deviation		126,25	9,12	179,21	76,99	899,59	651,25	161,03	339,66	82,12	142,29	37,01
Variance		15939,83	83,13	32116,12	5927,73	809255,17	424130,59	25930,03	115368,76	6743,79	20245,18	1369,43
Range		359	50	830	397	3112	2625	705	1517	360	683	232
Minimum		-1	1	300	0	0	4	0	0	4	0	55
Maximum		358	51	1130	397	3112	2629	705	1517	364	683	287
Percentiles	10	9	11	420	25	75	105	12	73	14	17	119
	20	27	15	474	49	211	167	110	127	39	41	136
	30	45	17	530	68	391	230	157	180	52	70	144
	40	83	20	620	87	834	322	189	239	75	107	150
	50	180	22	680	108	1206	459	209	306	96	145	157
	60	225	25	730	130	1607	609	242	386	122	189	165
	70	284	28	770	155	1857	804	286	478	152	241	174
	80	315	30	820	183	2114	1397	357	675	191	302	187
	90	333	34	880	221	2427	1825	481	977	238	382	208
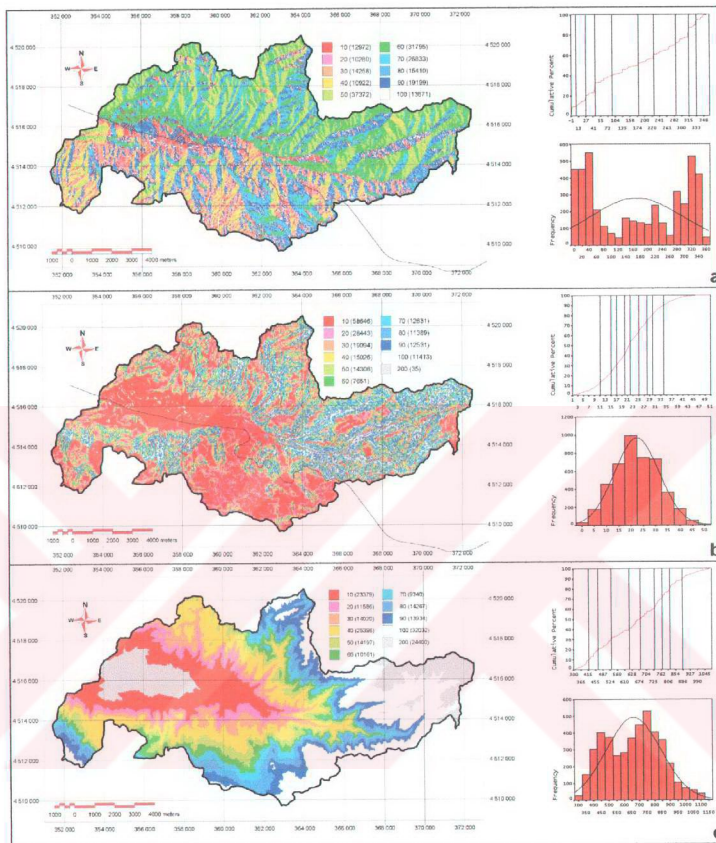
Figure 6.9. The percentile maps of morphology of Asarsuyu catchment, with frequency and cumulative histograms, a) Aspect, b) Slope, c) Elevation

144
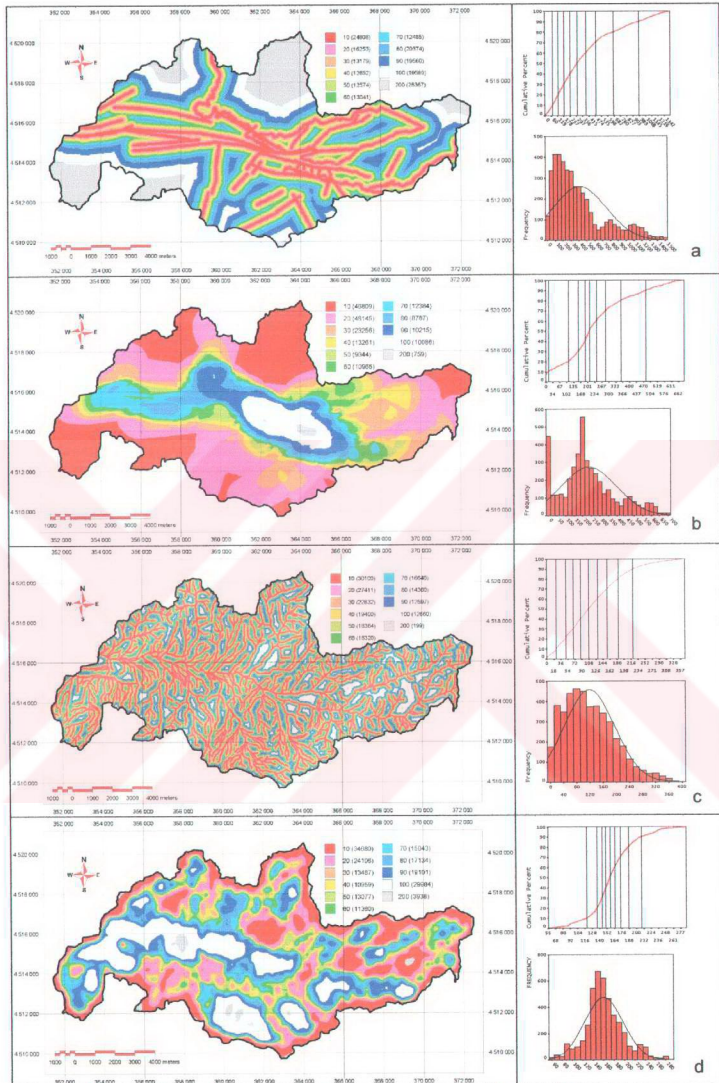
Figure 6.10. The percentile maps of lineament and density patterns of Asarsuyu catchment, with frequency and cumulative histograms, a) Distance to fault , b) Fault density, c) Distance to Drainage d) Drainage density
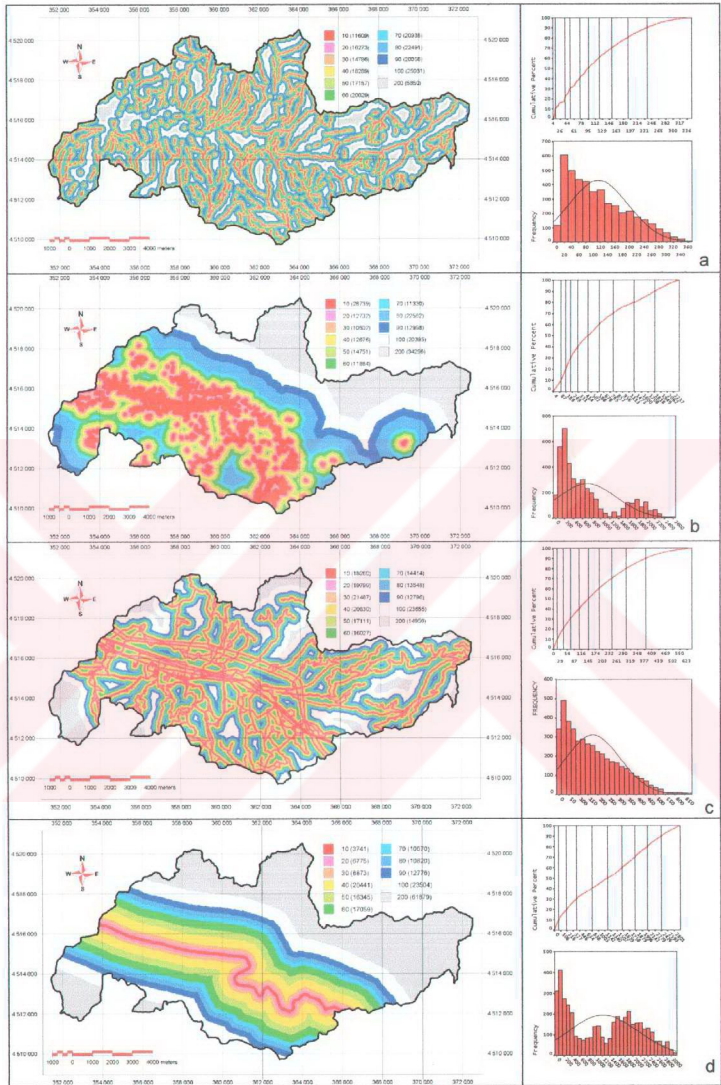
Figure 6.11. The percentile maps of infrastructure and distance to ridge of Asarsuyu catchment, with frequency and cumulative histograms, a) Distance to ridge , b) Distance to Settlement, c) Distance to Power and road network d) Distance to E-5 Highway.

6.4.1. Bi-variate analyses

In bi-variate analyses, as explained in Chapter 2, the core of the analysis is to get the densities of landslide occurrences in each parameter map and in each parameter map's classes, and to get some data driven weights based on the class density and the landslide density.

Two of the previously mentioned methods have been utilized in this study as: landslide susceptibility and information value method. In the landslide susceptibility method as the nodes of seed cells are representing 25x25 meter area in the map, the area density method is used. A brief recapitulation of the methods is given in Table 6.6.

Table 6.6. Methodological snapshot of used two methods

LANDSLIDE SUSCEPTIBILTY	INFORMATION VALUE
$$D_{area} = 1000 \frac{Npix(SX_i)}{Npix(X_i)}$$	$$I_i = \log \frac{S_i/N_i}{S/N}$$
in which D_{area} = Areal density per millage $Npix(SX_i)$ = number of pixels with mass movements within variable class X_i. $Npix(X_i)$ = number of pixels within variable class X_i. To evaluate the influence of each variable, weighting factors should have to be introduced, which compare the calculated density with the overall density in the area. The formula for the density-based area is: $$W_{area} = 1000 \frac{Npix(SX_i)}{Npix(X_i)} - 1000 \frac{\sum Npix(SX_i)}{\sum Npix(X_i)}$$	In which S_i: the number of land units or pixels with mass movements and the presence of variable X_i, N_i: The number of land units or pixels with variable X_i S: The total number of land units or pixels with mass movements N: The total number of land units or pixels. The degree of a hazard for a land unit or pixel j is calculated by the total information value I_j $$I_j = \sum_{i=0}^{m} X_{ij} I_j$$ in which m = number of variables, X_{ij} = 0 if the variable X_i is not present in the land unit or pixel j and 1 if the variable is present.

Simply, each area of class in the reclassified parameter maps (Figures 6.9, 6.10, 6.11) is divided by the landslide density in this class. As previously noted, the landslide densities are bound by the percentile method as approximately to %10 in each class. Hence the natural weights of each parameter class is solely dependent on its aerial coverage in the parameter map. Furthermore, it can be called as a weight standardization module as it is clear that weights of any arbitrary parameter classes are equal in concept but of course different in their values due to the areas of classes. However, they are still not comparable. In order to achieve the standardization of both the parameter classes in its parameter map and the natural weights of the parameter classes with respect to other classes, from each single parameter value the value of the sums of all of the weights of parameter classes in each map is subtracted. Hereforth, the weight of sixth percentile of slope class has a comparable weight with the weight of the third percentile of elevation as an example. On the other hand, a minor step to reduce the computational difficulties are made as the weight values of each class are normalized using the minimum weight in its parameter class. The example of this procedure is given in Figure 6.12 for both landslide susceptibility analysis and for information value method.

The steps shown in Figure 6.12 are carried out for all parameter maps of Asarsuyu catchment. The values of all parameter classes are shown in Table 6.7, which are then used to construct the hazard map.

The exploration of the weights yielded in expected values as well as some surprising weights, such as; in the fault density parameter the weights are decreasing when the density of fault lines decrease, of which it is expected that the rock units would stay intact enough in the absence of landslides. However, distance to fault has a surprising result as the first percentile (0 to 73 meter distance) gets 0 weight, which is probably due to the fact that some other factors in combination control the activity in those areas. Also the first percentile of distance to E-5 highway (0 to 75 meter distance) gets the maximum weight, which overrides some physical factors in the area. Another surprising fact was seen in the categorical parameters as the DSA class gets higher weight value than the flyschoidal sequence, which is due to the fact that they have the same amount of landslide but DSA is outcropping in a very restricted area.

Merely the same steps are encountered with the information value, however, after the analysis of the resultant weights, it was decided not to use as the method is based on log differences which do compress the mean values and speculating the extreme or outlier values. Although the weight values are different and more sensitive to the extremes, the signs of the values were similar to that of landslide susceptibility analyses, which is an affirmative condition for the validity of the method used.

Calculation of area coverages of parameter classes and respective seed cell count

Range of Parameter classes	Reclassified Value	Raster Cell Count	Parameter class Area in km2	Cumulative % of seed cells	% of seed cells
1-300	0	236630	148,0392	0	0
300-415	10	23379	14,6262	9,2	9,2
415,01-470	20	11586	7,2484	20	10,8
470,01-530	30	14020	8,7711	30,2	10,2
530,01-620	40	25396	15,8881	40,9	10,7
620,01-680	50	14197	8,8819	50,8	9,9
680,01-726	60	10161	6,3569	59,8	9
726,01-766	70	9340	5,8432	69,9	10,1
766,01-820	80	14267	8,9256	80,6	10,7
820,01-880	90	13934	8,7173	90,2	9,6
880,01-1130	100	32032	20,0397	100	9,8
1130,01-1580	200	24400	15,265	0	0
					100
	Total raster cells	192712	120,5634		
	used cells (0-100)	168312			

Calculation of landslide susceptibility index → Weight calculation of classes within the parameter map

Seed cell area / Parameter class area (SCA/PCA)	(SCA/PCA) * 1000
0,017433	17,43273878
0,041295	41,29466598
0,03223	32,2296719
0,018665	18,66475035
0,030892	30,89173769
0,039238	39,23826395
0,047905	47,90471092
0,033224	33,22422373
0,030521	30,5210277
0,013553	13,55332168
Total # of seed cells	4430
Total (SCA/PCA)*1000	26,32017

Parameter class weights with total density
-8,88743
14,9745
5,909505
-7,65542
4,57157
12,9181
21,58454
6,904056
4,20086
-12,7668

Normalization of weights

Parameter class weight normalization with minimum value
3,883
27,744
18,860
5,115
17,342
25,668
34,355
19,674
16,971
0

BEFORE NORMALIZATION

AFTER NORMALIZATION

Figure 6.12 The steps through landslide susceptibility analysis

Table 6.7. Weight values of the all available parameter classes

Parameter map	Parameter Classes									
	10	20	30	40	50	60	70	80	90	100
fault density	0,26	0,00	9,65	23,87	37,26	32,81	26,93	41,21	33,73	34,72
elevation	3,88	27,74	18,68	5,11	17,34	25,69	34,35	19,67	16,97	0,00
d to ridge	0,00	2,37	3,42	4,79	4,82	7,52	7,67	9,14	11,06	22,16
d to settle	0,00	19,38	27,19	19,18	14,61	21,92	23,68	4,20	18,66	6,30
d to pr	5,53	3,65	2,10	2,54	6,90	9,19	11,70	14,30	15,55	0,00
d to fault	0,00	9,67	15,76	17,16	17,02	16,11	17,62	3,89	4,79	4,76
d to e5	99,56	46,53	31,07	2,82	8,25	7,11	22,66	22,09	15,82	-0,01
d to drain	0,00	0,35	4,44	7,67	8,72	9,50	11,46	16,00	19,30	19,83
drainage dens	-0,01	5,89	20,14	26,17	22,53	22,06	18,18	12,04	10,15	1,59
slope	0,00	8,28	18,15	24,11	29,40	29,92	30,04	32,92	28,57	30,51
aspect	23,33	32,02	26,71	18,51	-0,01	3,68	3,52	11,20	15,72	22,84

	quat	BGY	DSA	tkbf	talus	bolugran	Eocay	sok	tkb	pliomen	asarsuyu	gypsum
geology	0,67	2,08	58,04	35,95	3,69	28,73	22,68	19,60	0,00	0,00	0,00	0,00

	dense forest	forest	soil moisture	mixed	settlement	road
landcover	4,69	29,03	0,00	14,43	0,16	124,04

After calculation of the weights, the weigh values are assigned to the initial parameter maps. Following this all of the 13 parameters maps are added up to create the hazard map. No extra weighting procedure for the parameter maps are used in the summation process as the classes have been normalized and they received their natural weightings from the data itself.

