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)

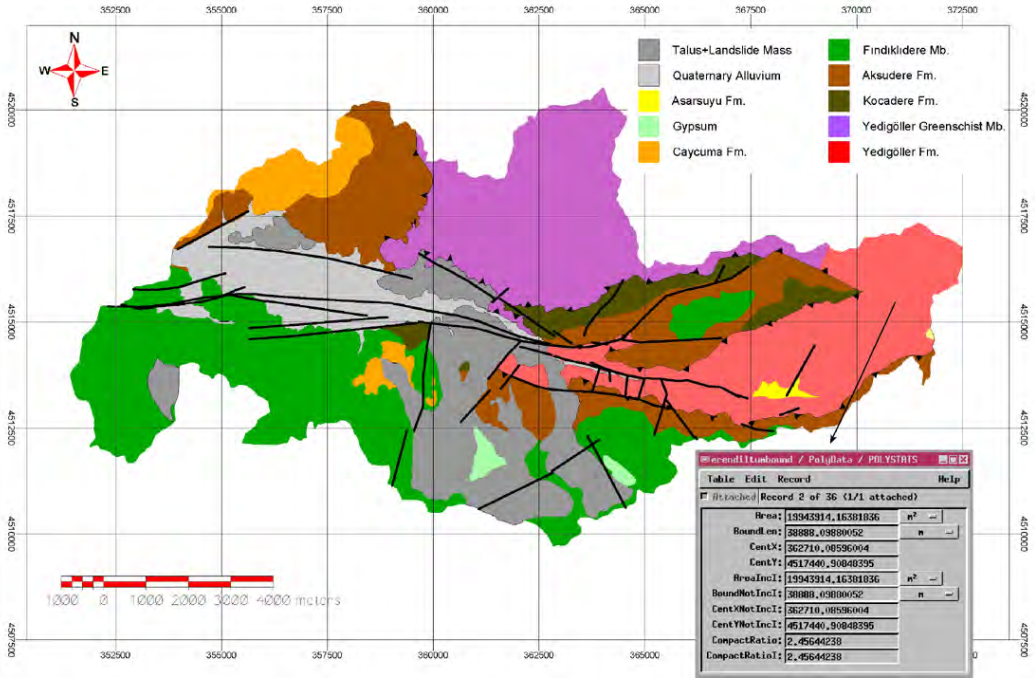

Figure 4.1. LITHOMAP and the attached topological table of the study area.

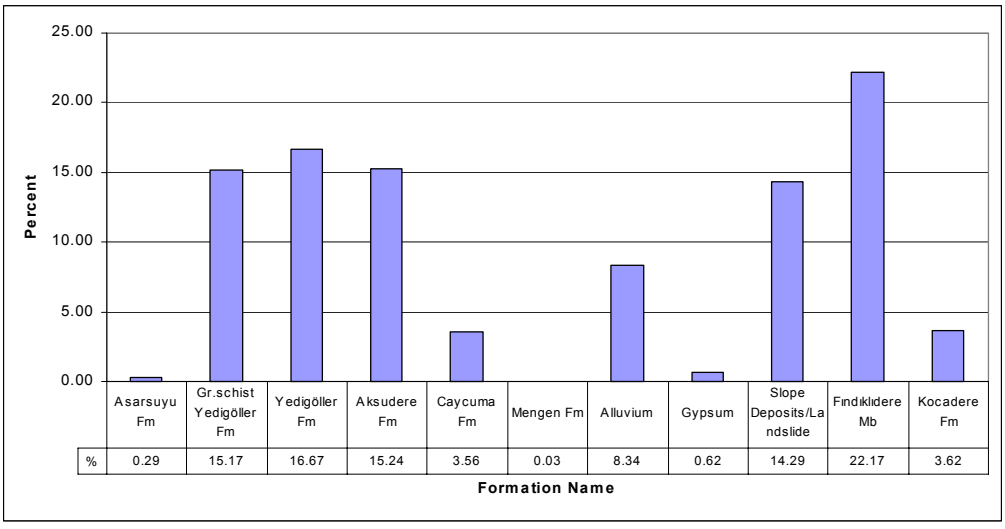

Figure 4.2. The area distributions of LITHOMAP.

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).

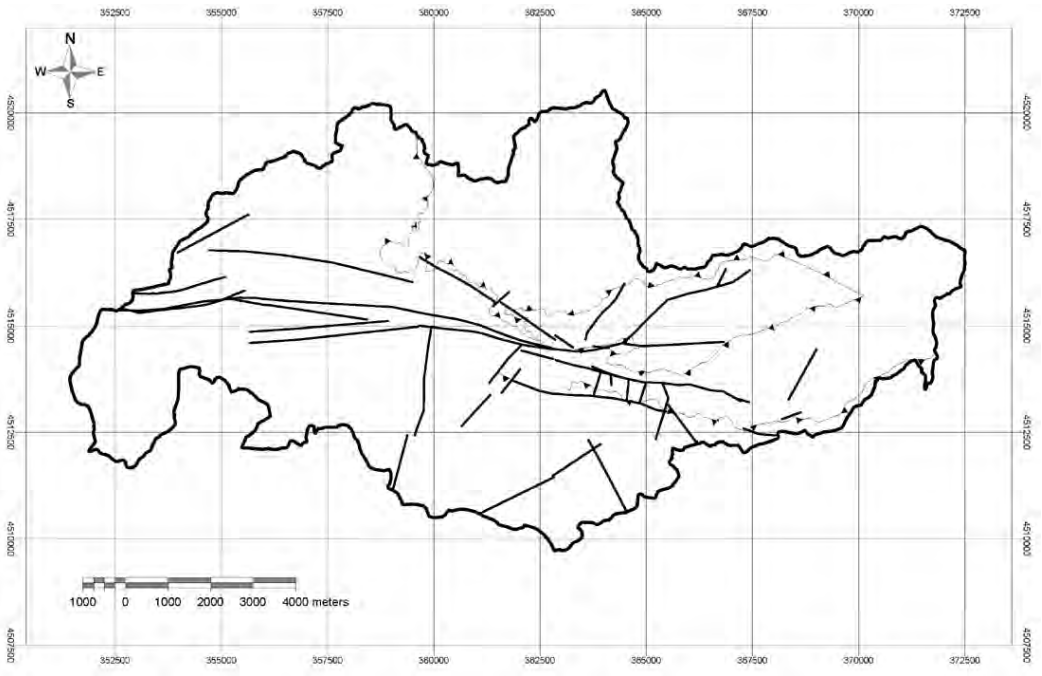

Figure 4.3. The FAULTMAP of Asarsuyu catchment.

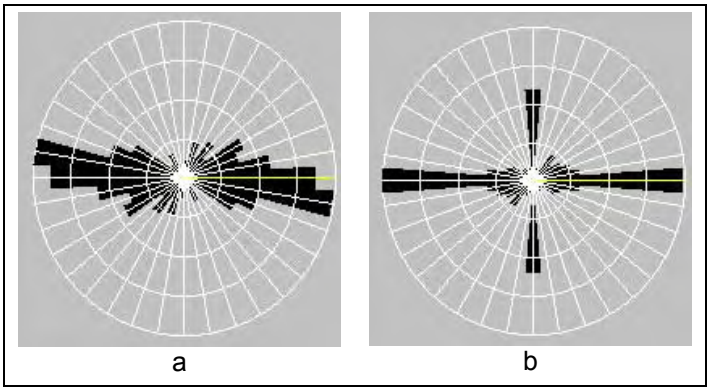

Figure 4.4. The rose diagram of FAULTMAP; a. weighted, b. non-weighted.

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.

