CHAPTER 6

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).

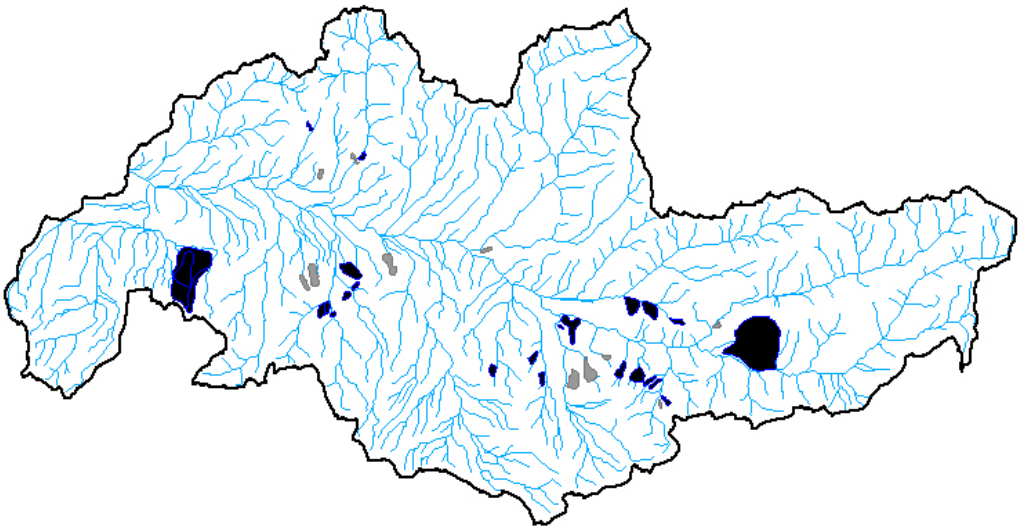


Figure 6.1. The Thematic Landslide Attribute spatial distribution of massinfo attribute of 1994 photo characteristics database, where black areas show Scarp & Body and grey areas show Scarp & Path attributes.

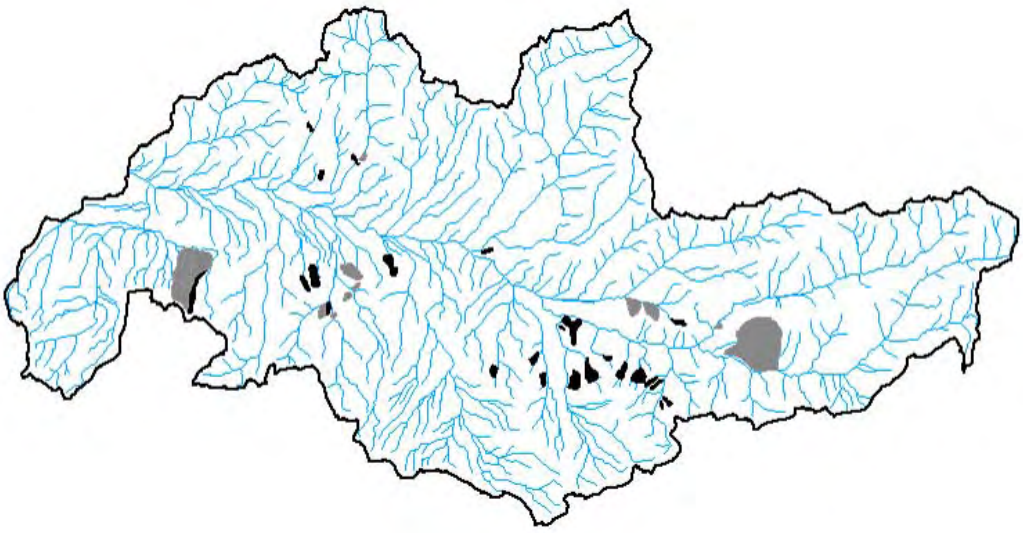


Figure 6.2. The Thematic Landslide Attribute spatial distribution of type attribute of 1994 photo characteristics database, where black areas show Flow and grey areas show Slide attributes.

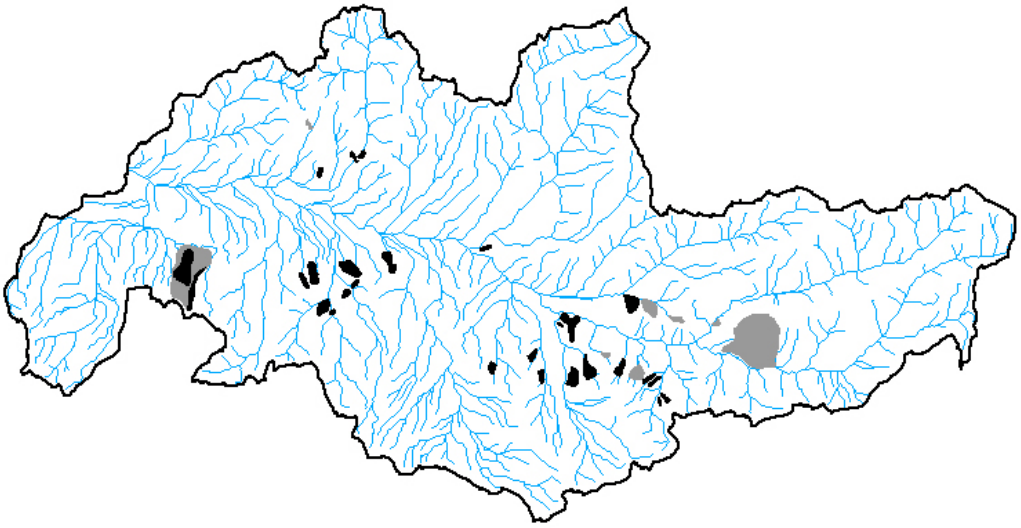


Figure 6.3. The Thematic Landslide Attribute spatial distribution of style attribute of 1994 photo characteristics database, where black areas show Single and grey areas show Multiple attributes.