The resultant hazard map is then reclassified into 4 hazard zones (very low, low, high, very high) using the hazard maps distribution parameters. The mean value of the hazard map is taken as the pivot point and classes are assigned to the + and - one standard deviations of the distribution. The resultant map and the landslide amounts in these hazard classes are given in Figure 6.13. It is seen that 48 % of the total area is classified as high and high hazard class and within these classes 93.3 % of seed cells are encountered. It should also be noted that the two giant landslide bodies (Bakacak and Bülbülderesi slides) are not taken into consideration at this stage, in order to show the real distribution of the hazard classes and to see if any information could be obtained from form the analyses within these landslide bodies. On the other hand 52% of the study area is classified as low and very low hazard, which in turn hosts only 6.7 % of total seed cells in the area. This distribution also validates that the classification is quite reasonable.
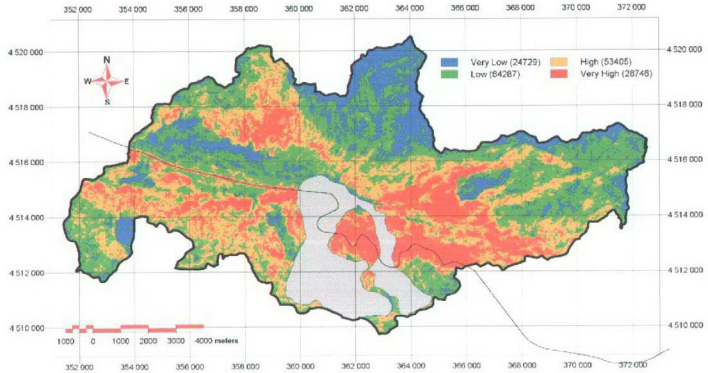
DN Value	Hazard class	% area covered	% landslide
0-93	Very low	14,45	0,5
93-146	low	37,56	6,2
146-199	High	31,20	26,7
199-481	Very High	16,79	66,6

Figure 6.13. The hazard map and the amounts of landslides in each class as a result of bivariate analysis. (red is very high hazard, orange is high hazard, green is low hazard and blue is very low hazard). The grey polygon in the figure is the huge landslide body.

6.4.2. Multivariate analyses

The multivariate statistical analyses of the important causal factors for landslide occurrence may indicate the relative contribution of each of these factors to the degree of hazard within a defined land unit. The analyses are based on the presence or absence of stability phenomena within these units (van Westen, 1993). Two major trends are recorded in the literature as the standard multiple regression and discriminant analyses. However, many of the authors who use these methods do override the necessity of the data to be normally distributed. It could have been said of an assumption, though this assumption is a vital one controlling the applicability of these methods. Instead of the common literature two other multivariate techniques are utilized in this study as factor analysis to understand better the interrelations of the parameter maps, and the logical regression analysis to figure out a multivariate dominated hazard map.

151

6.4.2.1. Factor analysis

Factor analysis is similar to principal components analysis in that it is a technique for examining the interrelationships among a set of variables (Afifi and Clark, 1998). Both of these techniques differ form multiple regression analyses, as there does not need to be a dependent and a series of independent variables. The factors obtained in a factor analysis are selected mainly to explain the interrelationships among the original variables. The major emphasis is placed on obtaining easily understandable factors that convey the essential information contained in the original set of variables.

In the initial stage all of the 13 parameters of the seed cells are included in the factor analysis, and principal axis factoring method with varimax rotation is selected as the factor analysis method. A number of tests should have to be performed for the validity of factor analysis with the given variables so a KMO-Bartlett test is conducted (Table 6.8). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy is a statistic which indicates the proportion of variance in the variables which is common variance, i.e. which might be caused by underlying factors. High values (close to 1.0) generally indicate that a factor analysis may be useful with the available data. If the value is less than 0.50, the results of the factor analysis probably won't be very useful. In this case it is nearly just over the limit as the value is 0.593. Bartlett's test of sphericity indicates whether the correlation matrix is an identity matrix, which would indicate that the variables are unrelated. The significance level gives the result of the test. Very small values (less than .05) indicate that there are probably significant relationships among the variables. A value higher than about .10 or so may indicate that the data are not suitable for factor analysis. Based on these critical values, factor analysis seems not to yield a very successful result.

Table 6.8. KMO and Bartlett's test with initial 13 variables.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,593
Bartlett's Test of Sphericity	Approx. Chi-Square	13926,75
	df	78
	Sig.	,000

The next step in the analyses is to figure out which variables are not fitting to the model, which is done by exploring the anti-image matrices. The anti-image matrices contain the negative partial covariances and correlations. They can give an indication of correlations that aren't due to the common factors. Small values indicate that the

variables are relatively free of unexplained correlations. Most or all values off the diagonal should be small (close to zero). Each value on the diagonal of the anti-image correlation matrix shows the Measure of Sampling Adequacy (MSA) for the respective item. Values less than 0.5 may indicate variables that do not seem to fit with the structure of the other variables. The anti image matrix of the initial 13 variables are presented in Table 6.9, and it is seen that distance to drainage and distance to ridge are the two variables that do not fit into the structure of the remaining variables. As this is the first iterative pass of the system it was decided to include these parameters in order to see the effects of these variables after and before their removal.

Table 6.9. The Anti-Image matrices of initial 13 variables.

Anti-image Matrices

		ASPECT	SLOPE	ELEV	landcover	D_DRAIN	DISTE5	DSETTLE	DEN_FAU	D_FAY	D_RIDGE	D_PRO	DEN_DR	Geocode
Anti-image Covariance	ASPECT	.914	-4.622E-02	4.994E-02	-9.060E-02	-8.535E-02	-7.857E-02	-7.283E-03	2.491E-02	3.996E-02	5.290E-02	3.701E-02	9.134E-02	-7.265E-04
	SLOPE	-4.622E-02	.819	2.417E-02	1.966E-02	-1.785E-02	2.930E-02	-7.419E-04	-2.450E-02	7.915E-02	-.247	-5.592E-02	-3.079E-02	-.127
	ELEV	4.994E-02	2.417E-02	.713	1.565E-02	-.117	.140	-.236	2.330E-02	7.471E-03	-6.583E-02	1.806E-02	-2.870E-02	-3.853E-02
	landcover	-9.060E-02	1.966E-02	1.565E-02	.832	-3.500E-02	.124	.108	-2.1706E-02	2.745E-02	-3.650E-02	.153	1.891E-02	-6.119E-02
	D_DRAIN	-8.535E-02	-1.785E-02	-.117	-3.500E-02	.788	-3.453E-02	8.459E-02	-4.981E-02	3.904E-02	.158	-7.490E-02	.263	-9.707E-02
	DISTE5	-7.857E-02	2.930E-02	.140	.124	-3.453E-02	.423	-.228	9.900E-02	-6.863E-02	1.803E-02	-4.477E-03	3.776E-02	-5.863E-02
	DSETTLE	-7.283E-03	-7.419E-04	-.236	.108	8.459E-02	-.228	.372	-.100	6.159E-02	7.264E-03	-3.253E-02	7.537E-02	-3.981E-02
	DENS_FAULT	2.491E-02	-2.450E-02	2.330E-02	-2.1706E-02	-4.981E-02	9.900E-02	-.100	.411	.237	8.426E-02	-5.751E-02	1.746E-02	.146
	D_FAULT	3.996E-02	7.915E-02	7.471E-03	2.745E-02	3.904E-02	-6.863E-02	6.159E-02	.237	.393	4.922E-02	-.128	.152	3.576E-02
	D_RIDGE	5.290E-02	-.247	-6.58E-02	-3.650E-02	.158	1.803E-02	7.264E-03	8.426E-02	4.922E-02	.829	3.522E-02	8.372E-02	3.546E-03
	D_PRO	3.701E-02	-5.592E-02	1.806E-02	.153	-7.490E-02	-4.477E-03	-3.253E-02	-5.751E-02	-.128	3.522E-02	.819	-1.981E-02	-6.681E-02
	DENS_DRAIN	9.134E-02	-3.079E-02	-2.870E-02	1.891E-02	.263	3.776E-02	7.537E-02	1.746E-02	.152	8.372E-02	-1.981E-02	.707	-9.759E-02
	Geocode	-7.265E-04	-.127	-3.85E-02	-6.119E-02	-9.707E-02	-5.863E-02	-3.981E-02	.146	3.576E-02	3.546E-03	-6.681E-02	-9.759E-02	.847
Anti-image Correlation	ASPECT	.545[a]	-5.340E-02	6.184E-02	-.119	-.101	-.126	-1.249E-02	4.064E-02	6.669E-02	6.075E-02	4.275E-02	.114	-8.254E-04
	SLOPE	-5.340E-02	.611[a]	3.163E-02	2.733E-02	-2.222E-02	4.980E-02	-1.345E-03	-4.324E-02	.140	-.300	-8.825E-02	-4.047E-02	-.153
	ELEV	6.184E-02	3.163E-02	.485[a]	2.332E-02	-.156	.255	-.459	4.305E-02	1.412E-02	-8.561E-02	2.362E-02	-4.043E-02	-4.957E-02
	landcover	-.119	2.733E-02	2.332E-02	.786[a]	-4.961E-02	.240	.222	-4.260E-02	5.511E-02	-5.043E-02	.213	2.829E-02	-6.997E-02
	D_DRAIN	-.101	-2.222E-02	-.156	-4.961E-02	.433[a]	-5.984E-02	.156	-8.757E-02	7.020E-02	.196	-9.325E-02	.352	-6.567E-02
	DISTE5	-.126	4.980E-02	.255	.240	-5.984E-02	.630[a]	-.570	.230	-.114	3.047E-02	-7.609E-03	5.080E-02	-9.800E-02
	DSETTLE	-1.249E-02	-1.345E-03	-.459	.222	.156	-.570	.540[a]	-.256	.151	1.309E-02	-5.895E-02	.147	-7.096E-02
	DENS_FAULT	4.064E-02	-4.324E-02	4.305E-02	-4.260E-02	-8.757E-02	.230	-.256	.579[a]	.680	.144	-9.913E-02	3.241E-02	.247
	D_FAULT	6.669E-02	.140	1.412E-02	5.511E-02	7.020E-02	-.114	.151	.680	.618[a]	8.826E-02	-.225	.289	6.201E-02
	D_RIDGE	6.075E-02	-.300	-8.56E-02	-5.043E-02	.196	3.047E-02	1.309E-02	.144	8.826E-02	.480[a]	4.273E-02	.109	4.231E-03
	D_PRO	4.275E-02	-8.825E-02	2.362E-02	.213	-9.325E-02	-7.609E-03	-5.895E-02	-9.913E-02	-.225	4.273E-02	.696[a]	-2.804E-02	-8.019E-02
	DENS_DRAIN	.114	-4.047E-02	-4.04E-02	2.829E-02	.352	5.080E-02	.147	3.241E-02	.289	.109	-2.804E-02	.589[a]	-.126
	Geocode	-8.254E-04	-.153	-4.95E-02	-6.997E-02	-6.567E-02	-9.800E-02	-7.096E-02	.247	6.201E-02	4.231E-03	-8.019E-02	-.126	.546[a]

a. Measures of Sampling Adequacy(MSA)

The amount of the total variance explained with the initial 13 variables are shown in Table 6.10. This table gives the amount of cumulative variance explained with the initial solution and initial rotation of the factor analyses. As in the first steps of factor analyses the decision rule for acceptance of the new factors are defined as "if a new factor is to be created it should have an effect of at least equal to that of an initial variable". Dependent on this rule eigenvalues of smaller than 1 are not included in the factor analysis, as a result, with the initial 13 variables only the first 4 factors are taken into consideration and they could only explain the 57% of the total variance observed (Figure 6.14).
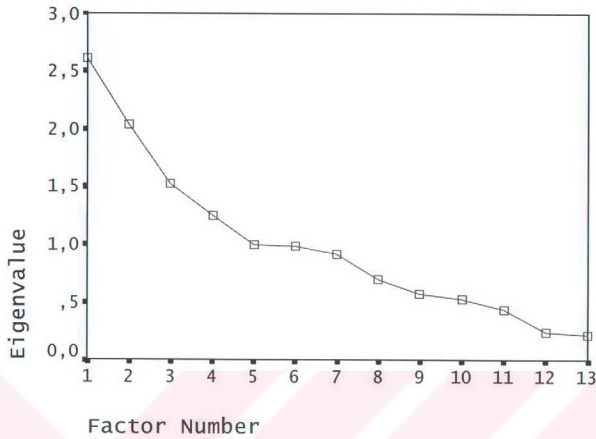
Figure 6.14. Eigenvalues of the factors.

Table 6.10. The amount of total variance explained via factors

Total Variance Explained

		Initial Eigenvalues			Extraction Sums of Squared Loadings			tation Sums of Squared Loadin		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Factor	1	2,616	20,121	20,121	2,176	16,738	16,738	1,985	15,269	15,269
	2	2,035	15,652	35,774	1,713	13,175	29,913	1,806	13,892	29,161
	3	1,518	11,679	47,453	,895	6,883	36,797	,850	6,536	35,697
	4	1,247	9,590	57,043	,563	4,334	41,130	,706	5,433	41,130
	5	,992	7,633	64,676						
	6	,987	7,590	72,266						
	7	,918	7,060	79,326						
	8	,694	5,337	84,663						
	9	,576	4,432	89,095						
	10	,523	4,025	93,119						
	11	,439	3,378	96,497						
	12	,242	1,863	98,360						
	13	,213	1,640	100,000						

Extraction Method: Principal Axis Factoring.

The next step is to figure out what variables are responsible for the selected 4 factors, this is done by analyzing the factor matrix after varimax rotation. The Table 6.11 reports the factor loadings for each variable on the components or factors after rotation.

Each number represents the partial correlation between the item and the rotated factor. The bold numbers represent the maximum correlations within these factors. Based on these maximum correlations, it could be concluded that the first factor is composed of human activities (Distance to settlement, Distance to E-5 highway, Distance to power lines and road network and the land cover), the second one is attributed to primarily to lineament pattern and minor to settlement and elevation, the third factors constituents are the drainage system and its resultant morphodynamic attributes (distance to drainage, drainage density and aspect), and the last factor the fourth one is attributed to the material properties and drainage dependent morphodynamic items (Slope, Distance to ridge and geology).

Upon the completion of this initial iterative pass, a step by step variable removal scheme is applied based on the rules of Anti-Image Matrices. The best solution is found after the second pass with the removal of distance to drainage and distance to ridge parameters. After the removal of these two variables the Kaiser-Meyer-Olkin Measure of Sampling Adequacy is increased from 0.0593 to 0,618, which is seen in Table 6.12.