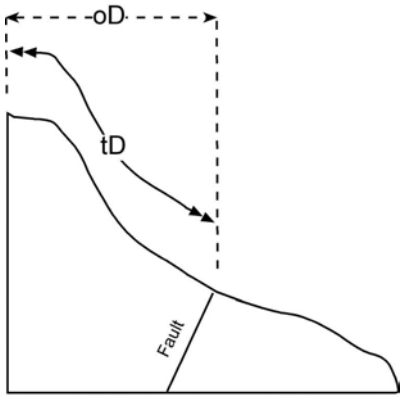

Figure 4.5. The distance calculations in distance raster map production.

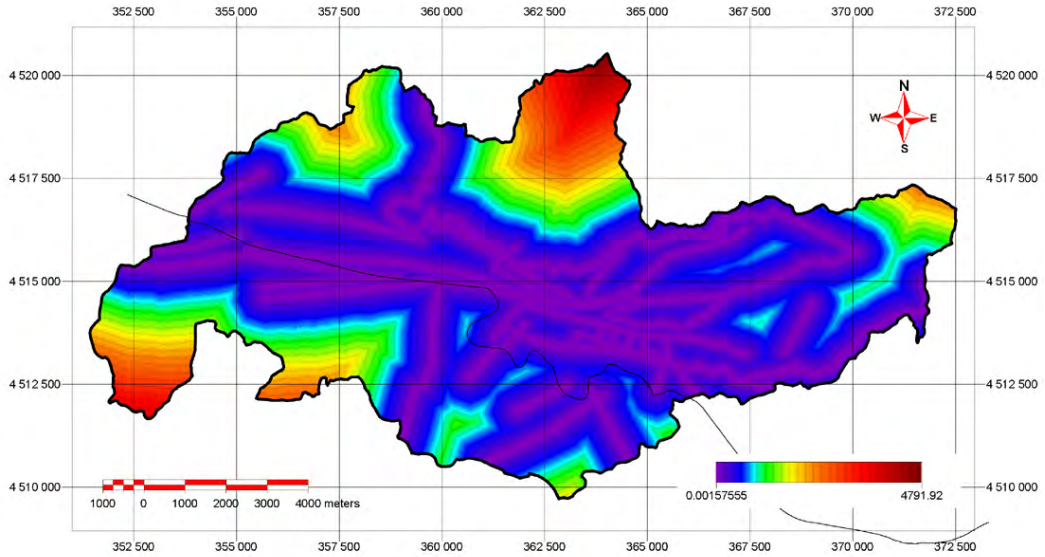

Figure 4.6. Distance raster image (DISTFAULT) produced from FAULTMAP

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

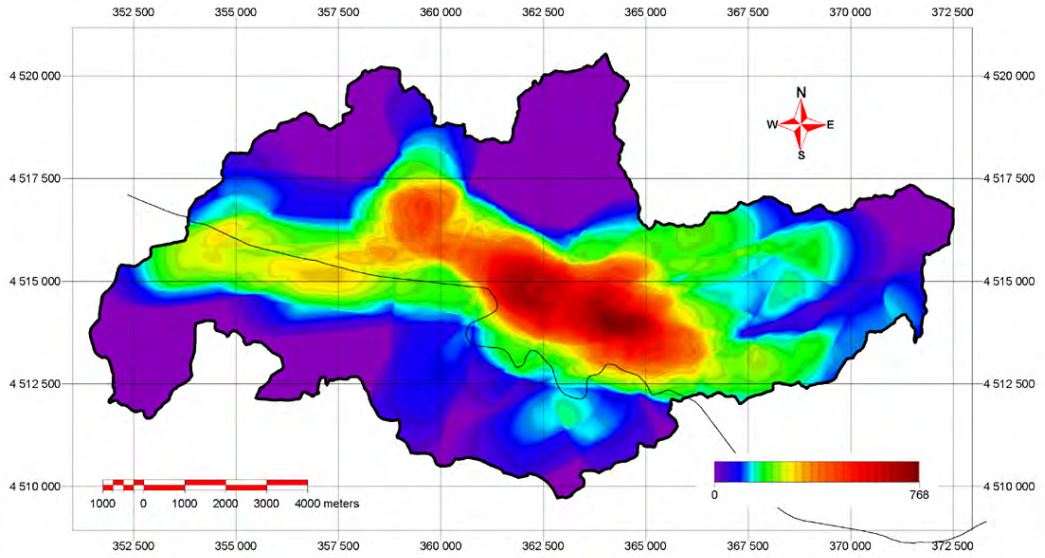

Figure 4.7. Fault density (FAULTDENS) of Asarsuyu Catchment

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.

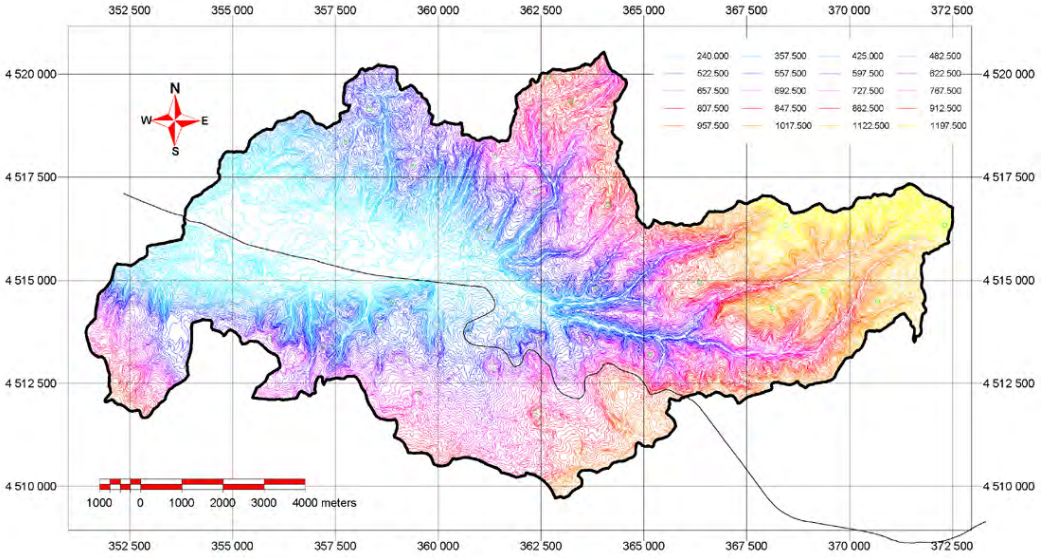

Figure 4.8. The color-coded CONTOURMAP.

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,

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.