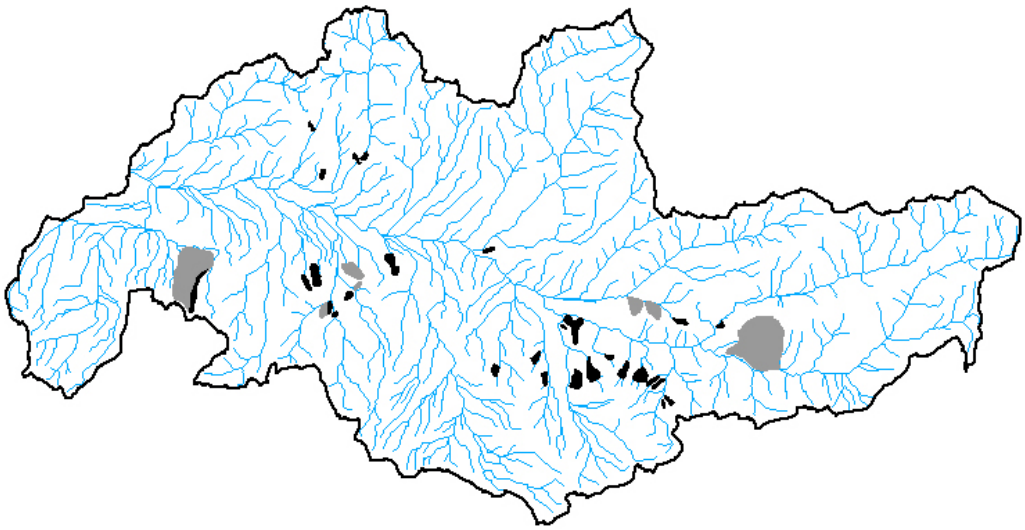


Figure 6.4. The Thematic Landslide Attribute spatial distribution of depth attribute of 1994 photo characteristics database, where black areas show Shallow and grey areas show Deep attributes.

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).

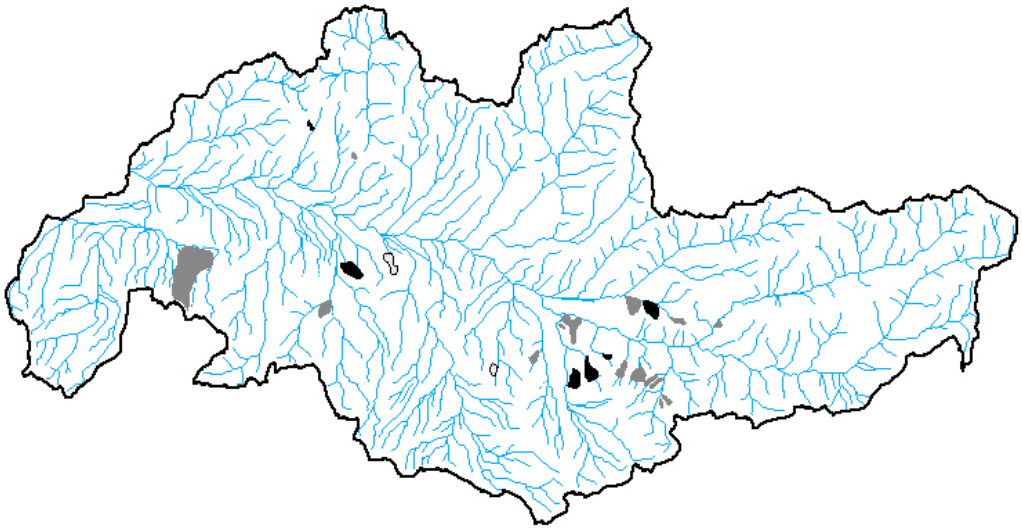


Figure 6.5. The Thematic Landslide Attribute spatial distribution of activity attribute of 1994 photo characteristics database, where black areas show Scarp, grey areas show Mass attributes, white polygons with black border show stable attribute. Polygons which do not have any attribute in this field 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 relicts. 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							
Activity in 1972 photo set		Scarp	Body	Present	Absent	Dormant	Σ
	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							
Activity in 1984 photo set		Scarp	Body	Present	Absent	Dormant	Σ
	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							
Activity in 1994 photo set		Scarp	Body	Present	Absent	Dormant	Σ
	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.