Table 6.11. The rotated factor matrix and the variable loadings

Rotated Factor Matrix [a]

	Factor			
	1	**2**	**3**	**4**
D_SETTLE	**,863**	,387	4,536E-02	5,018E-02
D_E5	**,726**	-,220	,130	-2,1E-02
landcover	**-,635**	7,658E-02	5,137E-02	6,173E-02
D_PRO	**,322**	-,143	8,227E-02	-3,8E-02
DENS_FAULT	-,123	**,834**	-1,0E-02	-,114
D_FAY	,182	**-,825**	,126	-,140
ELEV	,274	**,302**	4,949E-02	9,436E-02
D_DRAIN	-3,7E-02	9,862E-02	**,676**	-,114
DENS_DRAIN	-,190	,203	**-,505**	9,406E-02
ASPECT	1,652E-02	-1,0E-02	**,273**	1,308E-02
SLOPE	-6,8E-02	,199	-4,5E-03	**,579**
D_RIDGE	-7,0E-02	-2,0E-02	-,125	**,481**
Geocode	,187	-,150	1,867E-02	**,262**

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Table 6.12. KMO and Bartlett's test after removal of two variables.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,618
Bartlett's Test of Sphericity	Approx. Chi-Square	12215,87
	df	55
	Sig.	,000

Furthermore, the amount of the total variance explained with the variables are also increased from 57 % to 63 %, of which is shown in Table 6.13. Although 63% is still very low for such kind of analyses, further removal of the variables would probably yield in the degradation of the model success, hence the removal scheme is ended up with 11 variables.

Table 6.13. The amount of total variance explained via factors, after removal of two variables.

Total Variance Explained

		Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Factor	1	2,590	23,543	23,543	2,180	19,817	19,817	1,926	17,509	17,509
	2	2,033	18,485	42,028	1,711	15,557	35,374	1,865	16,953	34,462
	3	1,212	11,014	53,043	,638	5,800	41,174	,657	5,977	40,439
	4	1,125	10,225	63,267	,424	3,857	45,031	,505	4,592	45,031
	5	,921	8,373	71,640						
	6	,831	7,551	79,190						
	7	,740	6,727	85,918						
	8	,630	5,729	91,647						
	9	,449	4,080	95,727						
	10	,249	2,263	97,990						
	11	,221	2,010	100,000						

Extraction Method: Principal Axis Factoring.

The final factor loadings are seen in Table 6.14. Due to the different variables encountered into the analyses, the rotation scheme and the factor loading scheme is slightly changed, resulting in a more stable model. The stability of the model is also seen in the generic differentiation of the factors and their responsible variables. Such as the elevation parameter has promoted to the first factor which is more meaningful, as the second factor is solely remained for the effects of lineaments. It is obvious that the presence of lineaments are not controlled by elevation, on the other hand the presence of settlement, position of the highway, land cover and the power lines and the road network are dependent of elevation. Consequently, the first factor is dominated by human activities. As noted above the second factor is solely dependent on lineament pattern and density in the area. In the third factor geological units and slope is promoted

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

from 4th factor and drainage components and aspect fall into 4th factor, which in turn more logical as the author believes that the material properties control the slope and also the presence of landslides; hence third factor is attributed to material properties. The remaining two variables aspect and the density of drainage is responsible from the fourth factor and could be considered as the morphodynamical factor of the Asarsuyu catchment.

Table 6.14. The rotated factor matrix and the variable loadings, after removal of two variables

Rotated Factor Matrix [a]

	Factor			
	1	**2**	**3**	**4**
D_SETTLE	,870	,322	2,762E-02	,198
D_E5	,669	-,272	,103	,293
landcover	-,658	,146	1,409E-02	9,911E-02
D_PRO	,338	-,200	3,538E-02	-9,5E-02
ELEV	,316	,274	3,784E-02	-4,1E-02
D_FAY	,121	**-,909**	-4,9E-02	3,709E-02
DENS_FAULT	-2,0E-02	**,758**	-,218	-6,0E-02
SLOPE	-5,4E-02	,249	**,225**	-3,9E-02
Geocode	,140	-,112	**,714**	2,305E-02
ASPECT	-4,3E-02	-2,7E-03	3,079E-02	**,500**
DENS_DRAIN	-,187	,274	,177	**-,321**

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

6.4.2.2. Logical Regression

Logical regression allows forming a multivariate regression relation between a dependent variable and several independent variables which might affect the probability of the searched situation. If the searched variable is a dichotomous outcome the best method with free of predictor variable type is seem to be logistic regression (Afifi and Clark, 1998; Atkinson and Massari, 1998; Dai et al. 2001; Lee and Min, 2001).

Binomial (or binary) logistic regression is a form of regression which is used when the dependent is a dichotomy and the independents are continuous variables, categorical variables, or both. Multinomial logistic regression exists to handle the case of dependents with more classes. Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the dependent occurring or not). In this way, logistic regression estimates the probability of a certain event occurring. Note that logistic regression calculates changes

in the log odds of the dependent, not changes in the dependent itself as Ordinary Least Squares (OLS) regression does.

The logistic model can be written in its simplest form as:

$$P = \frac{1}{1+e^{-z}}$$

where P is the probability of an event occurring. P is the estimated probability of landslide occurrence. As z varies from $-\infty$ to $+\infty$, the probability varies from 0 to 1 on an s shaped curve. And where z is defined as:

$$Z=B_0+B_1X_1+B_2X_2+B_3X_3+\ldots\ldots\ldots+B_nX_n$$

Where B_0 is the intercept of the model, n is the number of independent variables, $\ldots B_i$ (I=1,2,3,.....,n) is the slope coefficient of the model and X_i(I=1,2,3,.....,n) is the independent variable.

In an extended form the equation of logistic regression could be written as:

Probability of belonging to population I (logit) = $\dfrac{1}{1+e^{B0+B1X1+B2X2+B3X3+\ldots\ldots\ldots+BnXn}}$

The advantage of the logistic regression over simple multiple regression is that, through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or categorical, or any combination of both types. In general the advantage of logistic regression modeling over the other multivariate statistical techniques, including multiple regression analysis and discriminant analyses, is that the dependent variable can have only two values – a dichotomous outcome -, and that predicted values can be interpreted as probability because they are constrained to fall into an interval between 0 and 1 (Kleinbaum, 1991). Logistic regression has many analogies to Ordinary Least Squares (OLS) regression: logit coefficients correspond to B coefficients in the logistic regression equation, the standardized logit coefficients correspond to beta weights, and a pseudo R^2 statistic is available to summarize the strength of the relationship. Unlike OLS regression, however, logistic regression does not assume linearity of relationship between the independent variables and the dependent, does not require normally distributed variables, does not assume homoscedasticity, and in general has less stringent requirements. The success of the logistic regression can be assessed by looking at the classification table, showing

correct and incorrect classifications of the dichotomous, ordinal, or polytomous dependent. (Afifi and Clark, 1998; Wrigley, 1984)

However, in a strict sense, it is not a probability because the extrinsic parameters triggering the landslides such as the rainfall and earthquake vibration are not accounted for. It might be appropriate to term it as landslide susceptibility based on the intrinsic physical parameters.

In order to carry out the logical regression the total number of seed cells (4430) are used. Moreover, 4430 random sample nodes are selected from the landslide free areas of Asarsuyu catchment, that are presented in Figure 6.15. Upon the selection of these random nodes, the values of the parameter maps are then transferred to the database of the random data set. Following the creation of the random data set database, the seed cells and the random set database is merged and a new column of a binary variable indicating the presence and absence of the landslides are added. This stage is repeated 4 times in order to have 4 different sets of random points, which in turn would let the user to see if there is any convergence in the success of logistic regression analyses.

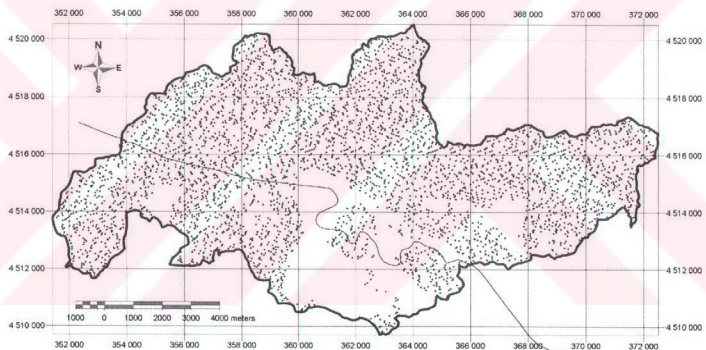**Figure 6.15.** The positions of selected 4430 random landslide free nodes.

The initial assumption of the variables by the logistic regression is shown in Table 6.15.

The system test reveals that the variables and the system constructed are valid. A Hosmer-Lemeshow test and Cox & Snell R square and Nagelkerke R square values are obtained and the statistical package supports that the system is still valid with these variables.

Table 6.15. The initial assumption of the variables.

Classification Table [a,b]

				Predicted		
				LANDBIN		Percentage Correct
				0	1	
Step 0	Observed	LANDBIN	0	0	4430	,0
			1	0	4430	100,0
		Overall Percentage				50,0

a. Constant is included in the model.
b. The cut value is ,500

After the validation of the system, logistic regression is applied to the data set and the resultant classification table is presented in Table 6.16. The system in overall has a success of classifying 77.3% of the pixels correctly, which is quite acceptable. The 77.3% is the highest classification success among the 4 different random sets, though the remaining three were oscillating around 75% with plus minus 1 %.

Table 6.16. The final classification of logistic regression

Classification Table [a]

				Predicted		
				LANDBIN		Percentage Correct
				0	1	
Step 1	Observed	LANDBIN	0	3387	1019	76,9
			1	985	3445	77,8
		Overall Percentage				77,3

a. The cut value is ,500

The loadings of the variables after logistic regression is presented in Table 6.17. Based on these values the logistic regression equation is compiled as follows:

Z= 0,773046364 + (0,130082590 * *GEOCODE*) – (0,004154044 * *DRAINAGE DENSITY*) – (0,000897442 * *DISTANCE TO POWERLINES AND ROAD NETWORK*) – (0,004813297 * *DISTANCE TO RIDGE*) + 0,000212306 * *DISTANCE TO FAULT*) – (0,000525944 * *DENSITY OF FAULT*) – (0,00114028 + *DISTANCE TO SETTLEMENT*) – (0,001257937 * *DISTANCE TO E-5 HIGHWAY*) + (0,001047155 * *DISTANCE TO DRAINAGE*) – (0,120726833 * *LANDCOVER*) + (0,002760724 * *ELEVATION*) + (0,52982916 * *SLOPE*) – (0,00296956 * *ASPECT*)

Table 6.17. The variables and their loadings after logistic regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B) Lower	Upper
Step 1[a]	ASPECT	-,000296956	,000	1,540	1	,215	1,000	,999	1,000
	SLOPE	,052982916	,003	318,496	1	,000	1,054	1,048	1,061
	ELEVATION	,002760724	,000	232,298	1	,000	1,003	1,002	1,003
	LANDCOVER	-,120726833	,024	25,276	1	,000	,886	,846	,929
	D_DRAIN	,001047155	,000	6,883	1	,009	1,001	1,000	1,002
	D_E5	-,001257937	,000	932,863	1	,000	,999	,999	,999
	D_SETTLE	-,000114028	,000	4,944	1	,026	1,000	1,000	1,000
	DENS_FAULT	-,000525944	,000	5,854	1	,016	,999	,999	1,000
	D_FAY	,000212306	,000	7,705	1	,006	1,000	1,000	1,000
	D_RIDGE	-,004813297	,000	210,251	1	,000	,995	,995	,996
	D_PRO	-,000897442	,000	33,496	1	,000	,999	,999	,999
	DENS_DRAIN	-,004154044	,001	31,743	1	,000	,996	,994	,997
	GEOCODE	,130082590	,017	59,496	1	,000	1,139	1,102	1,177
	Constant	,773046364	,229	11,390	1	,001	2,166		

a. Variable(s) entered on step 1: ASPECT, SLOPE, ELEV, LANDCOVE, DISTDERE, DISTE5, DSETTLE, FAYDENS, DFAY, DTEPE, DPRO, DENSDR, JEOCODE.

The observed groups and the predicted probabilities of these groups are presented in Figure 6.16. It is clearly seen that in the values larger than the cutoff value, are under dominance of binary variable 1 which is landsliding, the opposite of this argument is valid also for the values lower than cut off value as safe pixels. In the cut off value it is seen that both probabilities are nearly same, with a little emphasis on landslides side. Furthermore, the values smaller than 0.25 still have some landslide pixels which support the hazard classification scheme as very low to low hazard classes in lower values than cut off value.
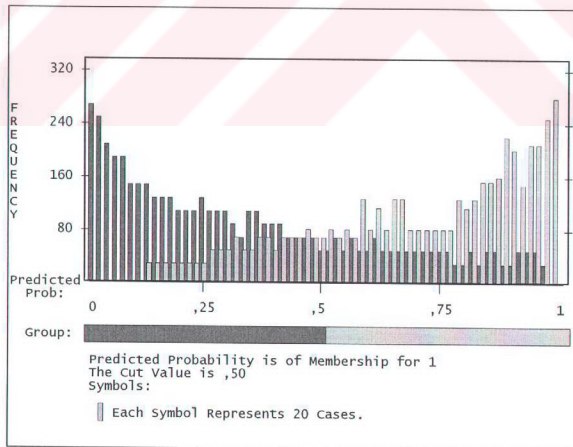

Figure 6.16. Observed groups and predicted probabilities

The logit of this z function is calculated for all of the pixels of Asarsuyu Catchment. The end members of the classification scheme is fixed as 0 being the no hazard class and 1 as the total hazard class. The foundation of the classification is based on these end members, hence the class boundaries are as follows: 0-0.25 very low hazard, 0.25-0.5 low hazard, 0.5-0.75 high hazard and 0.75-1 very high hazard. This re-classification of the hazard map is shown as the landslide hazard map of Asarsuyu catchment (Figure 6.17.)

The hazard map produced from logical regression results in more homogenous zones than that of bivariate analyses (Figure 6.13), especially in the end members of the zonations; in very low and very high hazard classes. The low and very low hazard classes constitute 72.31% of the area with corresponding 22.23% of the total landslide seed cells. On the other hand, the rest of the area is classified as high and very high hazard that yield in 27.68% of the area with corresponding 77.77% of the total landslide seed cells. Based on these numbers, the multivariate analysis results gives out a more comprehensive hazard classification in which the details of the comparisons of the produced two hazard maps will be further investigated in the next section.

It should also be noted that the grey polygons in the both produced hazard maps do belong to the two large landslide bodies. As previously mentioned, the boundaries of landslide polygons are digitized from the aerial photographs of 4 different periods. The union of these 4 polygons represents the final landslide polygon. The hazard class of this huge polygon is assigned as very high hazard based on field information and the aerial photographical interpretation. It was obvious from the photographs that these giant landslide polygons are formed by the aggradations of smaller landslides. The evidence was a relict landslide scarp at which the slided body of this relict scarp is acting as the host of newer landslides. Also the current slope morphology of these polygons suggests that there should be 4 or 5 different landslide associations. Furthermore, the work carried out by Işın (1999) in the Bolu mountain proved the slope movement by insitu monitoring. Based on these conditions it is concluded that these landslide polygons will behave in residual shear strength conditions, rather than peak conditions as that of the seed cells, consequently they will be put in to very high hazard class in the final hazard map.
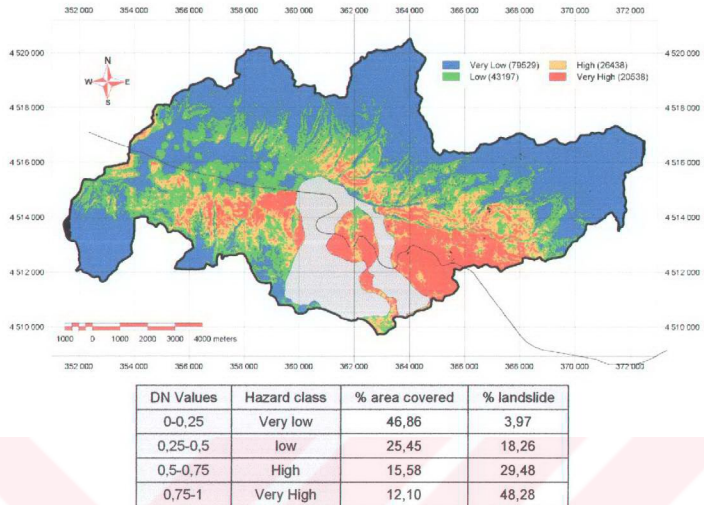
DN Values	Hazard class	% area covered	% landslide
0-0,25	Very low	46,86	3,97
0,25-0,5	low	25,45	18,26
0,5-0,75	High	15,58	29,48
0,75-1	Very High	12,10	48,28

Figure 6.17. The hazard map and the amounts of landslides in each class as a result of multivariate analysis. (red is very high hazard, orange is high hazard, green is low hazard and blue is very low hazard). The grey polygon in the figure is the giant landslide body

6.4.3. Comparison of two produced hazard maps

Two hazard maps are produced from bivariate analysis and logical regression analyses. Both of them produced acceptable results, as both of them classify the majority of the seed cells in high or very high hazard classes. However, they have to be analyzed in order to reveal which method is more successful and which method is more accurate. Therefore, two comparison schemes are developed and presented in the following sections.