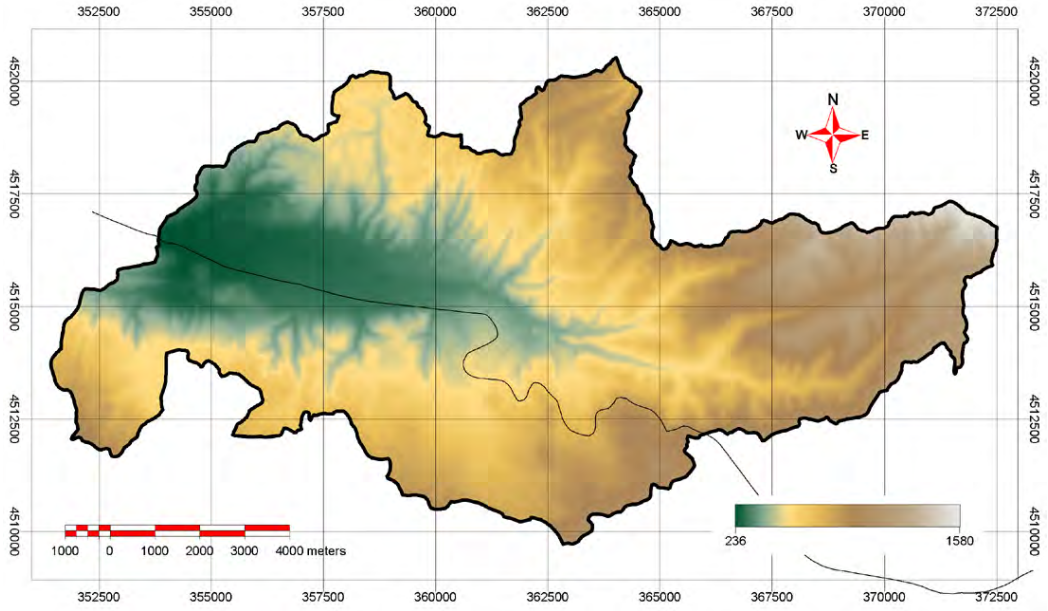

Figure 4.9. Color coded DEM of Asarsuyu catchment

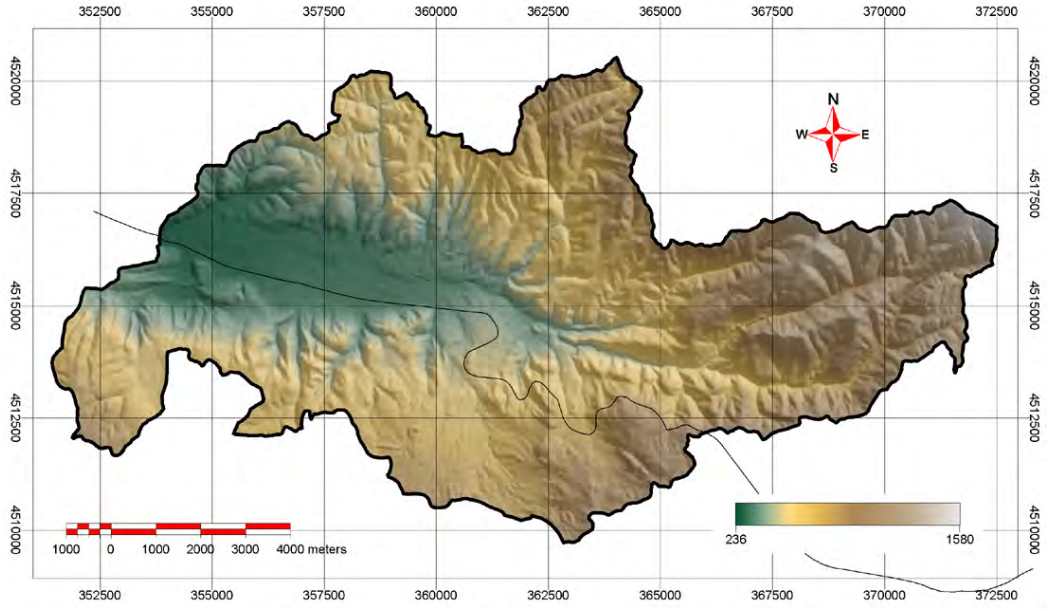

Figure 4.10. Color draped relief model of Asarsuyu catchment (illumination 045^{o}, vertical exaggeration x3).

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.

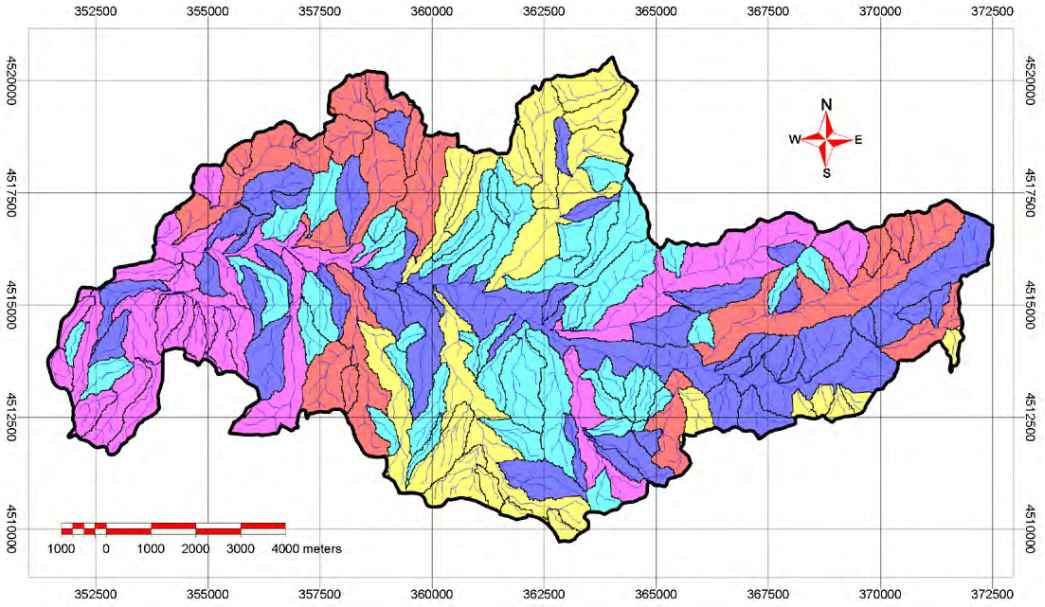

Figure 4.11. The drainage system of Asarsuyu catchment, including the microcatchments (MICROCATCHMAP), drainage-lines (DRAINAGEMAP), and ridges (RIDGEMAP). (Color coding is arbitrary).

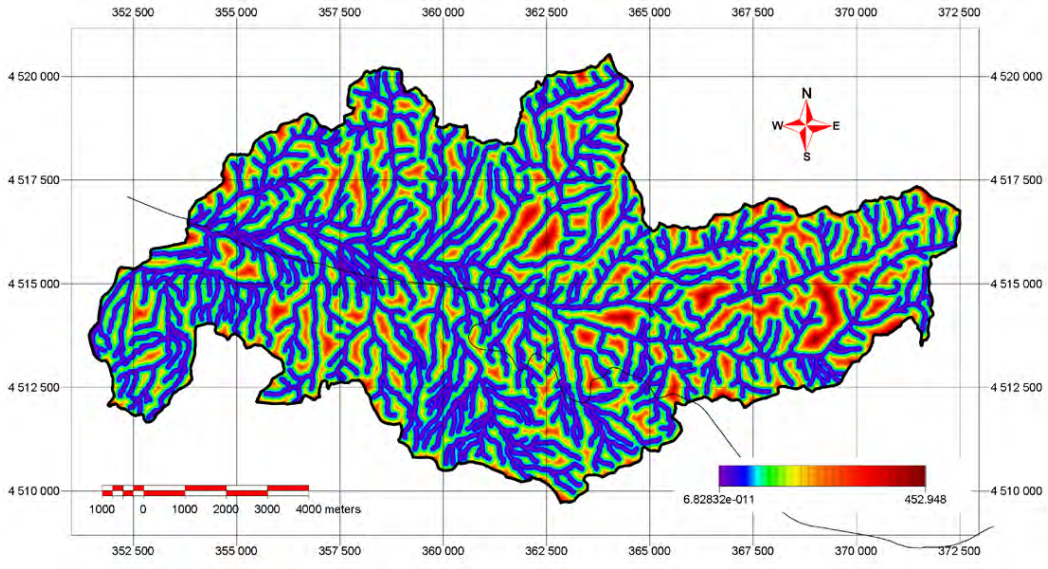

Figure 4.12. The distance raster of every pixel to the nearest drainage-line (DISTDRAINMAP).

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.