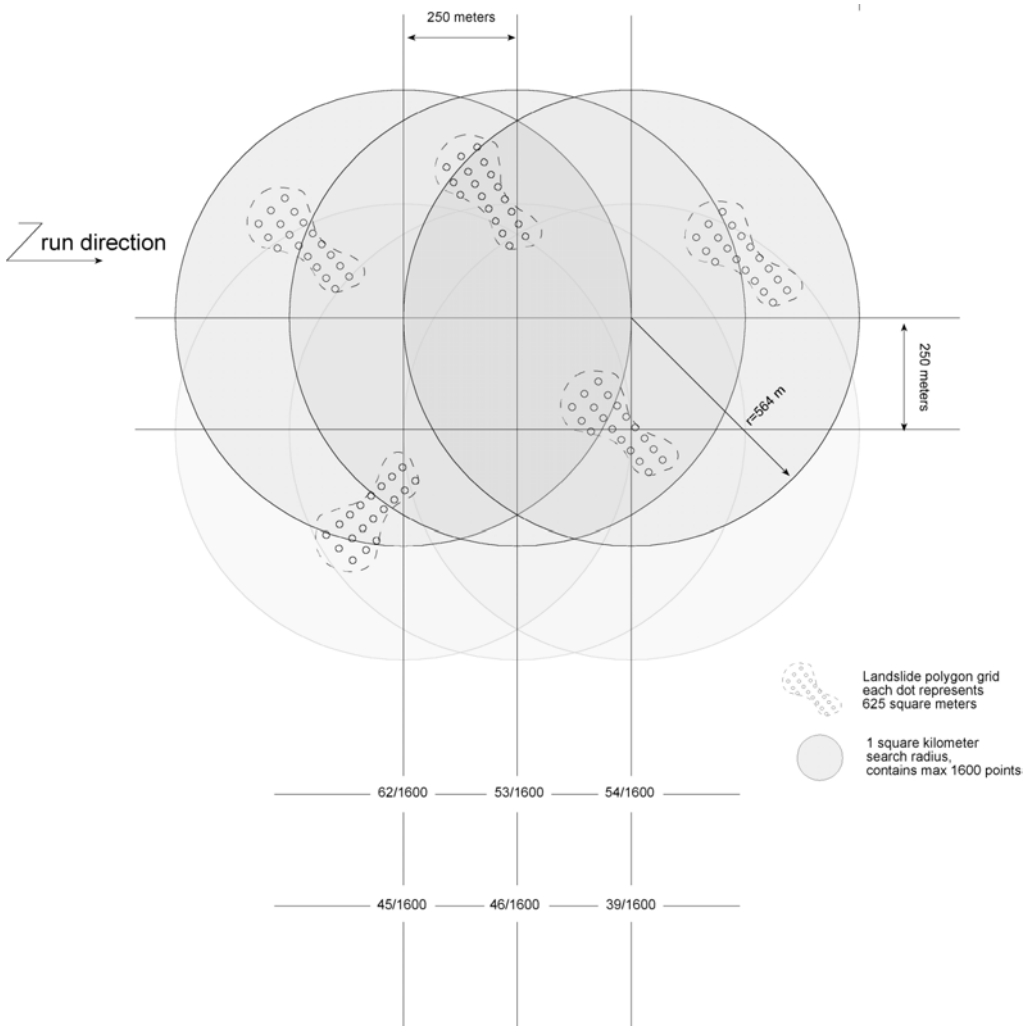


Figure 6.6. The mechanism of the isopleth analysis

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

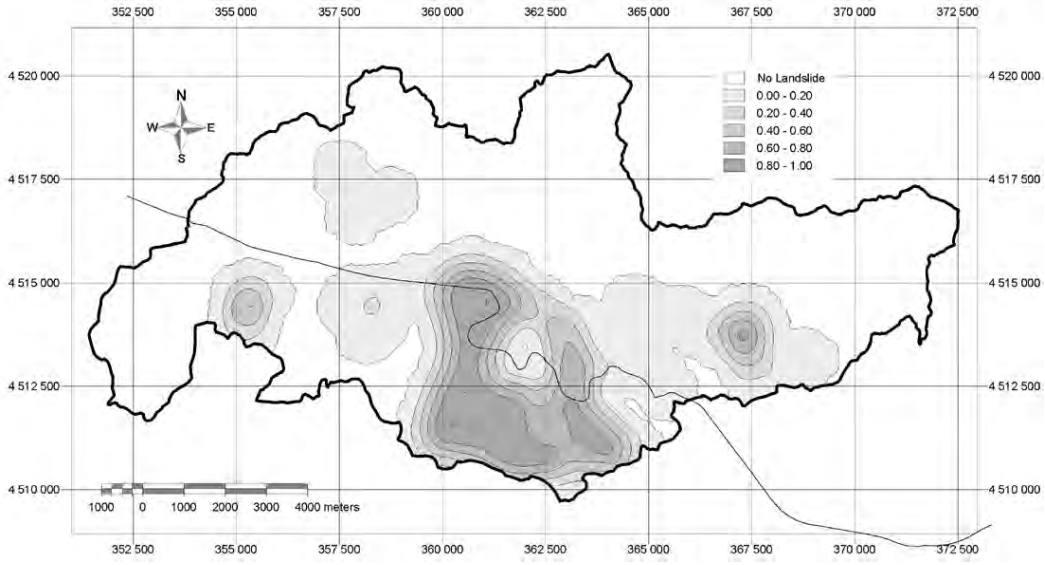


Figure 6.7. The isopleth map of the Asarsuyu catchment, each contour level represents 20 % landslide density.

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:

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

Where W is the width of the histogram bin, IQR is the inter quartile range (the 75^{th} percentile minus the 25^{th} 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

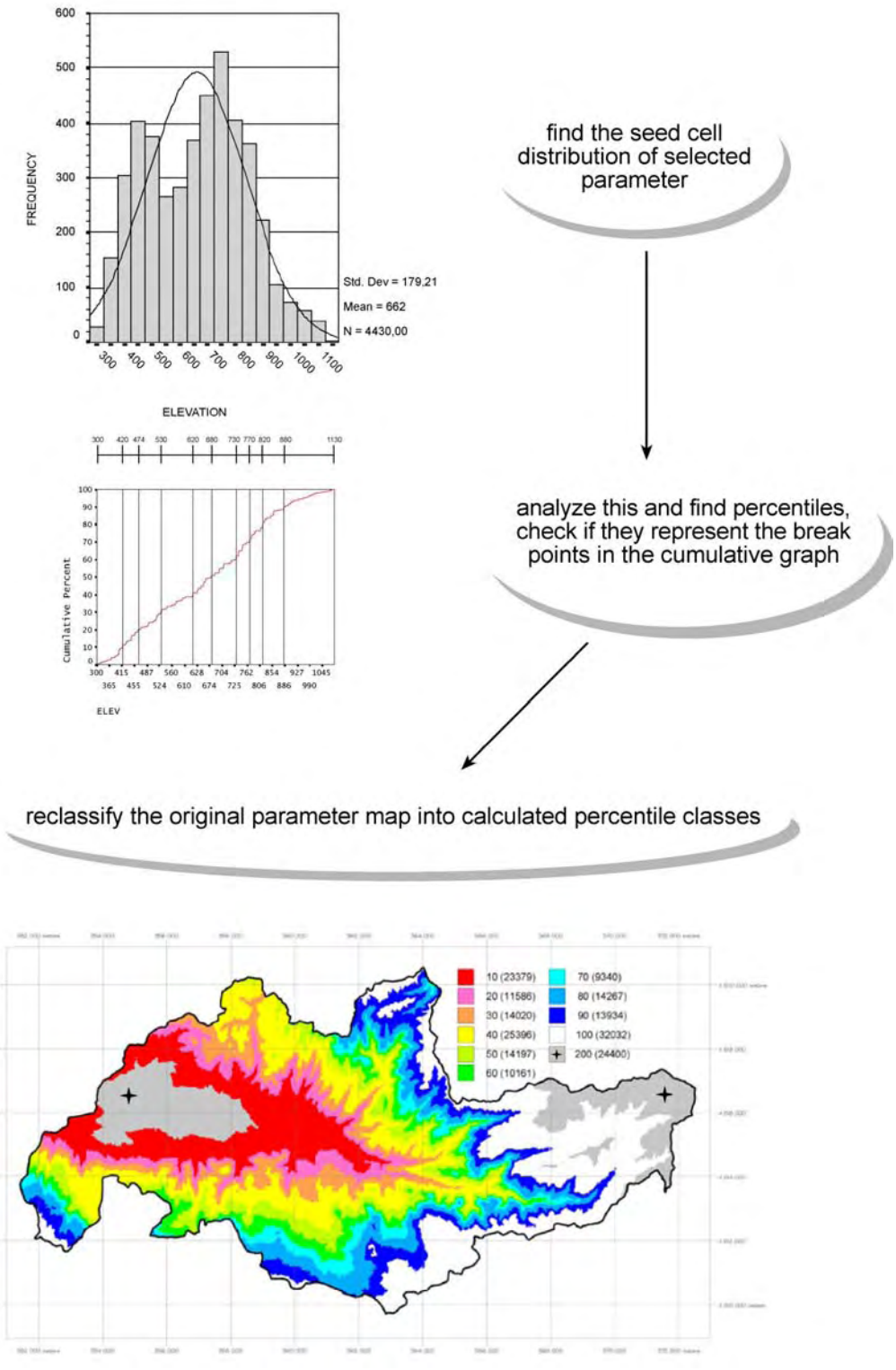


Figure 6.8. The snapshot of methodology of percentile method and reclassified parameter map production. The areas with a star are the nature's own decision rule and not taken into calculations as there is no landslide in these area.