6.4.3.1. The comparison of methods via their areas and corresponding landslide seed cells.

It has been shown that both methods classify less than half of the study area as high or very high hazard in conjunction more than two thirds of the seed cells in these areas, which could be said of a success. However, when the class areas are normalized with the landslide seed cell counts some important issues have aroused. In order to normalize the areas, the area percent values are divided with the landslide seed cell percent values, that is called as the seed cell area index (SCAI) density of landslides among the classes, which is presented in Table 6.18.

163

Table 6.18 The densities of landslides among hazard classes of both methods.

		Area %	Seed %	SCAI
Bivariate	Very Low	14,45	0,5	28,8945
	Low	37,56	6,4	5,8684
	High	31,20	26,8	1,1642
	Very High	16,79	66,8	0,2514
Logical	Very Low	46,86	3,97	11,79
	Low	25,45	18,26	1,3939
	High	15,58	29,48	0,5284
	Very High	12,10	48,28	0,2506

The logic behind SCAI lies in correct classification of seed cells within very conservative areal extent. As a result, it is desired that the high and very high hazard classes should have very small SCAI values and low and very low hazard classes to have higher SCAI values.

When the SCAI values of the two methods are compared it is found that the hazard map (LHM) produced from logical regression analyses has lower SCAI values than that of bivariate version. Only in very low hazard class bivariate hazard map (BHM) has a better result. In low and high hazard classes LHM has a clear superiority, however in very high hazard class they are quite close to each other but the logical one has a slight advantage as less area are classified as very high hazard and also its SCAI is slightly lower. Although the BHP has a high SCAI value which is desirable for the very low hazard class, the area classified is only 14.45 % of the total area which is very low for settlement planning purposes. The system should be a little more flexible rather than a mechanical rigid system, as considering the acceptable risk of the dwellers in the area. The 3.44% increase of landslide seed cells in the very low hazard class in the LHM ended up in 34.28 % of extra area with minimal hazard, which could be accepted.

6.4.3.2. The comparison of two methods in the spatial domain

The SCAI in the previous section does not reveal any information about the change of hazard score within a pixel. In order to achieve the pixel basis changes or mismatches both of the hazard maps are first re-classified into known numerical values. The bivariate hazard map is classified as 1,2,3,4 starting from very low hazard and ending up with very high hazard, correspondingly. The logical regression hazard map is classified by the 10 times multiplication of new class numbers of the bivariate hazard map, accordingly the class values are 10,20,30,40. After this re-coding process the two maps are added up. The available outcomes are presented in Table 6.19.

Table 6.19. The available combinations of re-coding process and their meanings

		L	O	G	I	C	A	L
BIVARIATE	0	10	20	30	40
	1	**11**	21	31	41
	2	12	**22**	32	42
	3	13	23	**33**	43
	4	14	24	34	**44**

Misclassification
Correct Classification
Acceptable Classification
Not Acceptable

As can be seen in Table 6.19, some combinations result in misclassified pixels. These are dominated by the absence of hazard score in both or in one of the hazard maps. The amount of these pixels compared to the whole classified area is merely 1.12 %, which is negligible, and shown in Figure 6.18 as the summation of 0,1,2,3,4,10,20,30 and 40 class id's. Although it is negligible, the spatial locations of these error prone pixels should be investigated (Figure 6.19). Generally they fall into the borders of the hazard map, which indicate a small acceptable resampling error.
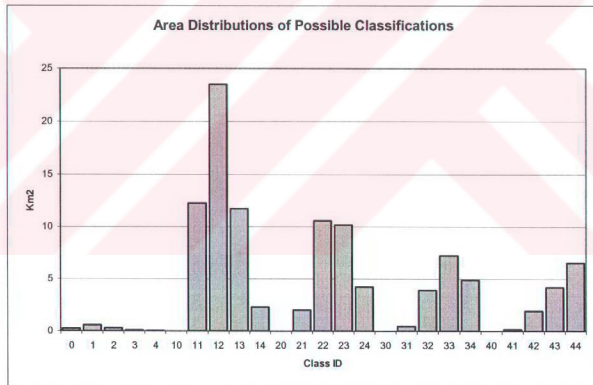

Figure 6.18. The areal distributions of classified pixels.

165
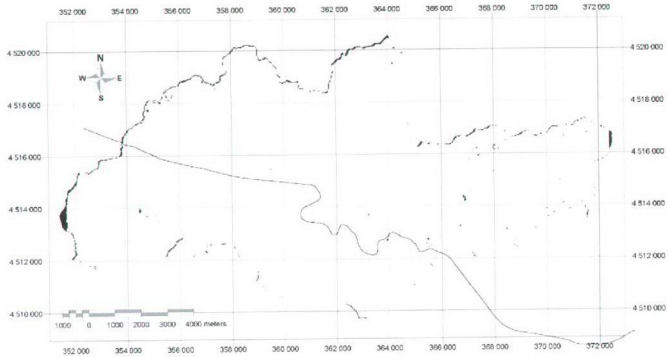
Figure 6.19. The locations of the misclassified pixels

If the two hazard maps converge in the same hazard classes after summation, such as they possess 11, 22,33 and 44 hazard ID's they are called as correctly classified pixels. The areal extent of these pixels are constituting 36,67 km^2 and 34.16 % of the total area (Figure 6.18). The locations of the correctly classified pixels are shown in Figure 6.20.
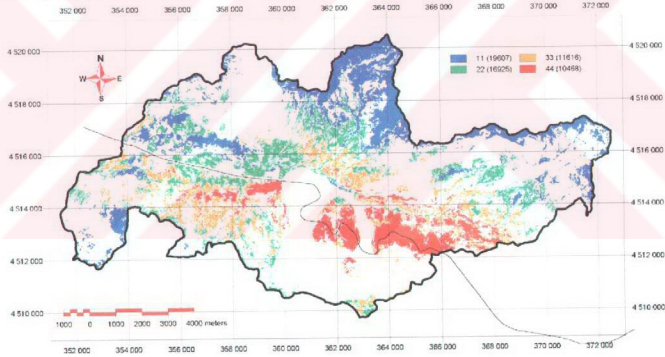
Figure 6.20. The locations of the correctly classified pixels (the numbers in the legend indicate the pixel counts).

Another pixel association is called the acceptable classification when the hazard classes in both of the hazard maps are differing from each other by one rank in the hazard classification scheme. Such as a change from very low hazard to low hazard, or change from high hazard to very high hazard is acceptable. This association is indicated in the summation map by the following id's 12,21,23,32,34 and 43. The area covered

with this association is 48,69 km^2 corresponding to 45.36 % of the total area. The locations of acceptable pixels are shown in Figure 6.21.

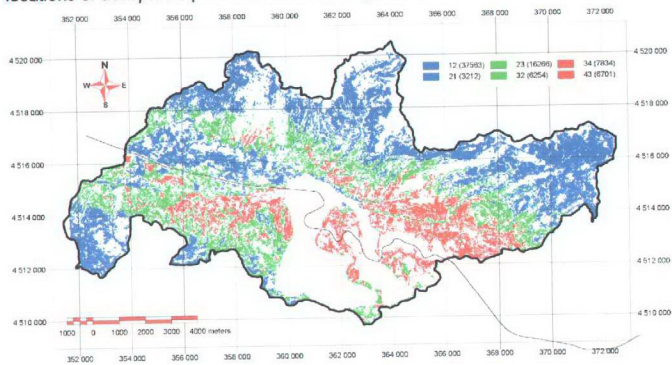**Figure 6.21.** The locations of the acceptable pixels (the numbers in the legend indicate the pixel counts).

The not acceptable pixels are defined as the difference of hazard ranks should be more than one rank. Such as changing from very low hazard to very high hazard is not acceptable. These pixels re-represented in the summation as 13, 14, 24, 31, 41 and 42. The area covered by these not acceptable pixels are 20.79 km^2 with corresponding 19.6 % of the area. The locations of these pixels are shown in Figure 6.22.

Upon the investigation of the not accepted pixels of the two methods, it is seen that the bivariate hazard map is overestimating the hazard classes relative to the logical hazard map. As shown in the legend of Figure 6.22, six pixel values are present and the first three of them has greater occurrences than the remaining three. The first three was 13, 14 and 24, indicating that they belong to high and very high hazard class in bivariate map and low to very low hazard class in logistic regression map. Although this is a relative comparison, it can be said that LHM is underestimating the hazard, however, the classification scheme fits well in the remaining three pixel values. If LHM was underestimating the hazard the remaining three pixel counts should be more that of the observed values. A further investigation is made in order to find the reasons why BHM overestimates the hazard with the aid of initial parameter maps, percentile maps and the hazard maps. It is found that most of the errors are dependent on the percentile division of the parameter maps. Such as the first percentile of distance from E-5 has the most weight among the other all percentiles, which is the result of the faint E-5 highway trace in the western part of Figure 6.22. Also the fault density and distance to fault percentile maps are responsible to the mismatch of the two hazard maps in the areas shown with

arrows in Figure 6.22. On the other hand, the good correlation in the high and very high hazard classes of both methods should not be underestimated.
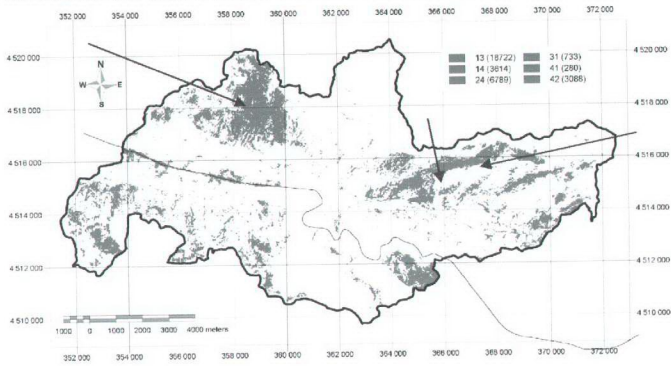
Figure 6.22. The locations of the not acceptable pixels (the numbers in the legend indicate the pixel counts).

In order to see where both hazard methods have the same or acceptable classifications, the acceptable pixels and the correct pixels are added up, which represents an acceptable classification of nearly %80 of the area. This result is presented in Figure 6.23.
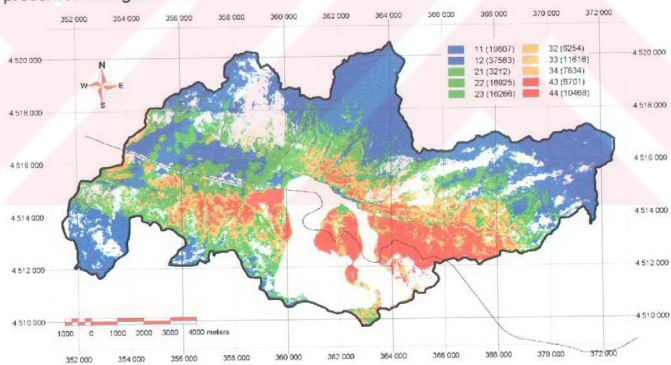
Figure 6.23. The locations of the correctly classified and the acceptable pixels united (the numbers in the legend indicate the pixel counts).

After a comparison of Figure 6.23 with other hazard and error maps, and taking into consideration about the reasons of erroneous pixels the hazard map produced from logical regression analysis is decided to be used for further analyses. The final hazard map with the infra structure overlaid is shown in Figure 6.24.
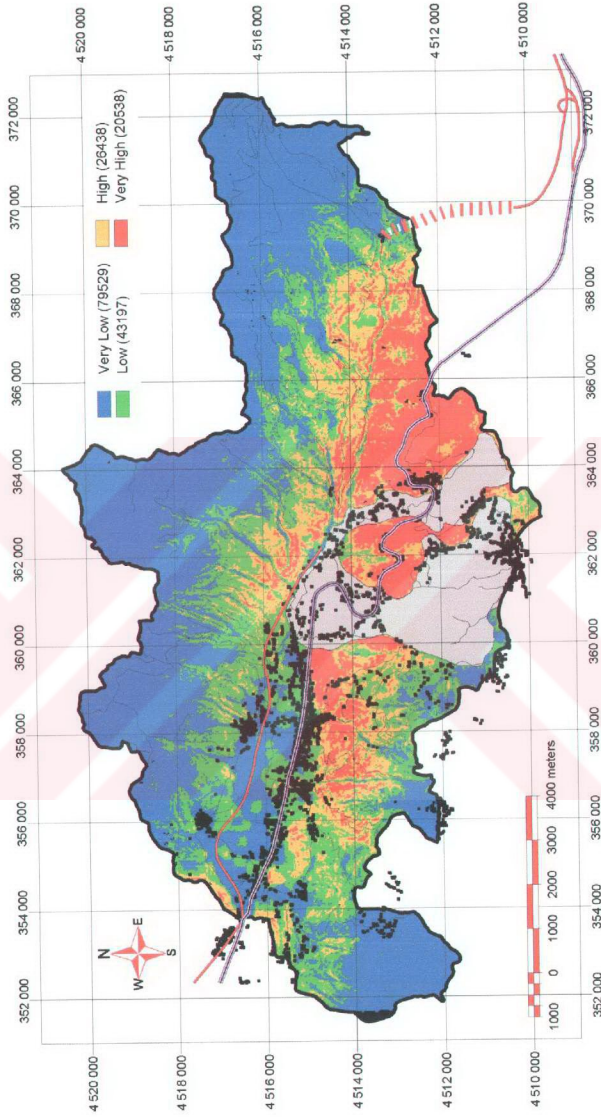
Figure 6.24. The final hazard map and the infrastructure of Asarsuyu catchment.

169

CHAPTER 7

DISCUSSION

The aim of this chapter is to endure step by step through the landslide assessment procedures while evaluating the possible errors and their reasons. Furthermore, a comprehensive flowchart will be produced while discussing and evaluating these stages.

7.1. Data Production

In harmony with the purpose and scope of the project three data domains were used to generate the needed data. These domains are remote sensing, topographical maps and existing geological maps.

The working size of the pixels for all of the products are selected as 25 by 25 meters. By rule of thumb, 1 map millimeter equivalent of the base map is adequate. Finer resolution would yield in larger errors as the locations of the pixels would be suspicious, moreover, the products other than topographic map should have to be resampled rigorously.

The first step through any hazard assessment scheme is the evaluation of current situation and preferably having information about the past states of the hazard. In order to achieve this goal, remote sensing technology is used, unless there are no archive information and no monitoring stations.