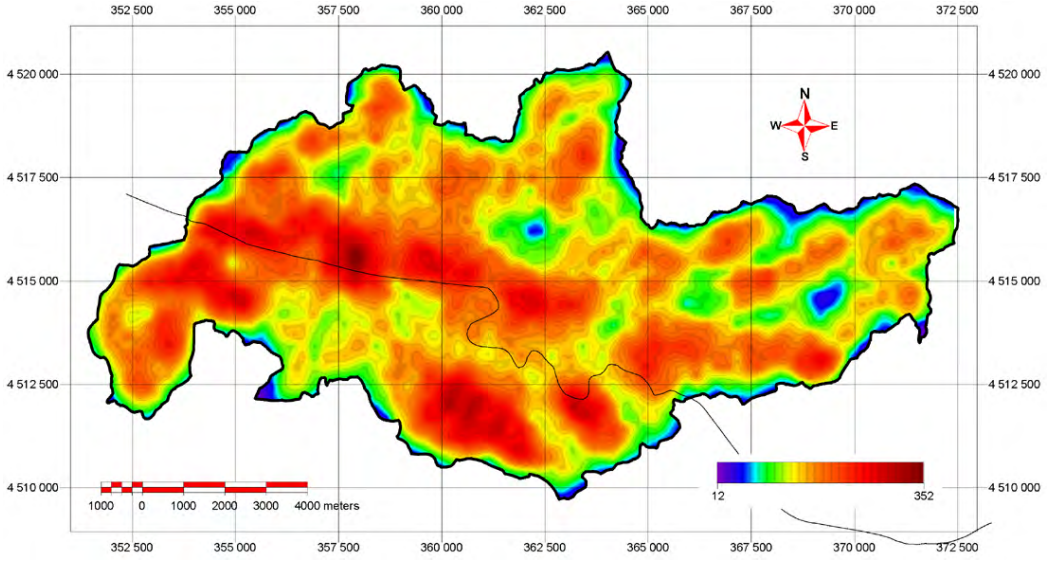

Figure 4.13. The drainage density of Asarsuyu catchment

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

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.

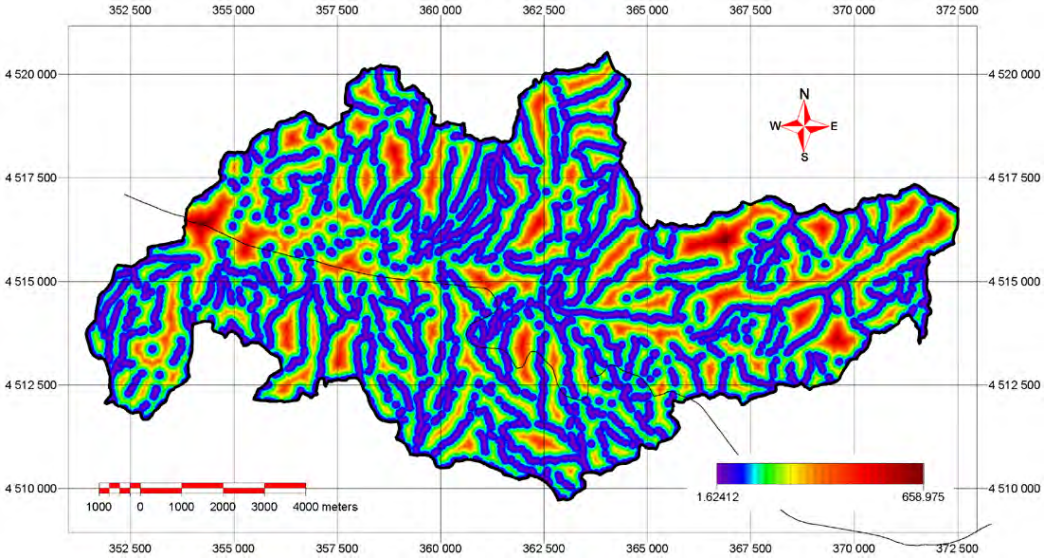

Figure 4.14. The distance raster of every pixel to the nearest ridge-line (DISTRIDGEMAP).

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.

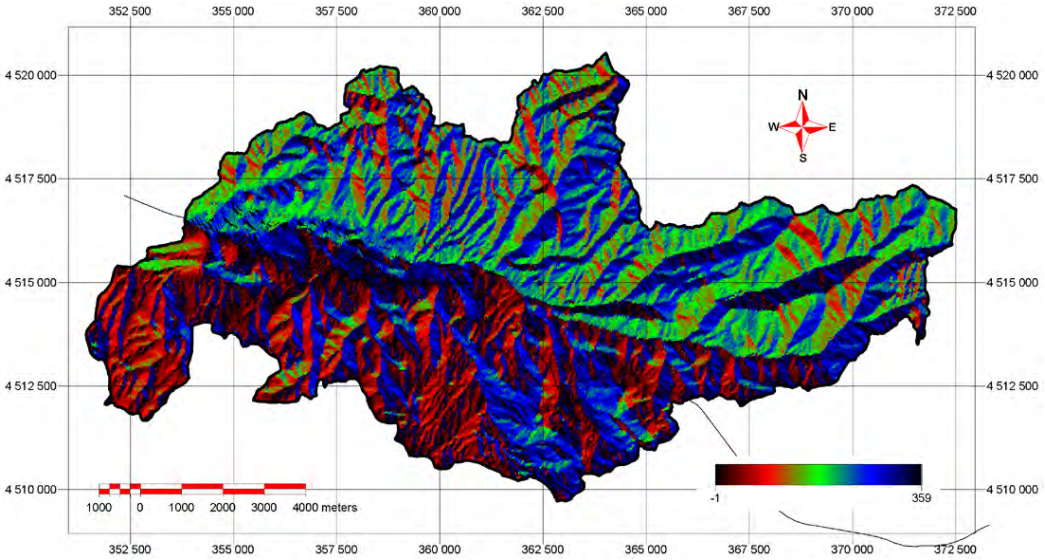

Figure 4.15. Aspect map of the Asarsuyu catchment (ASPECT).

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.

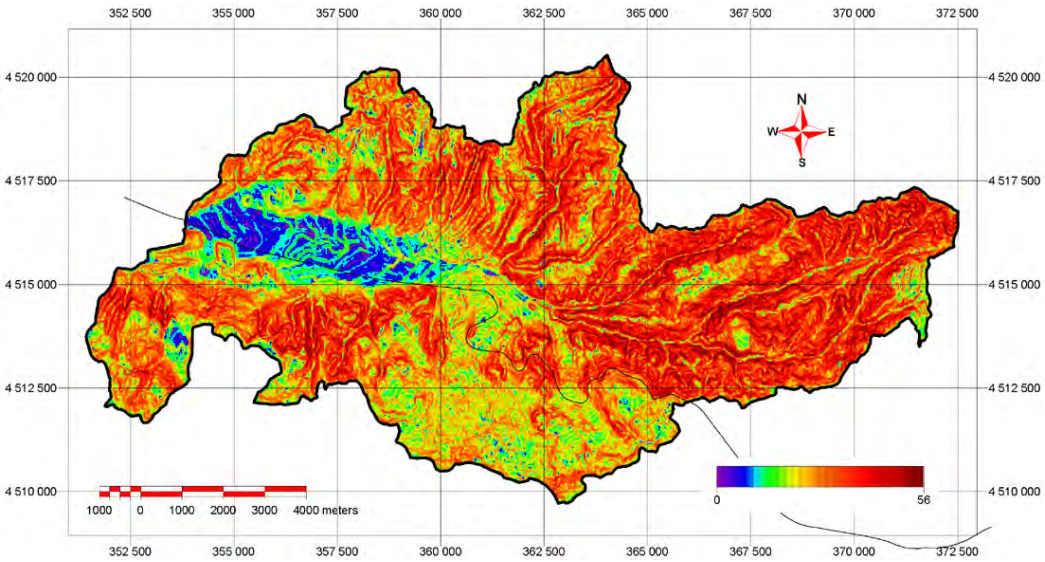

Figure 4.16. SLOPE map of Asarsuyu catchment.