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**

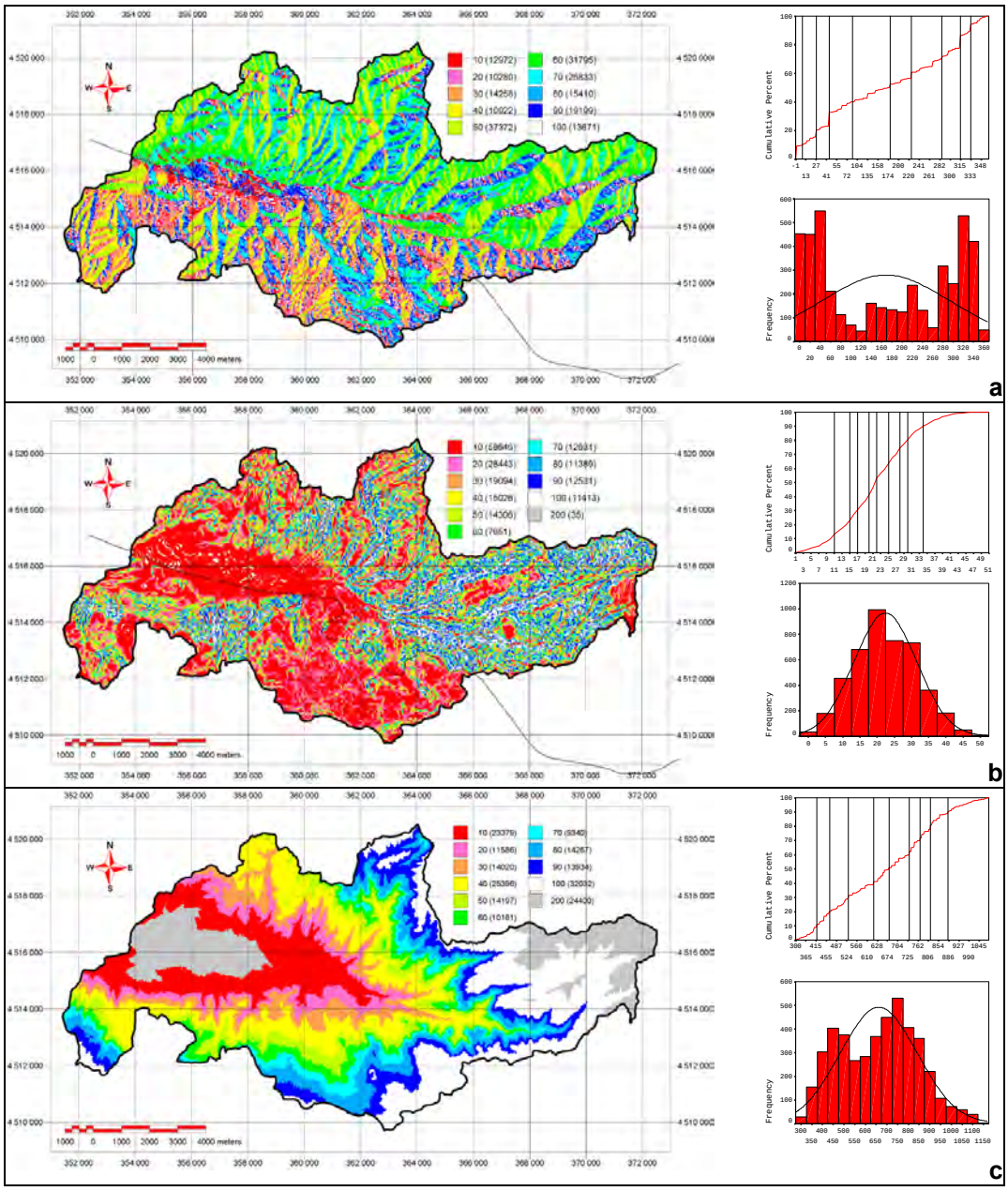


Figure 6.9. The percentile maps of morphology of Asarsuyu catchment, with frequency and cumulative histograms, a) Aspect, b) Slope, c) Elevation