The spatial resolution of the satellite images is becoming as good as comparable to the aerial photographs, however, the prices of the same resolution satellite products are incomparable to conventional aerial photographs. The aerial photographs are still the cheapest solution, in addition the military bloom of the cold war resulted in presence of at least one full coverage of the country almost in all nations across the world, which are hosting the true and non speculative information about the past situation of the hazards. On the other hand, the presence of distortions in the aerial photographs makes the information extraction stage quite painful, as the interpretation should have to be transformed into plain cartographic coordinates. Recent

developments in the computer based photogrammetry reduce the time required for these transformations but requires another investment item through the institute's budget. Nevertheless, with the aid of these systems the horizon of the researchers would be quite broadened as they will be able to derive the topographical and morphological decision rules of the period at which the aerial photographs are taken.

In this research, the spatial positions of the landslides in 4 different periods were extracted from the aerial photographs and cordially transformed into base maps manually (Figure 7.1). The interpretation is acquainted with a photo checklist regarding the observable characteristics of the landslides. This checklist in the further stages would be converted to a database attached to every single landslide.

This base map transformation, when done by experienced interpreters, is quite comparable with the results obtained from digital products. Yet, the manual transformation is always remaining as a large debate over one of the principal inputs of the hazard assessment procedures. This transformation inherits a minor subjectivity to the system depending on the experience of the interpreter.
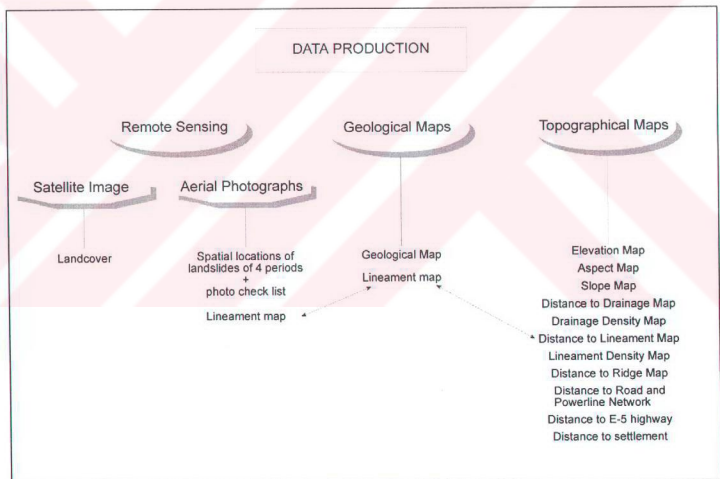

Figure 7.1. The elements of data production stage.

Though the aerial photographs are considered as the most capable and cheapest item within the remote sensing products, the great spectral resolution of satellite products still have superiority over the panchromatic aerial photographs.

However, their restricted superiority also restricts their usages as they could now only be used for land cover extraction or for gathering regional geological information. The new trends of researches point out that probably within a decade the prices of the satellite and historical aerial photographs would equalize. Furthermore, the researches in active sensors (radar systems) and laser-dominated systems (Lidar) are flourishing, which are promising activities for the next decade beyond the hazard assessment point of view. Landsat TM 5, due to its spectral superiority and price advantage, is used to extract the land cover information in Asarsuyu catchment, which would be treated as a parameter map in further stages (Figure 7.1).

The second and the most speculative domain is the geological map domain, as the maps have been prepared for general geological needs without considering the special needs of the hazard assessment procedures. Further groupings regarding the material are made rather than its stratigraphic content and age. These maps were compiled from existing literature, so coordinate mismatches and information inadequacies are frequent in the published maps. These geological unit mismatches also inherits a speculative perspective to the map prepared which is also dependent on the experience of the geologist. The lack of coordinate system in the compilation map creates a positional error, which should be eliminated by further resampling of the geological maps. Consequently the lineament information has same deficiencies, as though it was both compiled from existing maps and refined after aerial photographical interpretation (Figure 7.1)

The third and relatively the least speculative domain is the topographical map domain, assuming that there are not any errors present in the elevation information. In this thesis 11 variables out of 13 variables are derived from topographical maps (Table 5.16). The morphology, and infrastructure information are all derivatives from these topographical maps (Figure 7.1). The topographical maps have minimal processing errors as they have been digitized from the hardcopy templates by the Mapping Headquarters. The nature of the information present in the topographical maps generally consisted of vector coverages, as the representations of contour lines in point, line or polygon form. This form is not the most suitable form to be used in a GIS, yet it has to be converted into continuous maps. When the vector data has z attributes (topographical contours), this conversion is achieved through gridding operations. On the other hand if the vector object has no z attributes (infrastructure information), either the density of the feature or the nearest distance of the pixels to that feature is calculated to form a continuous map. A new approach is taken in this thesis, as taking the true distances to the objects rather than taking the map distances, which inturn duplicates the importance of the accuracy of the elevation model, as true distance is the function of DEM (Figure 4.5).

In order to achieve standardization, through the three domains, all of the maps were resampled by using a reference raster as the DEM. This resampling is vital to fix the extents and the centers of the pixels.

The products of aerial photograph and geological map domains have quite large subjectivities, when created by novice users, which should then have to be treated very carefully. The topographic map domain remains more objective than the others, as the errors might rise only from mechanical procedures, which could be validated more easily without thematic information about the maps.

Although the sequences presented here is quite objective the data domain and data type selection is still dependent on the available data and also dependent on the experience of the user, in which the assumption is that the selected maps will represent adequately the sliding conditions. Extreme caution should be applied in data set selection in order not produce redundant information. Although the distance to drainage and drainage density pair and distance to fault and fault density pairs seems to create duplicate information, they do possess different information. Distance to drainage maps are controlling the chance of any stream to erode the toe part of the landslide, while the drainage density yields in hydrological properties of the lithologies. Similarly the distance to fault map has information about the seismic activity and if not sealed would yield in increase in surface water with springs, while the fault density is a direct representation of how crushed the rock units are in the area. As noted above maps like distance to n^{th} Strahler order, or distance to epicenters are not used in order not to create redundant data sets and they are still not dependable as the errors in the epicenter locations are far beyond the resolution of such studies.

7.2. Data evaluation

The maps produced include the information of a particular parameter considering whole of the study area. The whole area is needed to be subdivided as the decision rules should have to be created from the landslide related pixels. In order to quantitatively discriminate the regions, a point mesh structure is laid over the region. The landslide polygons are cropped out and the nodes are stored in "slided mass" database with parameter map information.

A new approach is followed in the generation of decision rules of landsliding mechanism, as it is believed that the best undisturbed morphological conditions (conditions before landslide occurs) would be achieved by adding a buffer zone to the crown and side areas. These buffer areas are then extracted from the point mesh and the parameter map information is transformed. The resulted spreadsheet is converted into a database, which is named as "seed cells". The information in this database is used to create the decision rules (Figure 7.2).

The databases depending on "slided mass" are evaluated in Chapter 5. They are used to characterize the landslides and to figure out the after-slide morphology. Furthermore, the investigation of size and shape changes through 4 periods resulted in quite large contributions for the reasons of failure. On the other hand the "seed cell" database is used to create decision rules about sliding, as it reflects the most likely pre-sliding conditions.

The data mining results of both the "slided mass" database and the "seed cells" database are presented in detail in Chapter 5, hence only important issues will be raised here.

After the examination of 4 periods of polystats, fuzzystats and photo-characteristics (photo checklist) database (Figure 7.3), it is found that by using these three databases only very broad conclusions could have been achieved. Such as, the removal or change of land cover and construction of significant engineering structures. Among these databases the photo-characteristics database is found to be the most useful one, unless there are different types of landslides in the project area. Unlike to the literature (Carrara and Merenda, 1974; van Westen, 1993; Soeters and van Westen, 1996), it should have to be stressed that the single use of any of these databases would neither result in any factual conclusion about the events going on in the geomorphological history of the area nor would characterize the landslides.
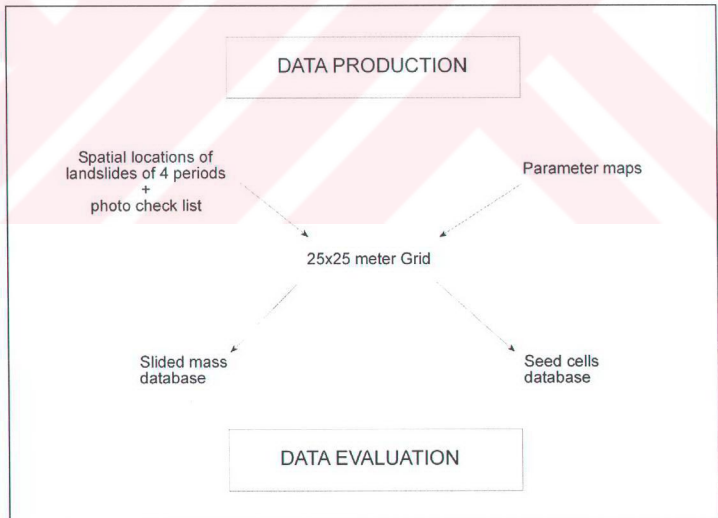
Figure 7.2. Snapshot methodology for information transformation.

The examination of the parameter distributions of "seed cell" database resulted in some conclusions about the contributions of every single parameter to the landsliding phenomena and /or the preferences of landslides on parameters. Other contributors in literature such as Lessing et al. (1983); Turrini and Visintainer (1998); Gupta and Joshi (1990), Carrara et.al (1991), Pachauri and Pant (1992); van Westen (1993), Chung et al. (1995), Soeters and van Westen (1996), Gupta and Anbalagan, (1997) and Guzzetti et al. (1999) tried to solve this parameter contributons by generalizing the whole event such as, getting averages parameters in slope facets or creating unique condition areas or creating dummy variables. However, none of these proposed methodologies are applicable to individual pixel analyses; furthermore, no factual reason is valid to generalize and to degrade the resolution of adequately detailed data. Although seed cells let the user to upgrade the analyses to individual pixels, the interactions of parameter maps could not still be explored efficiently.
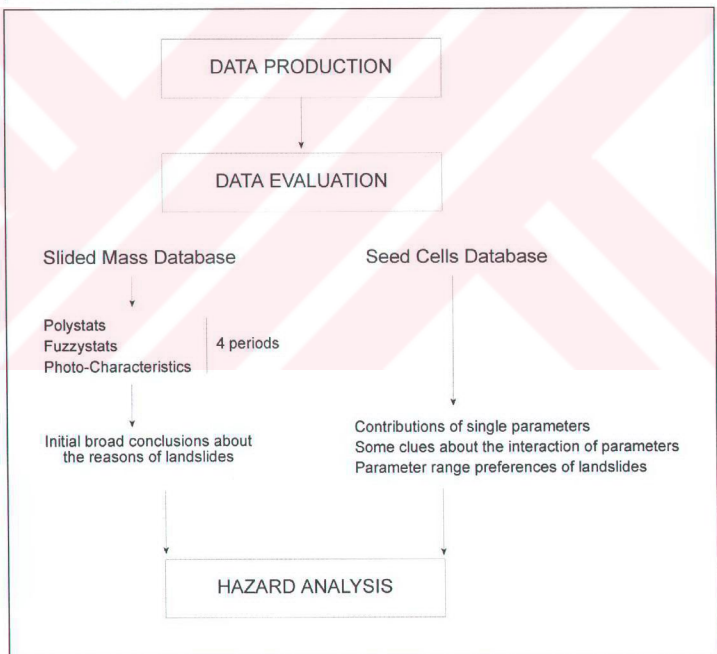
Figure 7.3 Components of data evaluation stage

7.3. Hazard Analysis

The aggregation of information through the steps of data production and evaluation yielded in a tailor-made hazard analysis. The analyses scheme started from the very primitive one, ending with the complex statistical models in order to show the limits of the proposed methodologies through the literature (Figure 7.4).

The most primitive analysis, Thematic Landslide Attribute Spatial Distribution Analysis (TLASDA), is explained in Section 6.1., which is extensively used by van westen and his colleaques (Wieczorek, 1984; van Westen, 1993; Soeters and van Westen, 1996). However, it is found that this methodology could only be useful in regional scale with very large number of landslides.

The second analysis, Landslide Activity Analysis (LACTA) is detailed in Section 6.2., which is found to be valuable, when used with polystats and fuzzystats databases, as this methods validates the conclusions about the land cover and landuse change in the study area. Single use of this analysis is of no use. However, the use of this methodology is encouraged as being the only method to evaluate the hazard with a historical perspective. Yet, it is still not advisable to base any hazard maps only on this analysis. Further examples could be seen in works of Canuti et al. (1979) and van Westen (1993).
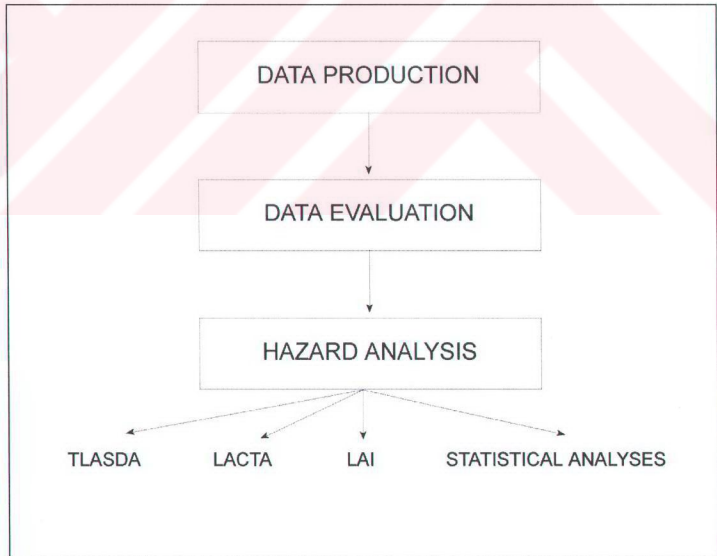
Figure 7.4. Components of Hazard Analysis.

The third analysis is Landslide Isopleth Analysis, which is explained in Section 6.3. This analysis reflects the current situation of the state, and to some extent shows preferences of parameters that is responsible of landsliding. Parameter map crossing could be exploited in order to show the relative preferences of landslides over parameter maps. Although the analysis seems to be robust and free of subjective results, it can turn into a very speculative map production method, as an experienced user can easily manipulate the size of the counting grid and the threshold of landslide density in order to show what he/she intends to emphasize. This analysis is also encouraged to use as an entry for initiation of statistical analyses in regional scales.

By the fourth analysis the hazard analysis scheme enters to a new realm, the statistical domain. Two main trends are observed in this realm as explained in Chapter 2, the bi-variate and multi-variate (Figure 7.5).
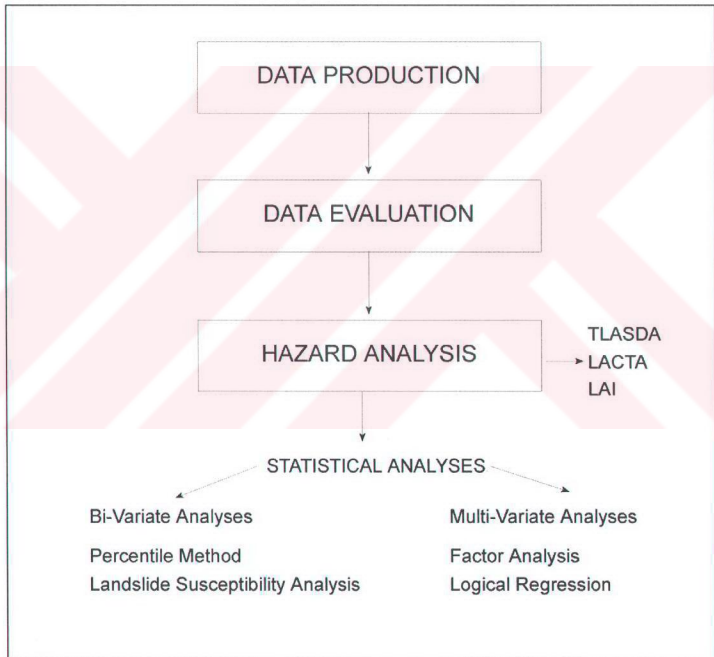
Figure 7.5. Components of Statistical analyses.