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.

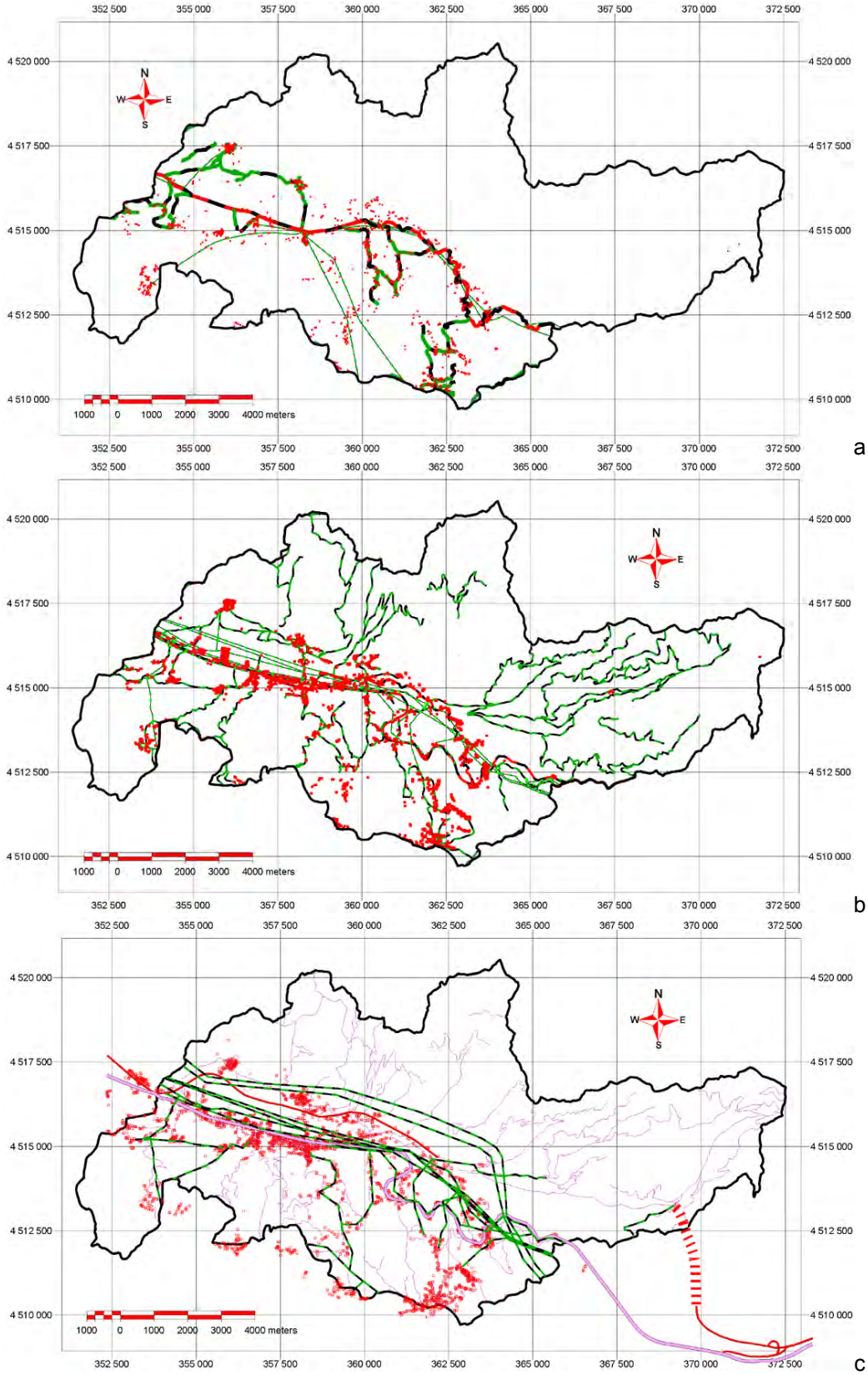

Figure 4.17. Historical infrastructure databases of Asarsuyu catchment a) 1952, b) 1972, c) 1994

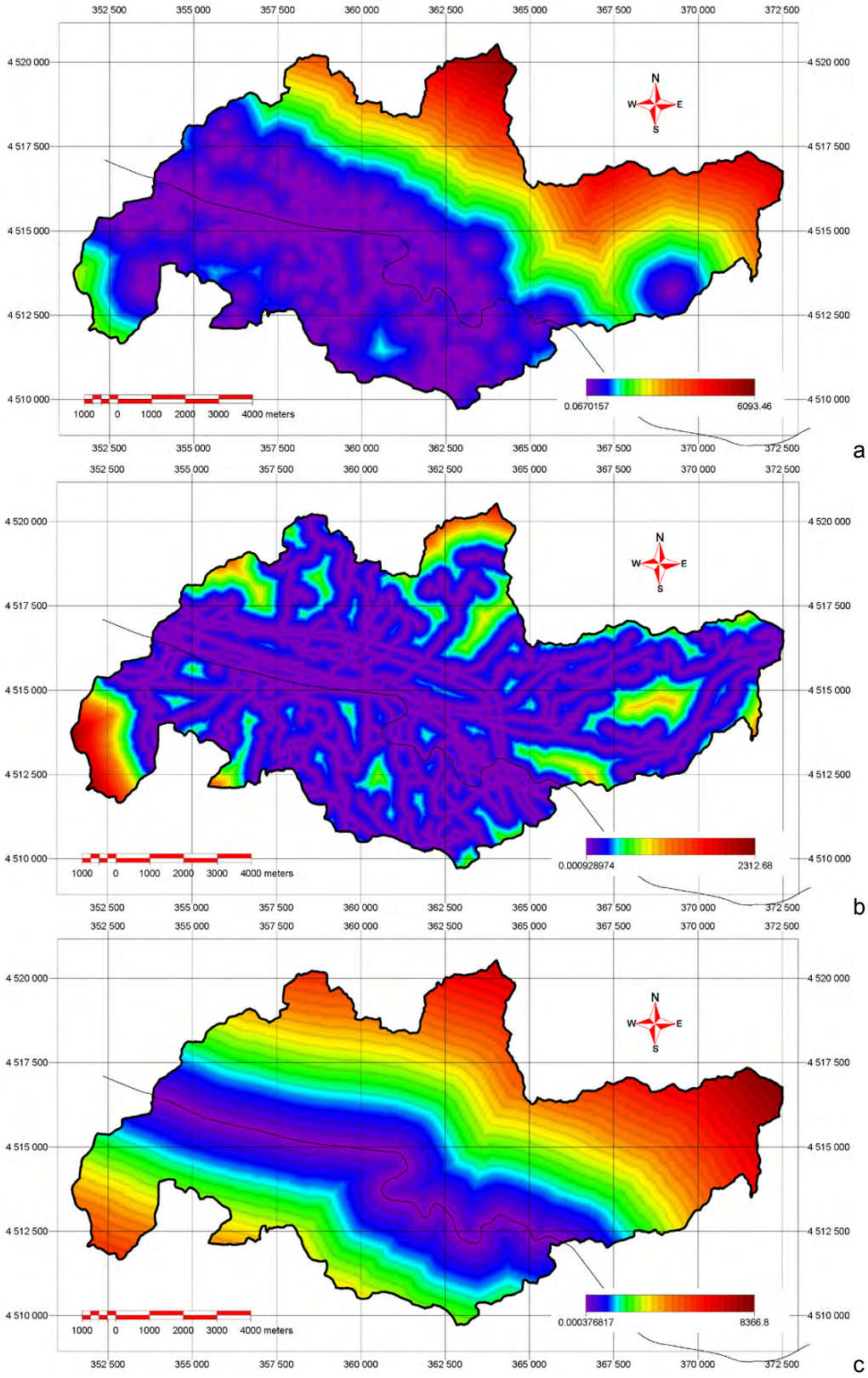

Figure 4.18. Distance rasters and frequency distributions of 1994 period a) distance raster of settlement of 1994, b) distance raster of road network and power lines of 1994, c) distance raster of E-5 road of 1994

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.