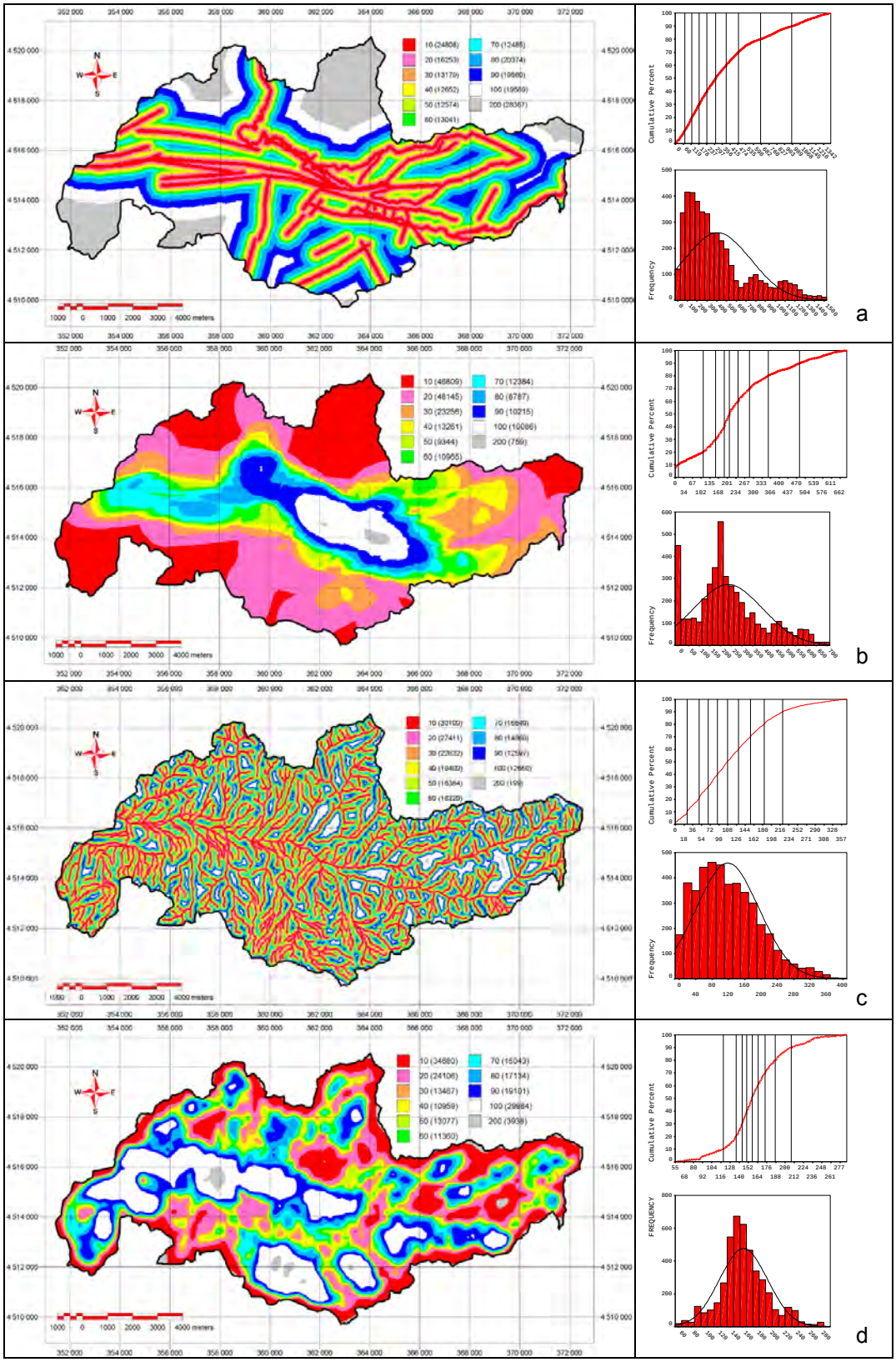


Figure 6.10. The percentile maps of lineament and density patterns of Asarsuyu catchment, with frequency and cumulative histograms, a) Distance to fault , b) Fault density, c) Distance to Drainage d) Drainage density

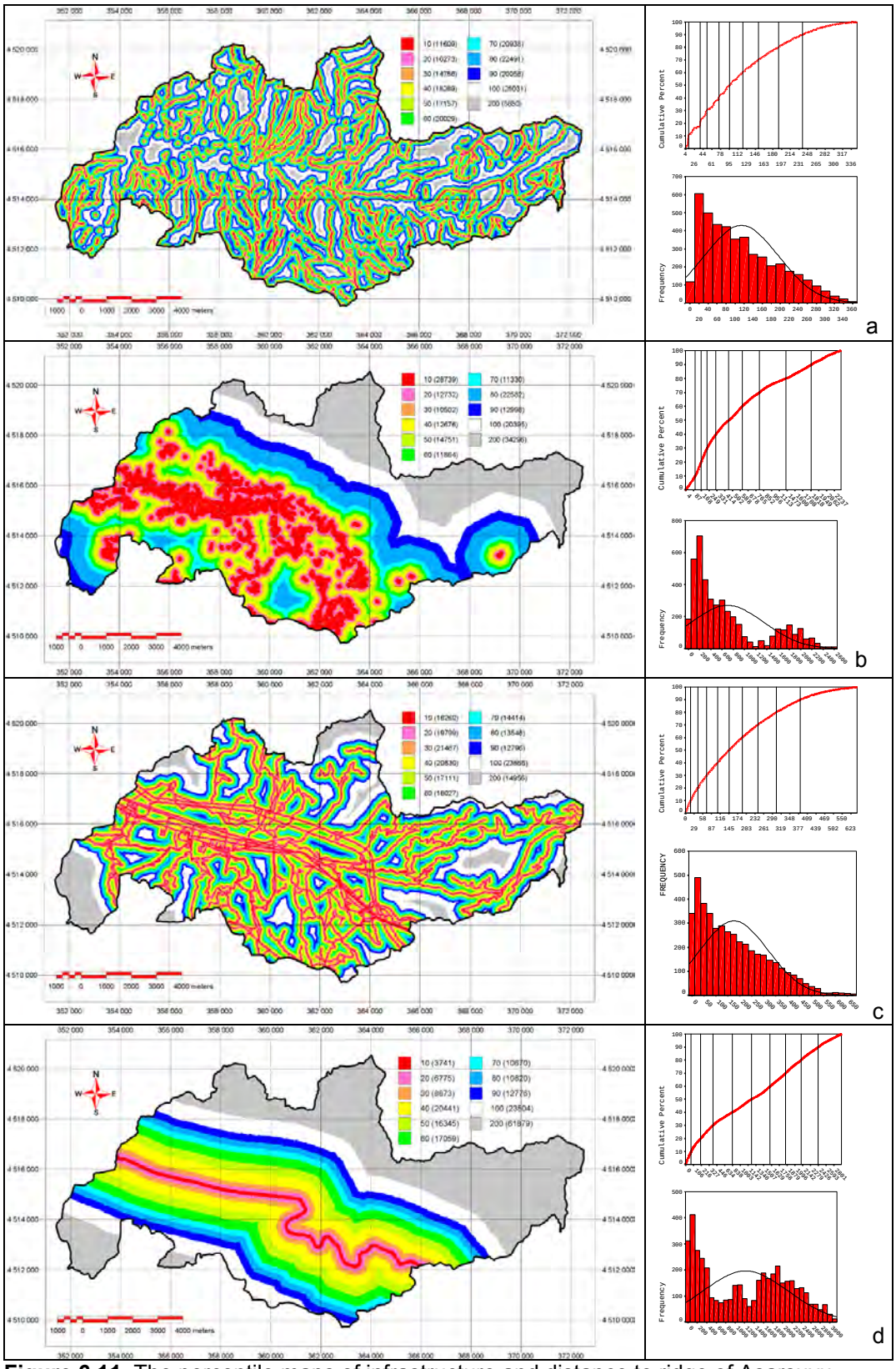


Figure 6.11. The percentile maps of infrastructure and distance to ridge of Asarsuyu catchment, with frequency and cumulative histograms, a) Distance to ridge , b) Distance to Settlement, c) Distance to Power and road network d) Distance to E-5 Highway.