In order to proceed with bi-variate analysis, the continuous parameter maps should have to be converted into discrete maps, as the landslide densities in each class of the parameter maps constitute the core of this analysis. However, this conversion issue remains always unclear in literature as most of the authors use their expert opinion for the boundaries of the classes (Carrara and Merenda, 1974; Meneraud and Calvino, 1976; Kienholz, 1977, 1978; Stevenson, 1977; Malgot and Mahr, 1979; Kienholz et al., 1983, 1988; Ives and Messerli, 1981; Rupke et al., 1988; Gupta and Joshi, 1990; Pachauri and Pant,1992; van Westen, 1993, Soeters and van Westen, 1996; Gupta and Anbalagan, 1997). The use of expert opinion results in subjectivity and removes the reproducibility of the proposed methods for different area in the globe. Based on this issue, a data driven methodology is proposed, as the classes should be selected according to the percentile divisions of seed cells. The data outside the seed cell range is discarded as it is also been discarded by the nature by not having a single landslide at these regions. In order to reduce complexity in the calculations stage, 10 classes are proposed to use, as even in 10 classes with 13 parameters, 130 different classes should have to be maintained. The magic number of 10 as the class number, is selected arbitrarily considering the complexity. However, it is believed that lower class numbers would result in large generalizations in the final hazard map, where higher class numbers would result in isolated pixels, that should have to be filtered out, where filtering would alter the objectivity of the hazard map.

Furthermore, this data dependent division reduces the problem, in bi-variate analyses, of what weight should be given to each parameter map, as each class acts like a map and the ratio of the landslide density over the class area gives its natural weight.

In bi-variate analysis, explained in Section 6.4.1., all continuous parameter maps are converted to categorical variables, and corresponding weight values of each class is calculated and added up to create the final hazard map. The complexity of this analysis is moderate and it could easily be implemented to any scales. However, the evaluation of parameter maps should be done in caution, as some parameters gets quite large weights. Further divisions of parameter maps might be needed, as discussed in Section 6.4.1, the distance to E-5 highway gets the maximum natural weight but it has two distinct divisions. The mountain pass gets its desired weight, on contrary the Kaynaşlı valley pass of this highway gets also the same weight without having a single landslide nearby. Moreover, it should also be checked in detail that duplicate information should not be present in the parameter weights, such as the distance to E-5 highway and the road class of land cover in Table 6.7.

Further in the analysis, factor analysis is carried out. Although factor analysis does not yield in a hazard map, the new factors could easily be used to validate the

natural weights of bi-variate analysis and could easily create a base for logical regression analyses. Such that the largest component of factor analysis points out that human activity is the most important factor, which is in concordance with the natural weight scheme as the percentiles of distance to E-5 highway, land cover and distance to settlement get quite large natural weights.

The last analysis performed in Asarsuyu Catchment is Logical Regression, which is presented in section 6.4.2.2. Although this analysis is fairly new in Landslide Hazard Assessment realm, a new approach is followed. The previous researches (Atkinson and Massari, 1998; Dai et al. 2001; Lee and Min, 2001) use categorical variables in the analysis, which creates doubts in the selection of classes of variables, which was discussed in the former paragraphs. The new approach is to use the continuous data as it is, in order not to alter the state and information present in the parameter maps.

Although the understanding of this analysis requires a good statistical background, the application is quite simple. It is encouraged to be used in medium and large scale applications, if fed with adequate parameter maps. It is believed that the analysis would end in more accurate results when the number of parameter maps increases.

In both hazard maps 4 classes are used to define the degree of the hazard. They are very low, low, high and very high. The moderate hazard class is not implemented here, as the definition of moderate is obscure and the question of "being moderate relative to what?" is still a great debate over hazard analysis. More willingly than using 3 classed hazard scheme (low, moderate, high), a four class scheme is used as dividing the moderate class into two (low and high), and stretching the end members to very low and to very high.

7.4. Hazard Map Comparison

Upon the completion of the hazard maps, a quantitative comparison scheme should have to be implemented. In order to do this, a two fold methodology is followed (Figure 7.6). First one is to compare the areas of hazard classes and the corresponding densities of landslides. This analysis is important because the idea in optimum zonation refers to allocating minimum areas for high hazard zones, while covering most of the landslides present in the area. Based on this optimum zonation concept, an index is defined as Seed Cell Area Index (SCAI) and the two hazard maps are compared. The logical regression hazard map is found to be more accurate and possesses acceptable results relative to that of bivariate analyses.

The second method is to compare the hazard classes of the two produced maps via their spatial locations. In order to achieve this a re-coded matrix is prepared and presented in Table 6.19. It is found that 80 % of the two maps are converging into acceptable results. The remaining mismatched 20% of the area is reflected by the deficiencies of bivariate analyses as discussed in previous sections. The overweight of percentiles of distance to E-5 highway, distance to fault, fault density and the geological map are the major sources of this mismatch.
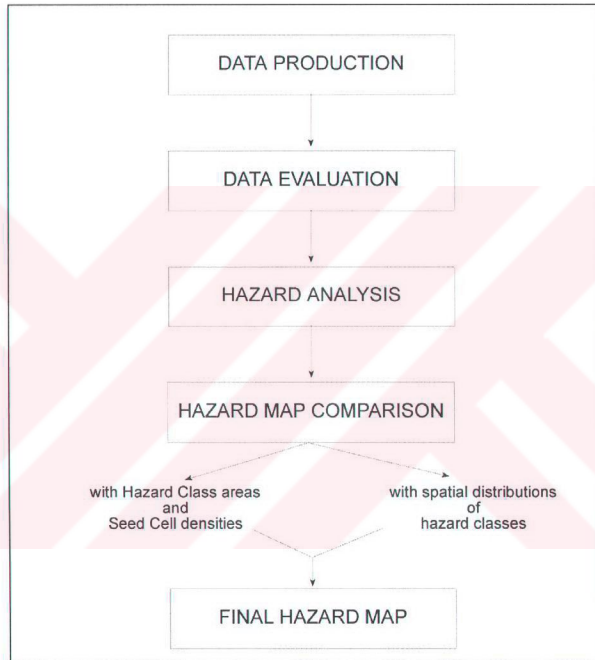
Figure 7.6. Components of Hazard Map Comparison

After it has been decided to use the logical regression hazard map as the final output, it is seen that, the northern slope of Asarsuyu catchment is definitely classified as very low hazard. The attributes responsible for this classification is quite reasonable as these areas are the least populated, the land cover is not disturbed and the cover is dense forest, very distant to E-5 highway and the major active fault, the lithology is

resistant enough, although the drainage density and slope values are higher than the rest of the area.

On the other hand the southeastern slopes are definitely on very high hazard class. The reasons could be listed as: the removal of lateral supports by E-5 highway cut slopes, fill areas of E-5 highway result in readily unstable unconsolidated material, close location to active faults, high disturbance to land cover, high activity of highway resulting in extra vibration and the presence of flyschoidal units.

CHAPTER 8

CONCLUSIONS

The additional arguments to the existing literature, obtained from the landslide hazard analysis of Asarsuyu catchment are grouped into two; conclusions about landslide phenomena in Asarsuyu catchment and new approaches generated for data driven landslide hazard assessment.

A. Landslide Phenomena in Asarsuyu Catchment:

1. The Asarsuyu catchment is investigated for four different periods. It is found that except Bakacak and Bülbülderesi landslides there exists 33 individual landslides in 1952, 45 in 1972, 39 in 1984 and 37 in 1994.

2. The major causes for generation and for reactivation of landslides is found to be the change in land cover, and the increasing interaction of human activity resulting in land use changes.

3. The northern slopes of Asarsuyu catchment is classified as very low hazard. The attributes responsible for this classification is quite reasonable as these areas are the least populated, the land cover is not disturbed, dense forest, very distant to E-5 highway and to the major active fault and the lithology is intact enough.

4. The Southeastern slopes, especially the Bolu Mountain Highway pass are definitely on very high hazard class. The reasons could be listed as: the removal of lateral supports by E-5 highway cut slopes, fill areas of E-5 highway resulting in readily unstable unconsolidated material, close location to active faults, high disturbance to land cover, high activity of highway resulting in extra vibration, and the presence of flyschoidal units.

B. New Approaches for Data Driven Landslide Hazard Assessment Procedures:

1. In the conversion of non attributed continuous data into categorical variables, true distances are used rather than map distances.
2. In the generation of decision rules of landsliding mechanism, a buffer zone is added to the crown and side areas, to reveal the best undisturbed morphological conditions to be stored as attributes of a new database called "Seed Cell Database".
3. The decision rules are directly based on Seed Cell Database such as the data outside the seed cell range is discarded as it is also been discarded by the nature by not having a single landslide.
4. In diving the classes of parameter maps, a new method is proposed and presented as percentile method, which is also dependent on "Seed Cell Database".
5. Factor analysis is used to figure out new obscure factors present in the current data set.
6. Logical regression is implemented and forced to be used with both categorical and continuous data sets.
7. For hazard map comparison a two fold methodology is proposed: 1).area comparison in which a new index SCAI is introduced and 2). Spatial comparison in which re-coding matrix is presented.

Following issues can be listed as recommendations for future research:

1. High resolution remote sensing products can be used to derive DEM's of any desired time period in order to check the system accuracy and to re-establish the hazard assessment at any time
2. New data sets such as: rainfall, geotechnical, hydrological and soil map related data, could be implemented in this hazard assessment procedures, if problems in data production stages are solved.
3. Seismic data could be implemented regarding different lithologies, if microtremor data, and P and S wave distribution maps are made available.

REFERENCES

Abdülselamoğlu, S., 1959, Almacıkdağı ile Mudurnu ve Göynük civarının jeolojisi, İstanbul Üniversitesi, Fen Fakültesi Monografileri, 14.

Afifi, A.A. and Clark, V, 1998, Computer aided multivariate analysis, Chapman-Hall, London, 455p.

Agterberg, F.P., Bonham-Carter, G.F., and Wright, D.F., 1990, Statistical pattern integration for mineral exploration, in: Proceedings COGEODATA Symposium, Espo Finland, Gaal G. ed.

Aktimur, T., Algan, Ü., Ateş, Ş., Oral, A., Ünsal, Y., Karatosun, H., Öztürk, V. and Sönmez, M., 1983, Bolu ve yakın çevresinin yerbilimleri sorunları ve muhtemel çözümleri, MTA Rap No:7387 (unpublished)

Akyüz, H.S., Barka, A., Altunel, E., Hartleb, R. and Sunal, G., 2000, Field observations and slip distributions of the November 12, 1999 Düzce earthquake (M=7.1), Bolu – Turkey. In 1999 Izmit and Düzce Earthquakes: preliminary results, Barka, A., Kozacı, Ö., Akyüz, S. and Altunel, E. Istanbul University Press. Istanbul. 63-70.

Ambraseys, N.N., 1970, Some characteristic features of the North Anatolian Fault Zone, Tectonophysics, 9, 143-165.

Ambraseys, N.N. and Finkel, C., 1995, The Anatolian Earthquake of 17 August 1668, Proc of the Symp.on Historical Seismograms and Earthquakes, 400-407.

Anderson, M.G. and Richards, K.S., 1987, Slope Stability. Geotechnical Engineering and Geomorphology, Wiley and Sons, New York, 648 p.

Archibald, P.D., 1987, GIS and remote sensing data integration, Geocarto International, Vol.3, pp. 67- 72.

Astaldi, 1990, Geological report: Preliminary design Anatolian Motorway Gümüşova-Gerede stretch No:2, Report number 2014.

Astaldi, 1995, Updating geological research results in the Bolu and Düzce area, Barka, A., Astaldi report no: 45,110/R/2097

Atkinson, P.M. and Massari, R., 1998, Generalized linear modelling of susceptibility to landsliding in the central Apennines, Italy. Computers and Geosciences, 24, 4, 373-385.

Aydan, Ö., Ulusay, R., Kumsar, H. and Tuncay, E., 2000, Site investigation and engineering evaluation of the Düzce-Bolu earthquake of November 12, 1999, Turkish Earthquake Foundation, TDV/DR 09-51, 220p.

Aydan, Ö. and Dalgıç, S., 1998, Prediction of deformation behaviour of 3-lanes Bolu Tunnels through squeezing rocks of North Anatolian Fault Zone (NAFZ). Regional Symp on Sedimentary Rock Engineering. Taipei, Taiwan, 1998, 228-233

Aydın, M., Serdar., S., Şahintürk, Ö., Yazman, M., Çokuğraş, R.,Demir, O. and Özçelik, Y., 1987, Çamdağ (Sakarya) – Sünnicedağ (Bolu) yöresinin jeolojisi, TJK bülteni, 30, 1-14.

Barka, A., and Erdik, M., 1993, Active Faults of Gümüşova-Gerede highway, Bolu Mountain, Astaldi report, 25p.

Batum, İ., 1968, 1/25,000 ölçekli Adapazarı G26-b3 paftasının Jeolojisi, MTA Rap No:4778 (unpublished).

Blumenthal, M., 1948, Bolu civarı ile Aşağı Kızılırmak mecrası arasındaki Kuzey Anadolu silsilelerinin jeolojisi, MTA Yayınları, Seri B, No:13, 265p.

Bonham-Carter, G.F., 1996, Geographic information systems for geoscientists, modelling with GIS. Pergamon Press, Canada, 398 p.

Bonham-Carter, G.F., Agterberg, F.P., Wright, D.F., 1990, Weights of evidence modelling: a new approach to mapping mineral potential. Geologcal Survey of Canada Paper 8-9. Agterberg, F.P. and Bonham and Carter, G.F. (eds). Ottowa, Canada 171-183 pp.

Brabb, E.E., 1984, Innovative approches to landslide hazard and risk mapping. In Proc., Fourth International Symposium on Landslides, Canadian Geotechnical Society, Toronto, Canada, vol: 1, pp. 307-324.

Bromhead, E.N., 1986, The stability of slopes. Surrey University Press, Surrey, 333 p.

Brunsden, D., Doornkamp, J.C., Fookes, P.G., Jones, D.K.C. and Kelly, J.M.H, 1975, Large scale geomorphological mapping and highway design. Quarterly Journal of Engineering Geology, Vol 8, 227-253.

Burrough, P.A., 1986, Principals of geographical information systems and land resources assessment, Clerandon Press, Oxford, England, 194 pp.

Canik, B., 1980, Bolu sıcak su kaynaklarının hidrojeolojisi, Selçuk Üniversitesi Fen Fakültesi Yayınları, 1, 74 p.

Canuti P., Frascati, F., Garzonio, C.A., and Rodolfi, C., 1985, Stabilita dei versanti nell area representative di Montespertoli (Firenze) S.E.L.C.A. Firenze. 2 map sheets.

Canuti P., Frascati, F., Garzonio, C.A., and Rodolfi, C., 1986, Slope stability mapping in Tuscany, Italy. In International Geomorphology, Part 1, Gardiner, V. (ed). Wiley and Sons, New York, 231-239.

Canuti, P., Frascati, F., Garzonio, C.A., and Rodolfi, C., 1979, Dinamica morfologica di un ambiente sogetto a fenomeni franosi e ad Intensa attiva agricola. Publ. No: 142. Consiglio Nazionalle delle Ricerche, Perugia, Italy, pp-81-102.

Carrara A., 1993, Uncertainity in evaluating landslide hazard and risk, in Prediction and Percention of Natural Hazards ed. Nemec, J., Nigs, J.M. and Siccardi, F., Kluwer, Dordrecht, The Netherlands, 101-111.

Carrara, A., 1983, Multivariate models for landslide hazard evaluation. Mathematical Geology, vol.15, no.3, pp.403-427.

Carrara, A., 1988, Landslide hazard mapping by statistical methods: A "Black Box" approach. In Workshop on Natural Disaster in European Medditerranean Countries, Perugia, Italy, Consiglio Nazionale delle Ricerche, Perugia, pp. 205-224.