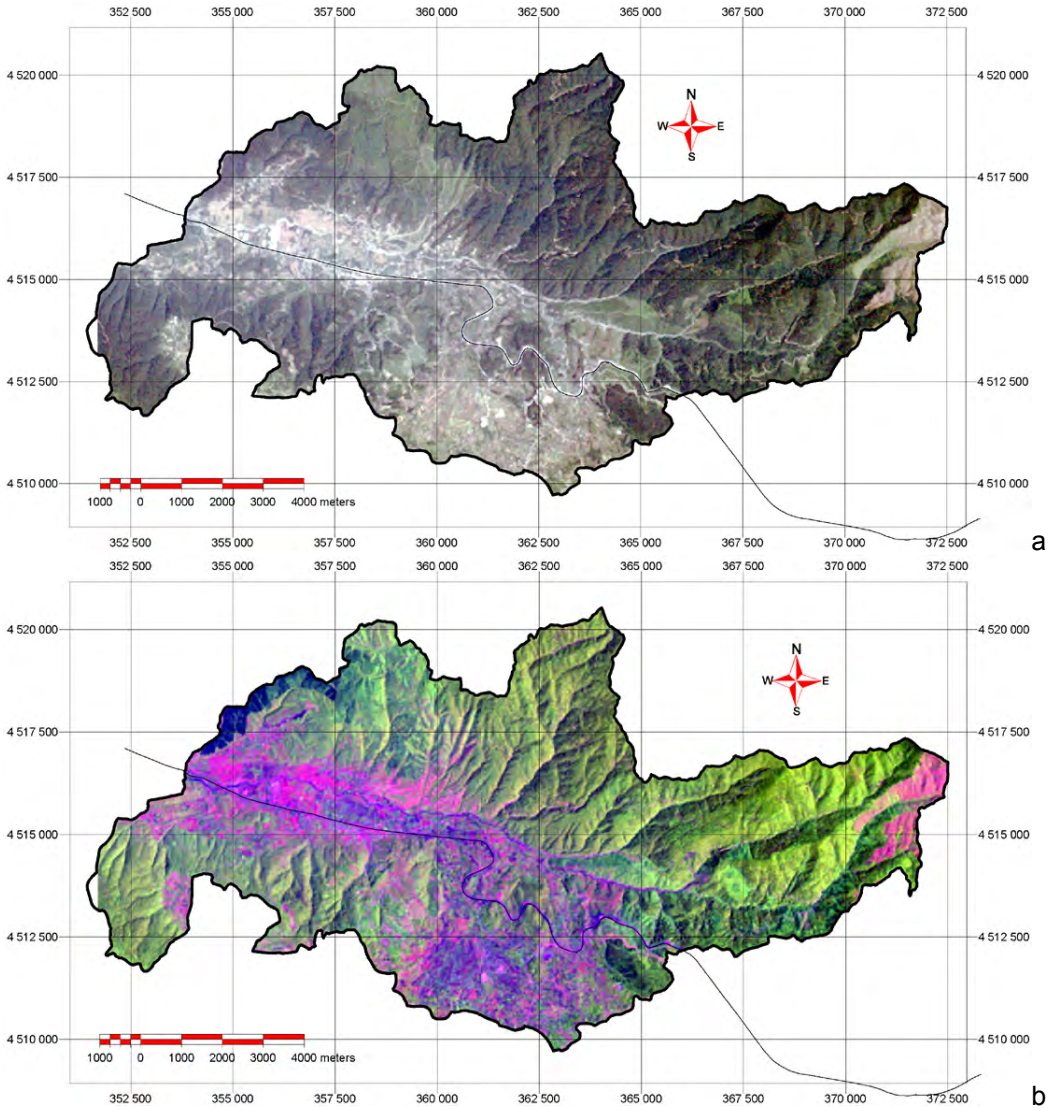

Figure 4.19. a) True color composite of Landsat TM 5 (R=3, G=2, B=1), b) false color composite of Landsat TM 5 (R=5, G=4, B=1).

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).

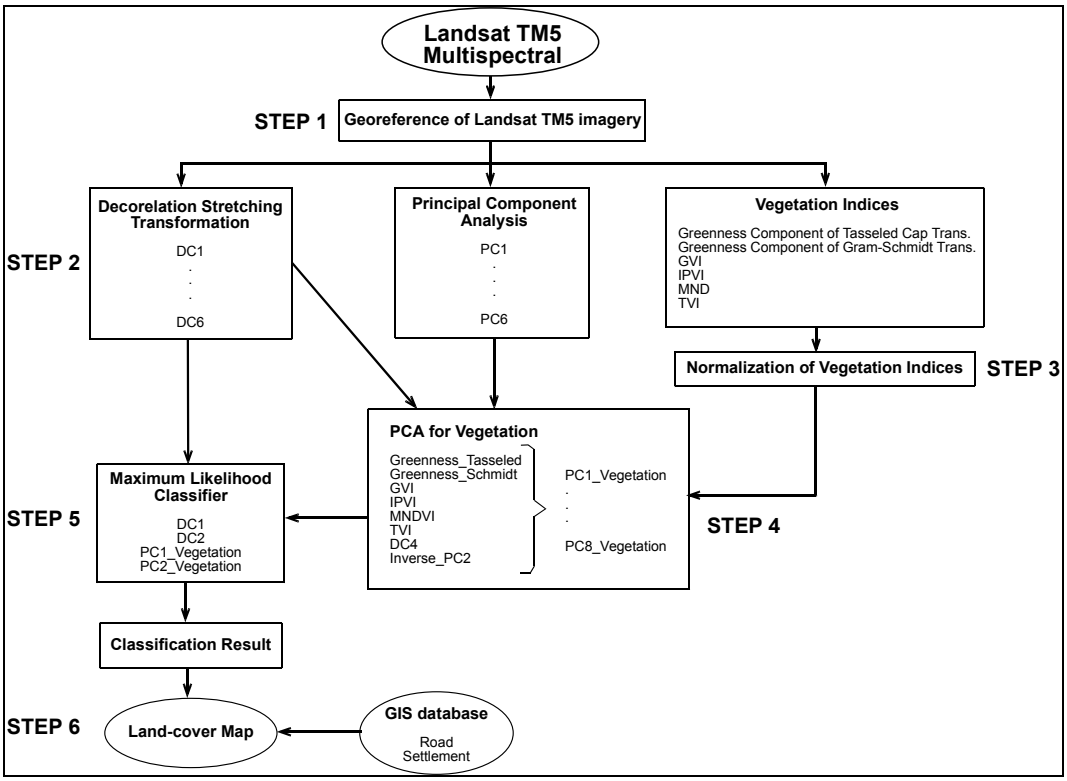

Figure 4.20. The methodological snapshot of land cover extraction scheme

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.