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

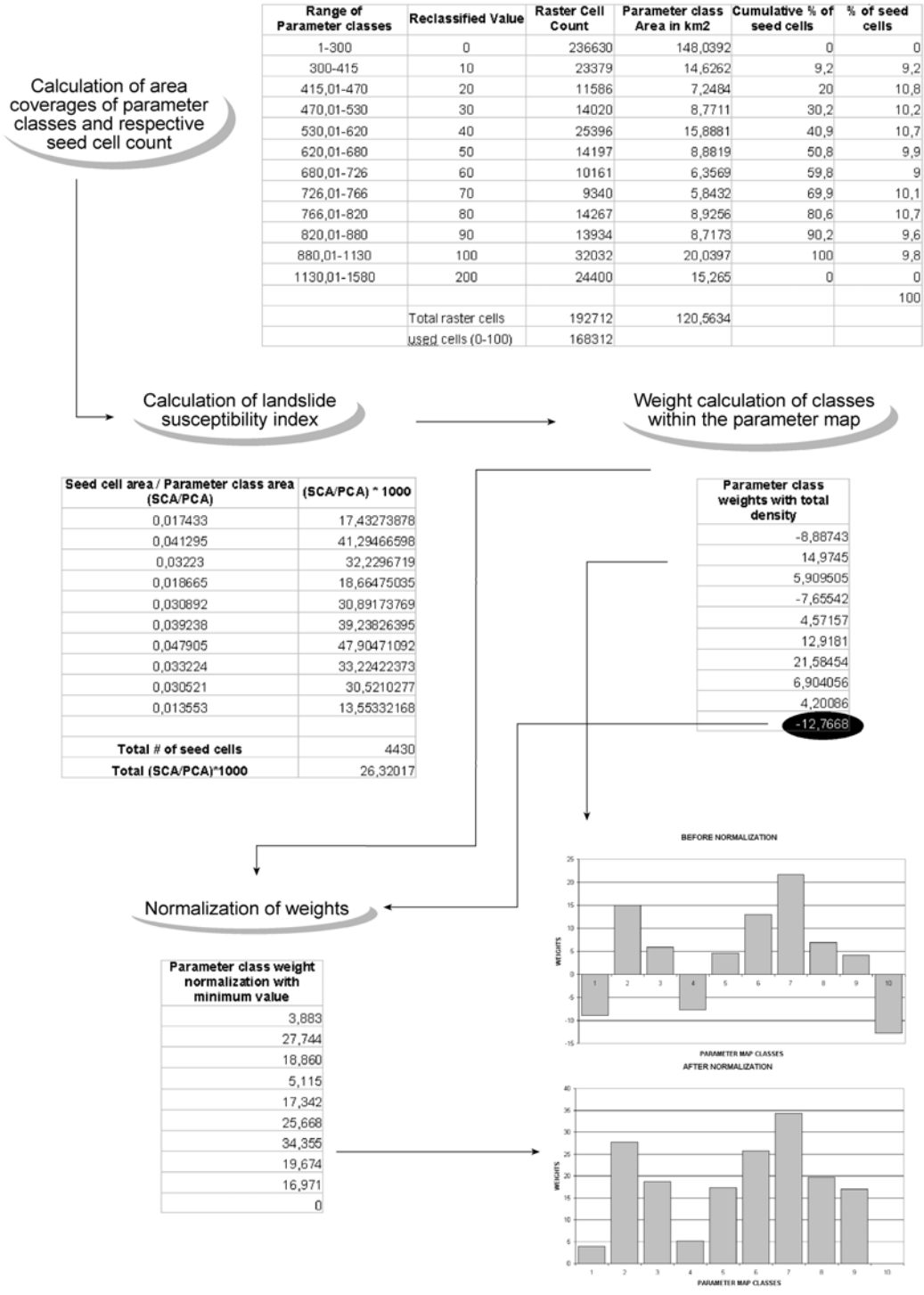
LANDSLIDE SUSCEPTIBILTY	INFORMATION VALUE
$$D_{area} = 1000\frac{Npix(SX_i)}{Npix(X_i)}$$ in which D_{area} = Areal density per millage $Npix(SX_i)$ = number of pixels with mass movements within variable class X_i. $Npix(X_i)$ = number of pixels within variable class X_i. To evaluate the influence of 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: $$W_{area} = 1000\frac{Npix(SX_i)}{Npix(X_i)} - 1000\frac{\sum Npix(SX_i)}{\sum Npix(X_i)}$$	$$I_i = \log\frac{S_i/N_i}{S/N}$$ 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.



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			

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

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

Figure 6.12 The steps through landslide susceptibility analysis

Table 6.7. Weight values of the all available parameter classes

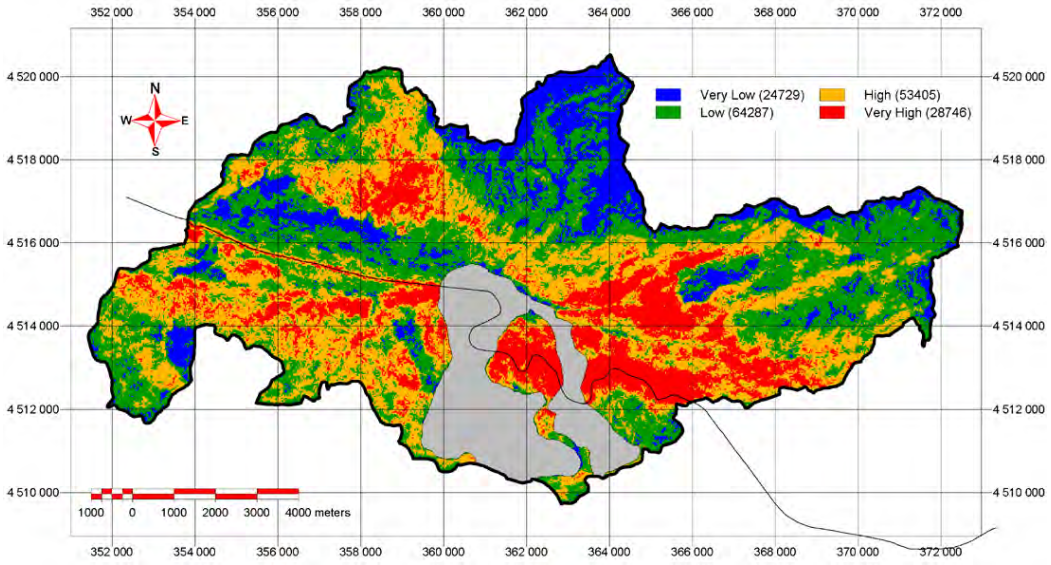
	Parameter Classes									
Parameter map	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
fault density	0,26	0,00	9,65	23,87	37,26	32,81	26,93	41,21	33,73	34,72
elevation	3,88	27,74	18,68	5,11	17,34	25,69	34,35	19,67	16,97	0,00
d to ridge	0,00	2,37	3,42	4,79	4,82	7,52	7,67	9,14	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.