Carrara, A. and Merenda, L., 1974, Metodologia per un censimento degli eventio franoso in Calabria. Geologia Applicata e Idrogeologia, Vol 10, 237-255.

Carrara, A., Cardinalli, M., Detti, R., Guzzetti F., Pasqui, V. and Reichenbach. P., 1990, Geographical information systems and multivariate models in landslide hazard evaluation. In ALPS 90 Alpine Landslide Practical Seminar, Sixth International Conference and Field Workshop on Landslides, Aug 31 - Sept 12, Milan, Italy, Universita degli Studi de Milano., pp.17-28.

Carrara, A., Cardinalli, M., Detti, R., Guzzetti F., Pasqui, V. and Reichenbach. P., 1991, GIS techniques and statistical models in evaluating landslide hazard, Earth Surface Processes and Landforms, vol.16, no. 5, pp. 427-445.

Carrara A., Cardinali M., and Guzzetti F., 1992, Uncertainity in assessing landslide hazard and risk. ITC Journal, 2, 172-183.

Carrara, A., 1989, Landslide hazard mapping by statistical methods: A black box approach. Proc. Int. Workshop Natural Disasters in Europ.-Mediterr. Countries, Perugia, June 27 – July 1, 1988, CNR-USNSF, 205-224.

Cerit, O., 1983, Mengen (Bolu NE) yöresinin jeolojik incelenmesi, Yüksek lisans tezi, Hacettepe Ü., 160 s.

Cerit, O., 1990, Bolu Masifinin jeolojik ve tektonik incelemesi, Doktora Tezi, Hacettepe Ü., 217p.

Chandler J.H. and Brunsden D., 1995, Steady state behaviour of the Black Ven Mudslide: the application of archival analytical photogrammetry to studies of landform change. Earth Surface Processes and Landforms, 20, 255-275.

Chandler J.H. and Moore, R., 1989, Analytical photogrammetry: A method for monitoring slope instability. Quarterly Journal of Engineering Geology, 22, 97-110.

Chung, C.J. and Fabbri, A.G., 1993, The representation of geoscience information for data Integration. Nonrenewable Resources, 2-3, 122-139.

Chung, C.F., Fabbri, A.G., van Westen, C.J., 1995, Multivariate regression analyses for landslide hazard zonation. In: Geographical Information Systems in Assessing Natural Hazards, Carrara, A. and Guzzetti, F. (eds). Kluver Academic Publishers, Netherlands, 107-134.

Congalton, Russell G., 1991, A review of assessing the accuracy of classifications of remotely sensed data, Remote Sensing Environment, Vol.37, pp. 35- 46.

Crippen, R.W. ,1990, Calculating the vegetation index faster. Remote Sensing of Environment, vol 34., pp. 71-73.

Crist, E.P. and Cicone, R.C., 1984, Application of the Tasseled cap concept simulated to Thematic mapper data. Photogrammetric engineering and remote sensing, 50, 3, 343-352.

Crozier, M.J., 1973, Techniques for the morphometric analysis of landslips. In: Slope Stability, Brunsden, D. and Prior, D.B. (eds), Wiley and Sons, New York, pp 103-142.

Crozier, M.J., 1986, Landslides: Causes, Consequences and Environment. Croom Helm, England, 245 p.

Cruden, D.M., 1991, A simple definition of a landslide, Bulletin of International Association of Engineering Geology, 43, 27-29.

Dai, F.C., Lee, C.F. and Xu, Z.W., 2001, Assessment of landslide susceptibility on the natural terrain of Lantau island, Hong Kong. Environmental Geology, 40,3, 381-391.

Dalgıç, S., 1994a, Anadolu otoyolu Bolu dağı geçişinin mühendislik jeolojisi, Doktora Tezi, İstanbul Üniversitesi, Fen bilimleri Enstitüsü, 213p. (unpublished)

Dalgıç, S., 1994b, Anadolu otoyolu Bolu dağı geçişinin mühendislik jeolojisi, Türkiye Jeoloji Kurultayı Bülteni, 9, 393-397.

Dalgıç, S., 1997, Distribution of lithological and structural features along Bolu tunnel, Anatolian Motorway, Geological Bulletin of Turkey, 40-2,39-48.

Dalgıç, S.,1998a, Slope stability problems of the weak rocks in the Asarsuyu pass of the Anatolian Motorway. Bull. Eng. Geol. Env., 57, 199-206.

Dalgic S, 1998b, Selection of crushed rock quarries for the construction of the Anatolian Motorway ,Environmental & Engineering Geoscience ,4: (4) 511-518

Dalgıç, S. and Gözübol, A.M., 1995, Bolu otoyol tünelinde stabilite problemleri. Geosound, 27, 73-80.

Dalgic, S., Şimşek, O. and Gözübol, A.M., 1995 Anadolu otoyolu Bolu Yumrukaya geçişinde heyelanların etkisi, İkinci Ulusal heyelan Sempozyumu, Adapazarı, 163-170.

Davis, J.C., 1986, Statistical and data analysis in geology, Wiley and Sons, New York, 546 pp.

Demirtaş., R. and Yılmaz, R., 1996, Türkiyenin Sismotektoniği, T.C. Bayındırlık ve İskan Bakanlığı, Ankara., 91 p.

Emerson, J.D., and Hoaglin, D.C., 1983, Stem and leaf displays. In: Understanding robust and exploratory data analysis. Hoaglin, D.C., Mossteiler, F. and Tukey, J.W., Wiley and Sons, New York.

Erendil, M., Aksay, S., Kuşçu, İ., Oral, A., Tunay, G. and Temren, A., 1991, Bolu masifi ve çevresinin jeolojisi, MTA Rap No:9425 (unpublished)

Faust, N.L., Anderson, W.H., Star, J.L., 1991, Geographic information systems and remote sensing future computing environment, Photogrammetric Engineering & Remote Sensing, Vol.57, No. 6, pp. 655- 668.

Fookes, P.G., Dale, S.G., Land, J.M., 1991, Some observations on a comparative aerial photography interpretation of a landslipped area. Quarterly Journal of Engineering Geology, 24, 249-265.

Freedman, D., and Diaconis, P., 1981, On the histograms a density estimator L_2 theory. Zeitschrift für Wahrscheinlichkeustheorie und verwandte Gebette, 57, 453-476.

Gagon, H., 1975, Remote sensing of landslide hazards on quick clays of Eastern Canada. Proceedings 10th International symposium on remote sensing of environment. ERIM, Ann Arbor, Michigan, 803-810.

Gedik, A., ve Alkaş, İ, 1996, Bolu yöresinin Jeolojisi ve karbondioksid olanakları, MTA rap No: 10147. (unpublished).

Gee, M.D., 1992, Classification of landslide hazard zonation methods and a test of predictive capability. Proocedings 6th International symposium of landslides, Christchurch, New Zealand, Vol 2, 947-952.

Gillespie, A.R., Kahle, A.B. and Walker, R.E., (1987), Color enhancement of highly correlated images. I.Decorrelation and HSI contrast stretches. *Remote Sensing of Environment*, 20, pp. 209-235.

Görmüş, S., 1980, Yığılca (Bolu NW) yöresinin jeolojik incelemesi, Doktora Tezi, Hacettepe Üniversitesi. 210s.

Görmüş, S., 1982b, Yığılca (Bolu KB) yöresinin tektoniği ve jeolojik evrimi, H.Ü. Yerbilimleri, 9, 133-140.

Görmüş, S., 1982a, Yığılca (Bolu KB) yöresinin stratigrafisi, H.Ü. Yerbilimleri, 9, 91-110.

Gözübol, A.M., 1978, Mudurnu-Dokurcun-Abant (Bolu ili) alanının jeoloji incelemesi ve Kuzey Anadolu yarılımının yapısal özellikleri, Doktora tezi, İstanbul Üniversitesi Fen Fakültesi Tatbiki Jeoloji Kürsüsü.

Graham, J., 1984, Methods of stability analysis. In: Slope Stability, Brunsden, D. and Prior, D.B. (eds), Wiley and Sons, New York, pp 171-215.

Grunder, M., 1980, Beispiel einer anwendungsorientierten gefahrenkartierung 1:25.000 für forstliche sanierungsprojeckte im berner oberland (Schweiz). Proceedings INTRAPRAEVENT, Bad Ischl,Austria, Band 4, p. 353-360.

Güler, B., 1999, Bolu Dağı Bakacak (Bolu) yöresi jipslerinin jeolojik incelemesi, Yüksek Lisans Tezi, Hacettepe Ü., 46 s.

Gupta, P. and Anbalagan, R.,1997, Slope stability of Tehri dam reservoir area, India using landslide hazard zonation (LHZ) mapping. Quarterly Journal of Engineering Geology, 30, 27-36

Gupta, R.P. and Joshi, B.C., 1990, Landslide hazard zoning using the GIS approach – A case study from the Ramganga catchment, Himalayas. Engineering Geology, 28, 119-131.

Guzzetti, F., Carrara, A., Cardinalli, M., Reichenbach, P.,1999, Landslide hazard evaluation: A review of current techniques and their application In a multi-case study, central Italy. Geomorphology, 31, 181-216

Hammond, C.J., Prellwitz, R.W., and Miller, S.M., 1992, Landslide hazard assessment using monte carlo simulation. In Proc. Sixth International Symposium on Landslides (Bell, D.H. ed), Christchurch, New Zealand, A.A. Balkema, Rotterdam, Netherlands, vol.2, pp 959-964.

Hansen, A., 1984, Landslide hazard analysis, In: Brunsden D. and Prior, D.B., Slope Instability, p. 523-602. Wiley, Newyork.

Hartlen, J. and Viberg, L., 1988, General report: evaluation of landslide hazard. In Proc., Fifth International Symposium in Landslides (C.Bonnard ed), Lausanne, A.A. Balkema, Rotterdam, Netherlands, Vol.1, pp. 3-35.

Hearn, G.J., 1992, Terrain hazard mapping at Ok Tedi mine, Papua New Guinea. Proocedings 6th International symposium of landslides, Christchurch, New Zealand, Vol 2, 971-976.

Hermelin, M., 1990, Bases fisicas para los planes de desarollo de los municipios de Riseralda. AGID report no:13. Environmental geology and natural hazards of the Andean region, 269-274.

Hermelin, M., 1992, Medio ambiente, planes de desarolloy toma de decisions. Proceedings 1st Simposio Internacional sobre sensors remotos y sistemas de informacion geografica (SIG) para el estudşo de riesgos naturals, Bogota, Colombia, 646-663.

Hoek, E. and Bray, J.W., 1981, Rockslope engineering, Institute of mining and Metallurgy, London, 358 p.

Huang, S.L. and Chen, B.K., 1991, Intergration of Landsat and terrain information for landslide study. Proceedings 8th Thematic Conference on Geological Remote Sensing (ERIM), Denver, Colorado, Vol 2, 743-754.

International Association of Engineering Geology (IAEG), 1976, Engineering geological maps: A guide to their preparation, UNESCO Press, Paris, 79pp.

Işın, Z.S., 1999, D-100 (E-5) Karayolu, Bolu Dağı geçişi, Kaynaşlı – Elmalık arasının mühendislik jeolojisi ve Taşaltı Heyelan Bölgesinin jeoteknik değerlendirilmesi, Yüksek Lisans Tezi. İ.T.Ü., 145s.

Ives, J.D. and Messerli, B., 1981, Mountain hazard mapping in Nepal. Introduction to mountain research project. Mountain Research and Development, Vol 1, No: 3-4, 223-230.

Izenman, A.J., 1991, Recent developments in non-parametric density estimation. Journal of the American Statistical Association. 86, 413, 205-224.

Jenson, S.K. and Domingue, J.O. (1988). Extracting topographic structure from digital elevation data for geographic information system analysis. Photogrammetric Engineering and Remote Sensing, 54, 1593-1600.

Jones, F.O., Embody, D.R. and Peterson, W.C., 1961, Landslides along the Colombia river valley, North-eastern Washintgon. U.S.G.S. prof. Paper 367: 98 p. Washington.

Kam, Tin-Seong , 1995, Integrating GIS & RS techniques for urban land cover and land use analysis, Photogrammetric Engineering & Remote Sensing, Vol.57, 655- 668.

Kauth, R.J. and Thomas, G.S., 1976, The tasseled cap – A graphic description of spectral – Temporal development of agricultural crops as seen by Landsat. Proceedings 2nd International Symposium on Machien Processing of Remotely Sensed data, Purdue University, West Lafayette, IN. 41-51

Kaya, O., 1973, The Devonian and Lower Carboniferous stratigraphy of İstinye, Bostancı and Büyükada subareas, Paleozoic of İstanbul, ed. Kaya,O., 40, 1-36.

Kaya, O., 1982, Ereğli, Yığılca, Bolu Kuzey, Mengen alanlarının stratigrafisi ve yapı özellikleri. TPAO Rap No: 1639. (unpublished).

Kaya, O. and Dizer, A., 1981-1982 a, Mengen (Bolu) Eosen kömür havzasının stratigrafisi. MTA dergisi 97/98, 123 – 133.

Kaya, O. and Dizer, A., 1981-1982 b, Bolu Kuzeyi Üst Kretase ve paleojen kayarının stratigrafisi ve yapısı. MTA dergisi, 97/98, 57-77.

Kaya, O., Dizer, A., Tansel, İ. and Özer, S., 1986, Yığılca (Bolu) alanı üst Kretase ve Paleojeninin stratigrafisi, MTA derg., 107, 13-31.

Ketin, İ, 1955, Akçakoca-Düzce bölgesinin jeolojik lövesi hakkında memuar, MTA Rap No:2277 (unpublished)

Ketin, İ., 1966, Anadolunun Tektonik birlikleri. MTA Enstitüsü Dergisi, 66

Ketin, İ., 1967, Bolu – Gerede – Mengen ve Yığılca bölgesindeki Paleozoyik oluşuklara ait jeolojik rapor, TPAO Arama Grubu Rap No: 379. (unpublished).

Ketin, İ., 1969, Kuzey Anadolu Fayı hakkında. MTA Enstitusü Dergisi, 72, 1-27.

Kienholz, H., 1977, Kombinierte geomorphologische gefahrenkarte 1:10,000 von Grindelwald. Catena, vol.3, pp 265-294.

Kienholz, H., 1978, Maps of geomorphology and natural hazards of Grindelwald, Switzerland, scale 1:10.000, Arctic and Alpine Research, 10, 169-184.

Kienholz, H., 1980, Zur Anwendung des luftbildes bei der mittelmasstaabingen gefahrentkartierung für regionalplanereische zwecke in sclecht erschlossenen gebirgsraumen anhand von erfahrungen aus kartierungen in den Colorado Rocky Mountains. Proceedings INTRAPRAEVENT, Bad Ischl,Austria, Band 3, p. 155-172.

Kienholz, H., 1984, Hangstabilitatas un gefahrenbeurteilung im nepalischen mittelgebirge. Proceedings INTRAPRAEVENT, Villach, Austria, Band 2, p. 331-342.

Kienholz, H., Bichsel, M., Grunder, M. and Mool, P., 1983, Kathmandu Kakani area, Nepal: Mountain hazards and slope stability map. United Nations University, Mountain Hazards mapping Project. Map No :4, Scale 1:10.000.

Kienholz, H., Mani, P and Klay, M., 1988, Rigi Nordlene, Beurteilung der naturgefahren und waldbaliche prioritatenfestlegung. Proceedings INTRAPRAEVENT, Graz, Austria, Band 1, p. 161-174.

Kleinbaum, D.G., 1991, Logistic regression a self learning text. Springer Verlag, New York.