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

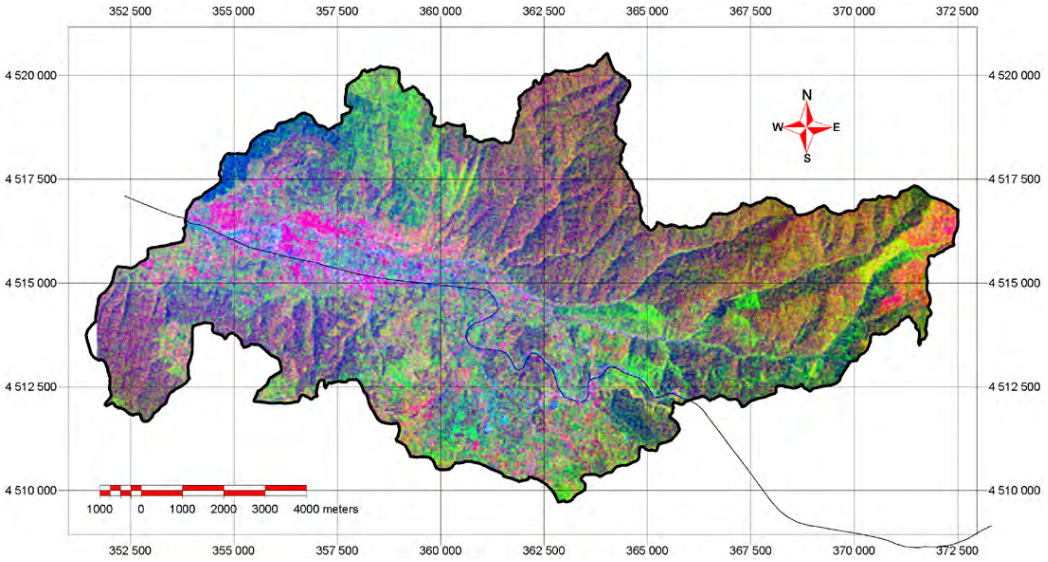

Figure 4.21. Decorrelation stretching results (R: decor_4, G: decor_3:B: decor_1).

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

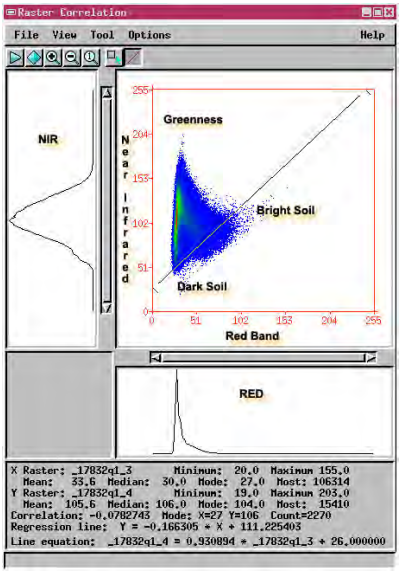

Figure 4.22. Near Infra Red versus Red band Raster Correlation Graph

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).

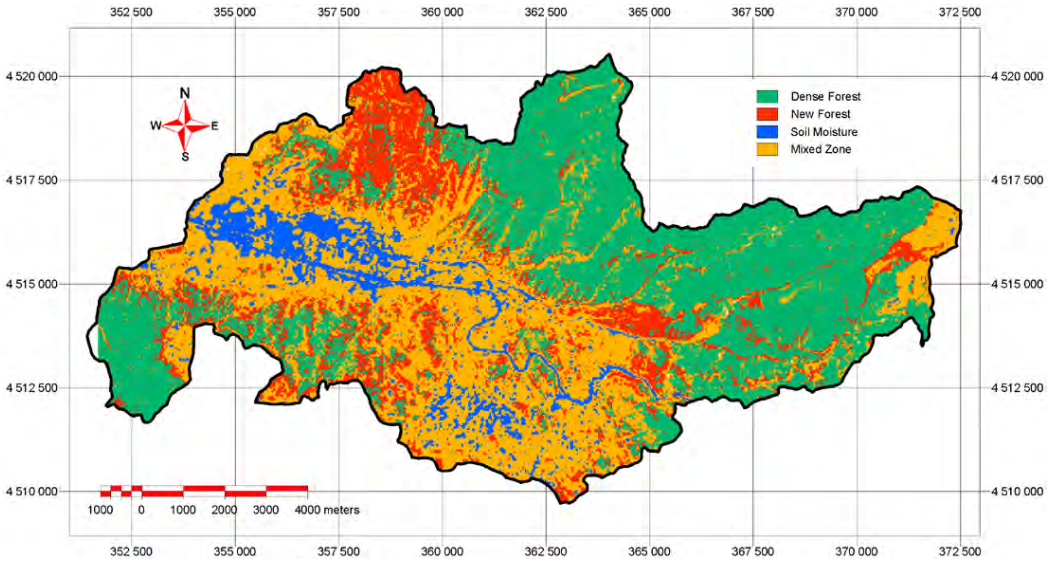

Figure 4.23. Product of Maximum Likelihood Classification

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

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

		Ground Truth Data					
Classif	Name	Forest	Y_Forest	Moisture	Mixed	Total	Accuracy
	Forest	62	0	0	0	62	100.00%
	Y_Forest	2	22	0	1	25	88.00%
	Moisture	0	0	49	7	56	87.50%
	Mixed	6	0	1	76	83	91.57%
	Total	70	22	50	84	226	
	Accuracy	88.57%	100.00%	98.00%	90.48%		
Overall Accuracy = 92.48% Khat Statistic = 89.45%							

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.

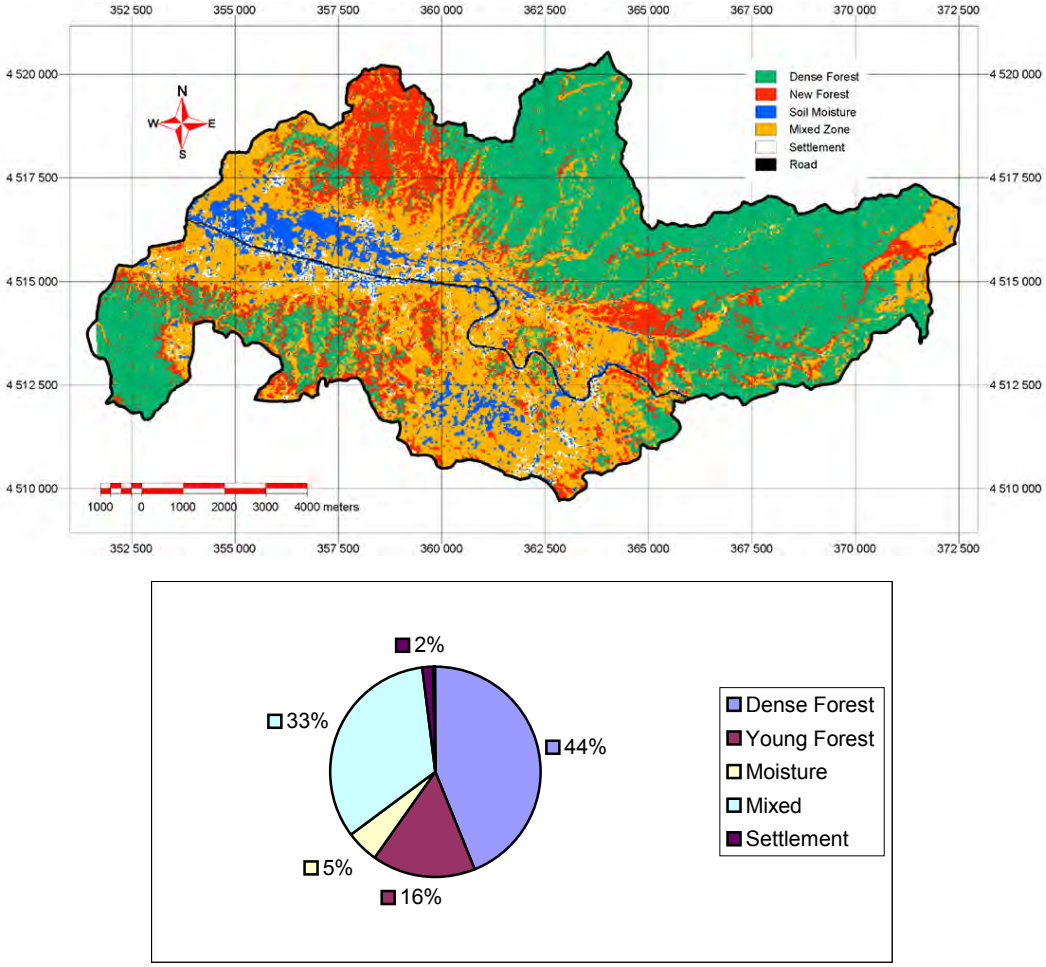

Figure 4.24.a) Land-cover map of the study area, b) areal distributions of land cover.