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	,632	-3,500E-02	,124	,108	-2,170E-02	2,745E-02	-3,650E-02	,153	1,891E-02	-5,119E-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	-5,707E-02
	DISTE5	-7,857E-02	2,930E-02	,140	,124	-3,453E-02	,423	-,226	9,600E-02	-4,663E-02	1,803E-02	-4,477E-03	2,776E-02	-5,863E-02
	DSETTLE	-7,283E-03	-7,419E-04	-,236	,108	8,459E-02	-,226	,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	-2,170E-02	-4,981E-02	9,600E-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	-4,663E-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	-2,87E-02	1,891E-02	,263	2,776E-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	-5,119E-02	-5,707E-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	-4,224E-02	,140	-,300	-6,825E-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	-9,323E-02	,352	-6,987E-02
	DISTE5	-,126	4,980E-02	,255	,240	-5,984E-02	,698[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	,546[a]	-,256	,161	1,309E-02	-5,895E-02	,147	-7,096E-02
	DENS_FAULT	4,064E-02	-4,224E-02	4,305E-02	-4,260E-02	-8,757E-02	,230	-,256	,579[a]	,590	,144	-9,913E-02	3,241E-02	,247
	D_FAULT	6,669E-02	,140	1,412E-02	5,511E-02	7,020E-02	-,114	,161	,590	,618[a]	8,626E-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	8,626E-02	,488[a]	4,273E-02	,109	4,231E-03
	D_PRO	4,275E-02	-6,825E-02	2,362E-02	,213	-9,323E-02	-7,609E-03	-5,895E-02	-9,913E-02	-,225	4,273E-02	,696[a]	-2,604E-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	-2,604E-02	,589[a]	-,126
	Geocode	-8,254E-04	-,153	-4,96E-02	-6,997E-02	-6,987E-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).

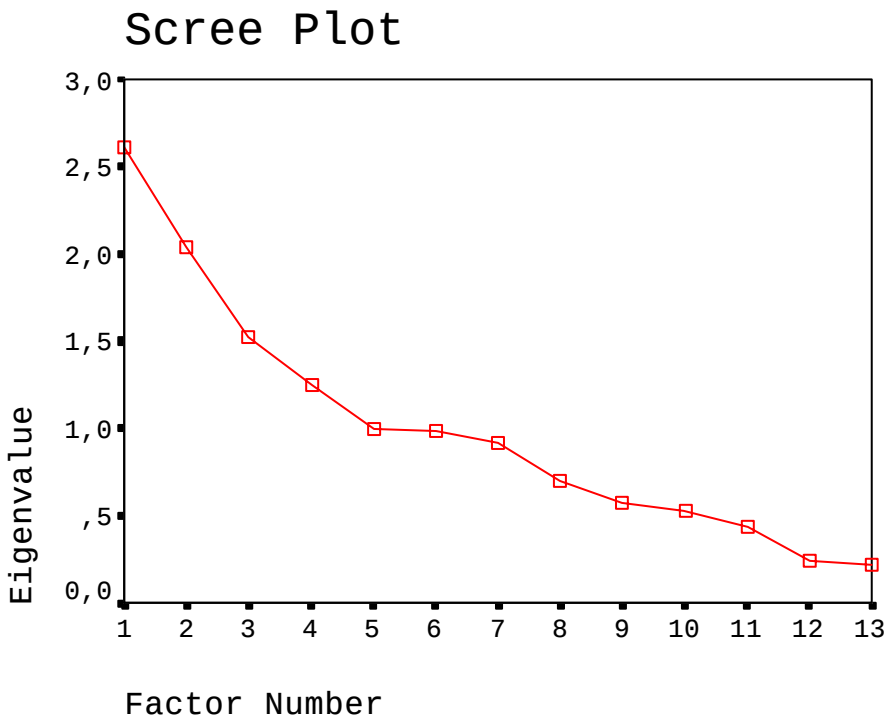


Figure 6.14. Eigenvalues of the factors.

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

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 -∞ to +∞, the probability varies from 0 to 1 on an s shaped curve. And where z is defined as:

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

Where B_0 is the intercept of the model, n is the number of independent variables, …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:

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

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.

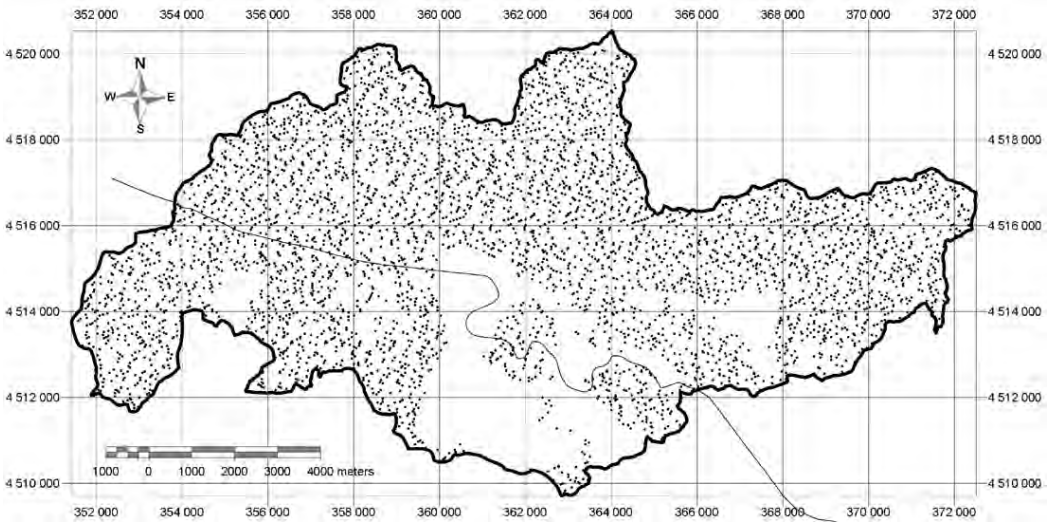


Figure 6.15. The positions of selected 4430 random landslide free nodes.