Kobashi, S., and Suzuki, 1988, Hazard index for the judgement of slope stability in the Rokko Mountain region. In Proc., Interpraevent, 1988, Graz, Austria, vol 1., pp. 223-233.

Koral, H., Dalgıç, S. and Gözübol, 1994, Bolu masifi gabroyik kayalarında mikrofabrik çalışma. Türkiye Jeoloji Kurultayı Bülteni, 9, 183-187.

Lambe, T.W. and Whitman, R.V., 1969, Soil Mechanics, Wiley and Sons, New York, 553 p.

Lee, S. and Min, K., 2001, Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology, 40, 1095-1113.

Lessing, P., Messina, C.P. and Fonner, R.F., 1983, Landslide risk assessment. Environmental Geology, Vol 5, No 2, 92-99.

Lillesand, T.M. and Kiefer, R.W., 1996, Remote sensing and image interpretation. Wiley and Sons, New York, 750p.

Lunetta, Ross S., Congalton, and Russell G., 1991, Remote sensing and geographic information systems data integration: error sources and research issues, Photogrammetric Engineering & Remote Sensing, Vol.57, No. 6, pp. 677- 687.

Malgot, J. and Mahr, T., 1979, Engineering geological mapping of the west Carpathian landslide areas. Bull.Int. Assoc. Eng. Geologists, 19, 116-121.

Mather, P.M., 1999, Computer processing of remotely sensed data: An Introduction, Chichester, New York, Wiley and Sons, 292 p.

Mc Donald, H.C. and Grubbs, R.C., 1975, Landsat imagery analysis: an aid for predicting landslide prone areas for highway construction. NASA Earth Resource Survey Symposium, Houston, Texas, Vol 1-b, 769-778.

Meneraud, J.P. and Calvino, A., 1976, Carte ZERMOS, Zones exposees a des risques lies aux mouvements du sol et du sous-sol a 1:25.000, region de la Moyenne Vesubie (Alpes Maritimes). Bureau de Recherches Geologiques Minieres, Orleans, France, 11p.

Mollard, J.D., 1986, Early regional photointerpretattion and geological studies of landslide terrain along the south Saskatchewan and Qu'appelle River valleys. Canadian Geotechnical Journal, 23, 79-83.

Mulder, H.F.H.M, 1991, Assessment of landslide hazard, Ph.D. Thesis, University of Utrecht, Netherlands, pp.150.

Mulder, H.F.H.M, and van Asch, T.W.J, 1988, A stochastic approach to landslide hazard determination in a forested area. In Proc., Fifth International Symposium in Landslides (C.Bonnard ed), Lausanne, A.A. Balkema, Rotterdam, Netherlands, Vol.2, pp. 1207-1210.

Neugebauer, J., 1994, Closing-up Structures, alternatives to pull-apart basins; the effect of bends in the North Anatolian Fault, Turkey, Terra Nova, 6, 359-365.

Neugebauer, J., 1995, Structures and kinematics of the North Anatolian Fault zone, Adapazarı-Bolu region, northwest Turkey. Tectonophysics, 243, 119-134.

Neugebauer, J., Löflerr, M., Berckhemer, H., Yatman, A., 1997, Seismic observations at an overstep of the western North Anatolian Fault (Abant-Spanca region, Turkey). Geol Rundsch, 86, 93-102.

Neuland, H., 1976, A Prediction model for landslips. Catena, 3, 215-230.

Nurlu, M., 1993, Kuzey Anadolu Fay Zonunda, (Bolu-Sapanca gölü arası) etken olan gerilimlerin fay analizleri ve uydu görüntüleri yardımı ile saptanması. 46. Türkiye Jeoloji Kurultayı Bildiri özleri. p. 121.

Okimura, T. and Kawatani, T., 1986, Mapping of potential surface-failure sites on granite mountain slopes. International Geomorphology, Wiley, New York, 121-138.

Orkan, N.İ., Aktimur, H., Sungur, G. and Işıklar, İ.S., 1977, Ankara-İstanbul otoyolu güzergahı Bolu dağı jeoloji incelemesi, MTA Rap No:2277 (unpublished)

Öztürk A., İnan, S. and Tutkun, Z., 1984, Abant-Yeniçağa (Bolu) yöresinin tektoniği, Cumhuriyet Universitesi, Yerbilimleri, 1, 1-18.

Pachauri, A.K. and Pant, M., 1992, Landslide hazard mapping based on geological attributes. Engineering Geology, 32, 81-100.

Perrot, A., 1988, Cartographie des risques de glissement en Lorainne. Fifth International Symposium in Landslides (C.Bonnard ed), Lausanne, A.A. Balkema, Rotterdam, Netherlands, Vol.2, pp. 1217-1222.

Rengers, N., 1986, Use of (stereo-) orthophotography prepared from aerial and terrestrial photographs for engineering geological maps and plans. Proc. 7 Th Int Symp on Remote Sensing for Resources Development and Environ. 665-668.

Rengers, N., Soeters, R. and van Westen, C.J., 1992, Remote Sensing and GIS applied to mountain hazard mapping. Episodes, 15, No:1, 36-45.

Rib, H.T. and Liang, T., 1978, "Recognition and identification", In: Schuster, R.L. and Krizek, R.J.,(eds.). Landslides analysis and control. National Academy of Science, Trans. Resear. Board Special Report, 176, p.34-80, Washington D.C.

Rupke, J., de Graaf, L.W.S., de Jong, M.G.G. and Verhofstad, J., 1987, A geomorphological Mapping system at scale 1:10.000 for mountainous areas. Zeitschrift für Geolomorfologie N.F., 31, No: 2, 229-242.

Rupke, J., Cammeraat, E., Seijmonsbergen, A.C., van Westen, C.J., 1988, Engineering geomorphology of the Widentobel catchment, Apenzell and Sankt Gallen, Switzerland. Ageomorphological inventory system applied to geotechnical appraisal of slope stability. Engineering Geology, 26, 33-68.

Sabto, M., 1991, Probabilistic modeling applied to landslides in Central Colombia using GIS procedures. Unpublished Msc thesis, ITC, Enschede, Netherlands, 26 p.

Saner, S., Taner, İ., Aksoy, Z., Siyako, M. and Burkan, K.A., 1979, Karabük-Safranbolu yöresinin jeolojisi, TPAO rap no: 1322 (unbpublished).

Sauchyn, D.J. and Trench, N.R., 1978, Landsat applied to landslide mapping. Photogrammetric Engineering and Remote Sensing, 44, No: 6, 735-741.

Scott, D.W., 1979, On optimal and data-based histograms. Biometrika, 66, 3, 605-610.

Scott, D.W., 1992, Multivariate density estimation. John Wiley, New York.

Seijmonsbergen, A.C., 1992, Geomorphological evolution of an Alpine area and its application to geotechnical and natural hazard appraisal. Ph.D thesis, University of Amsterdam, 109p.

Serdar, H. and Demir, C., 1983, Mengen-Bolu-Abant dolayı jeolojisi ve petrol olanakları. TPAO raporu. (unpublished).

Sissakian, Kh. V, 1986,The study of mass movement from aerial photographs. Proc. 7 Th Int Symp on Remote Sensing for Resources Development and Environ. 675-678

Soeters, R and van Westen, C.J., 1996, Slope instability recognition analysis and zonation, In Landslides: Investigation and Mitigation (Turner, K.T, and Schuster, R.L. eds), Transportation Research Board National Research Council, Special Report no: 247, Washington D.C., pp 129-177.

Sözen, A., Erbayar, M., Çamaşırcıoğlu, A. ve Çeltek, N., 1996, Düzce (Bolu) - Devrek (Zonguldak) yöresinin genel jeokimya raporu, MTA Rap No: 9922. (unpublished).

Spiegelhalter, D.J., 1986, Uncertainty in expert systems, in artificial intelligence and statistics (Gale, W.A., ed), Addison Wessey, Reading, Mass., pp. 17-55.

Stephens, P.R., 1988, Use of satellite data to map landslides. Proceedings 9th Asian Conference on Remote Sensing, Bangkok, 1-7.

Stevenson, P.C., 1977, An emprical method for the evaluation of relative landslide risk. Bulletin of Int.Assoc. of Engineering Geologists, 16, p.69-72

Strandberg, C.A., 1967, Aerial Discovery manual. Wiley and Sons, New York, 249 p.

Sturges, H.A., 1926, The choice of a class interval. Journal of the American Statistical Association, 21, 65-66.

Sucuoğlu, H.,Bakır, S., Tankut, T., Erberik, A., Gülkan, P., Özcebe, G., Gür, T., Yılmaz, Ç., Ersoy, U., Yılmaz, T. and Akkar, S., 2000, Marmara ve Düzce depremleri mühendislik raporu, Deprem Mühendisliği Araştırma Merkezi, 174 p.

Şaroglu, F., Boray, A. and Emre, O., 1987, Active Faults of Turkey, MTA Rap No: 8643 (Unpublished).

Şaroglu, F., Boray, A. and Emre, O., 1992, Active Faults of Turkey, MTA Publication

Şaroğlu, F., Herece, E., Sarıaslan, M., Emre, Ö., 1995, Yeniçağa-Gerede-Eskipazar arasının jeolojisi ve Kuzey Anadolu fayıın genel özellikler. MTA Rap No: 9873 (Unpublished)

Şengör, A.M.C. and Canıtez, N., 1982, The North Anatolian Fault. In: Alpine Mediterranean Geodynamics, Berekhemer, H. and Hsu, K. (eds). Deuticke, Vienna, 205-216.

Şengör, A.M.C., Görür, N., and Şaroğlu, F., 1985, Strike-slip faulting and related basin formation in zones of tectonic escape; Turkey as a case study, In Strike slip deformation, basin formation and sedimentation, edited by Biddle, K.T. and Blick, N.C., SEPM Spec.Pub., 37.

Şimşek, O. and Dalgıç, S., 1997, Consolidation properties of the clays at Düzce plainand their relationship with geological evaluation, Geological Buletin of Turkey, 40(2), 29-38.

Taymaz, T., 2000, Seismotectonics of Marmara Region: Source characteristics of 1999 Gölcük – Sapanca – Düzce Earthquakes. In 1999 Izmit and Düzce Earthquakes: preliminary results, Barka, A., Kozacı, Ö., Akyüz, S. and Altunel, E. Istanbul University Press. Istanbul. 79-97

Terrill, W.R., 1994, A FAQ on Vegetation in Remote Sensing, http://asio.jde.aca.mmu.ac.uk/new_gis/data/rsvegfaq.htm.

TNT Mips Manual, 2000, Combining Rasters, pp. 4-10.

Tokay, M., 1952, Karadeniz Ereğlisi, Alaplı, Kızıltepe, Alacaağzı bölgesi jeolojisi, MTA derg, 42-43, 1-35.

Tokay, M., 1964, 1/500.000 ölçekli Türkiye Jeoloji Haritası Bolu Paftası, MTA Yayınları

Tokay, M., 1973, Kuzey Anadolu Fay Zonunun Gerede ile Ilgaz arasındaki kısmında jeolojik gözlemler. Kuzey Anadolu Fayı ve deprem kuşağı sempozyumu. 12-29.

Turrini, M.C. and Visintainer, P., 1988, Proposal of a method to define areas of landslide hazard and an application to an area of the dolomites, Italy. Engineering Geology, 50, 255-265

Unterberger, W. and Brandl, J., 2000, The effect of recent earthquake in Turkey on the Bolu Tunnels, Felsbau, 18-2, 50-54.

Ustaömer, P.A. and Rogers, G., 1999, The Bolu Massif: remnant of a pre-Early Ordovicien active margin in the west Pontides, northern Turkey. Geological Magazine, 136, 5, 579-592.

Uysallı, H., 1959, Bolu - Merkeşler bölgesinin jeolojisi ve linyit imkanları. MTA Enstitusü dergisi sayı 52, 107 p.

Unesco, Working Party on World Landslide Inventory, 1993, A suggested method for describing the activity of a landslide. Bull. of the Int Eng Geol, 47, 53-57.

Van Westen, C.J.,1992, Medium scale landslide hazard analysis using a PC based GIS: A case study from chinchina. In Proc., 1er Simposio Internacional sobre Sensores Remotes y Sistemas de Informacion Geografica (SIG) para el Estudo de Riesgos Naturales, Bogota, Colombia (Alzate, J.B. ed), Instituto Geografico Agustin Codazzi, Bogota, vol 2, pp20.

Van Westen, C.J., 1993, Application of geographic information systems to landslide hazard zonation. ITC Publication no: 15. International Institute for Aerospace and Earth Resources Survey, Enschede, The Netherlands, 245 pp.

Vargas, G.C., 1992, Methodologia pour l'etablissement de cartes de sensibilite aux movements de terrain fonde sur l'utulisation d'un couple stereographique SPOT XS/TM. Application a la region de Palz del Rio (Colombie). Proceedings 1st Simposio Internacional sobre sensors remotos y sistemas de informacion geografica (SIG) para el estudio de riesgos naturals, Bogota, Colombia, 201-220.

Varnes, V., 1978, Slope movement types and processes, Landslides Analysis and Control: Natiopnal Academy of Sciences Transportation Research Board Special Report 176, p. 12-33.

Varnes, D.J.,1984, Landslide Hazard Zonation: a review of principles and practice, UNESCO, Natural Hazards, No:3, 61pp

Ward, T.J., Ruh-Ming Li and Simons, D.B., 1982, Mapping landslide hazards in forest watershed. Journal of Geotechnical Engineering Division, Proceedings of The American Society of Civil Engineers, vol 108, No: GT2, 319-324.

Wieczorek, G.F., 1984, Preparing a detailed landslide inventory map for hazard evaluation and reduction, Bulletin of the Association of the Engineering Geologists, vol.21, no:3, pp.337-342.

Wright, R.H., Campbell, R.H., and Nilsen, T.H., 1974, Preparation and use of isopleth maps of landslide deposits, Geology, vol.2 pp 483-485.

Wrigley, N., 1984, Categorical data analysis for geographers and environmental scientist. Longman. Harlow.

Yalçın, H. and Cerit, O., 1991, Bolu masifi örtü kayaçlarında diyajenetik ve çok düşük dereceli metamorfik kil minerallerinin mineralojisi ve jeokimyası. C.Ü. Müh. Fak. Der. Seri A. Yerbilimler C.8,1.

Yılmaz, Y., Gözübol A.M., Tüysüz, O. and Yiğitbaş., 1981, Abant (Bolu)-Dokurcun (Sakarya) arasında Kuzey Anadolu Fay Zonunun kuzey ve güneyinde kalan tektonik birliklerin evrimi, İ.Ü.Yerbilimleri der., 2, 231-261.

Yin, K.L., and Yan, T.Z., 1988, Statistical prediction model for slope instability of metamorphosed rocks. In Proc., Fifth International Symposium in Landslides (C.Bonnard ed), Lausanne, A.A. Balkema, Rotterdam, Netherlands, Vol.2, 1269-1272.

VITA

MEHMET LÜTFİ SÜZEN

suzen@metu.edu.tr
www.metu.edu.tr/~suzen

1972	Born in Ankara, TURKEY
1989 - 1994	B.Sc. in Geological Engineering, Department of Geological Engineering, Middle East Technical University (M.E.T.U.), Ankara, Turkey
1994 - 1996	M.Sc. in Geological Engineering, Department of Geological Engineering, Middle East Technical University (M.E.T.U.), Ankara, Turkey
1996 - 1997	Post Graduate Diploma Course at International Institute for Aerospace and Earth Sciences (I.T.C.), Enschede, The Netherlands.
1997 - 2002	Ph.D in Geological Engineering, Department of Geological Engineering, Middle East Technical University (M.E.T.U.), Ankara, Turkey

Areas of Interest
Geographical Information Systems, Remote Sensing, Natural Hazard Assessment

Others
Stock Markets, Portfolio Management, Scuba Diving, Yatchting, Motorsports