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".

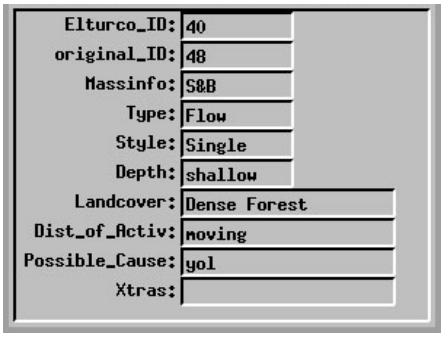

Figure 4.25. The landslide attribute database

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.

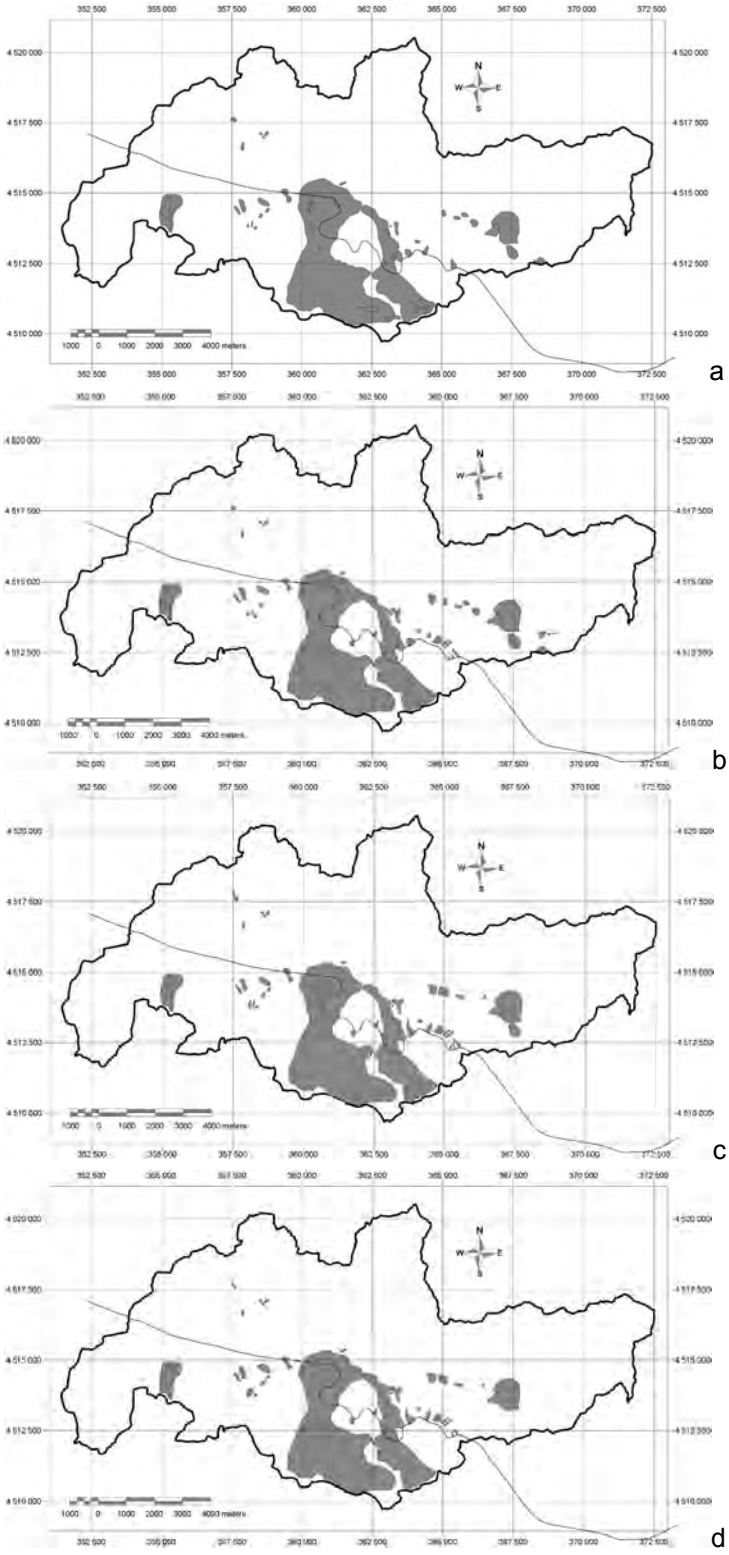

Figure 4.26. Landslide inventories of the four time periods, a. 1952, b. 1972, c. 1984, d. 1994

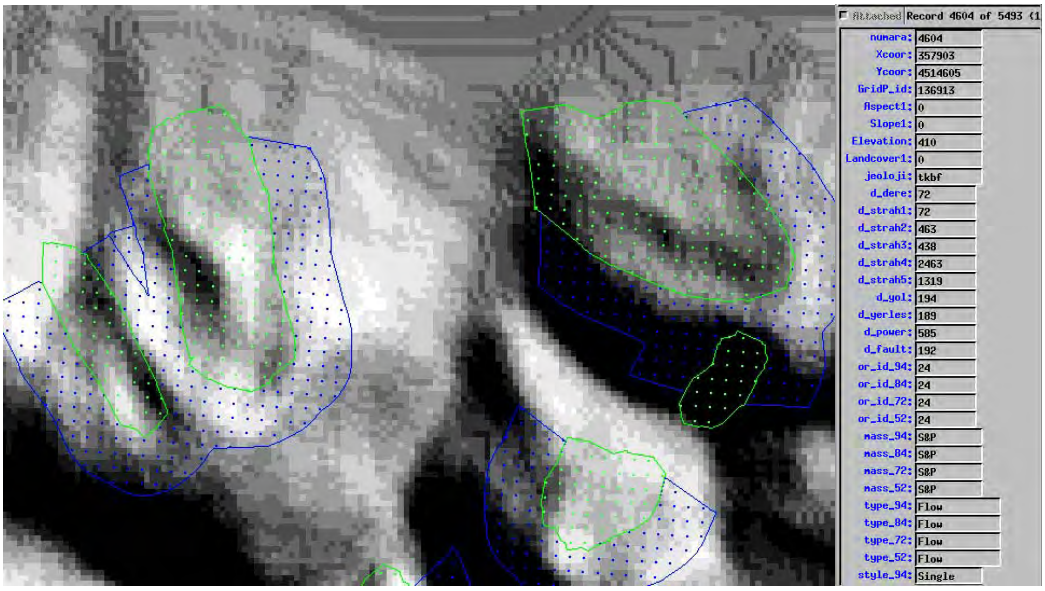

Figure 4.27. Polygon grid midpoints and transferred attributes (white points are merged polygons, black ones are buffered zone)