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.

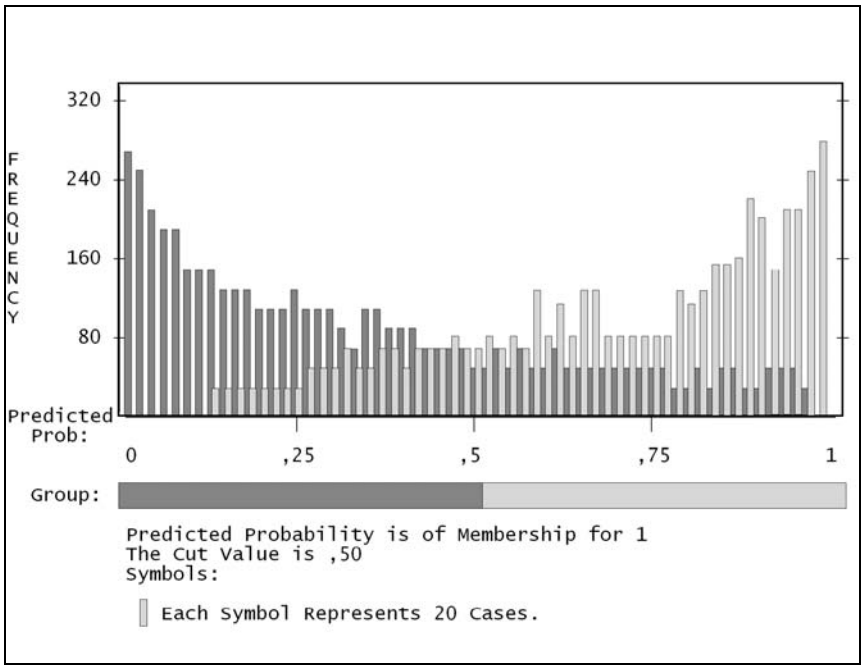
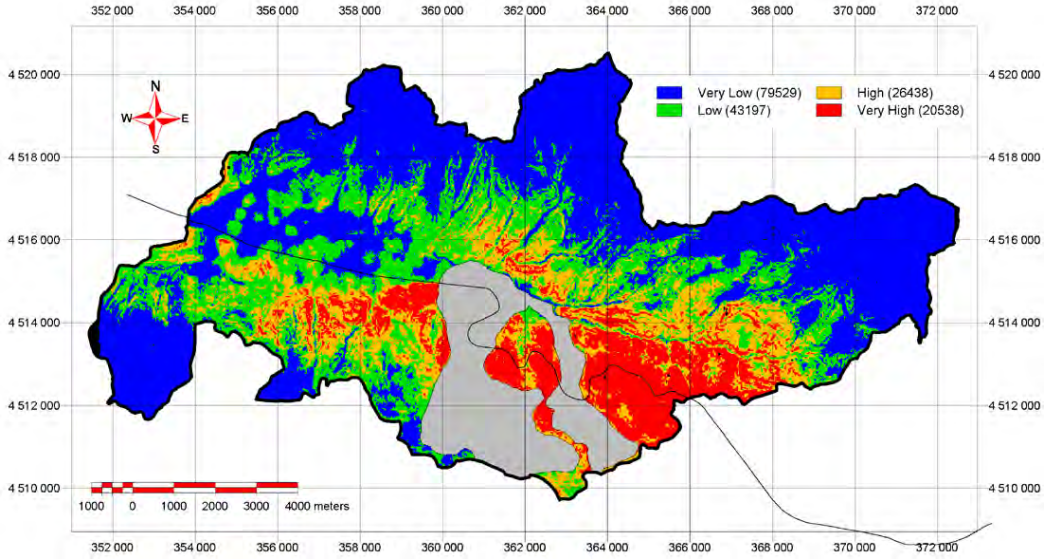


Figure 6.16. Observed groups and predicted probabilities

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.

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

BIVARIATE	LOGICAL			
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.

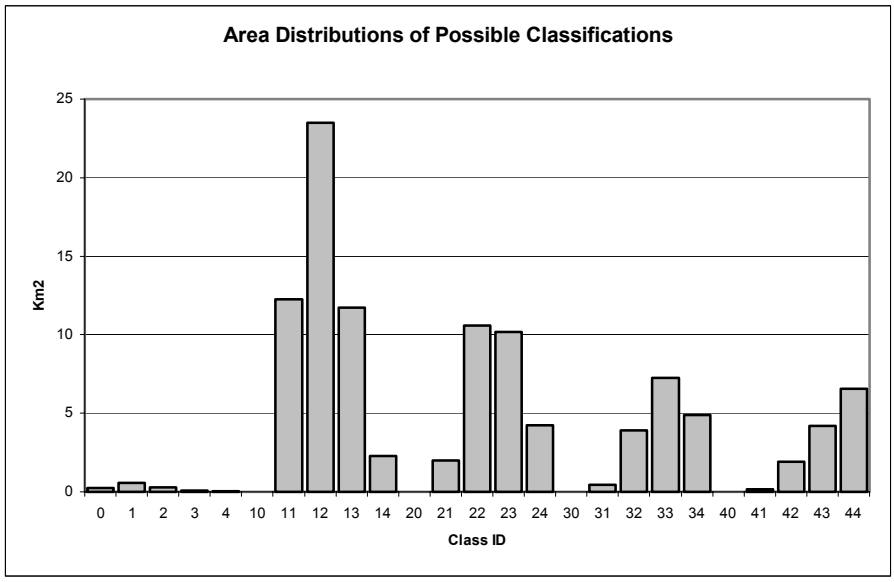


Figure 6.18. The areal distributions of classified pixels.

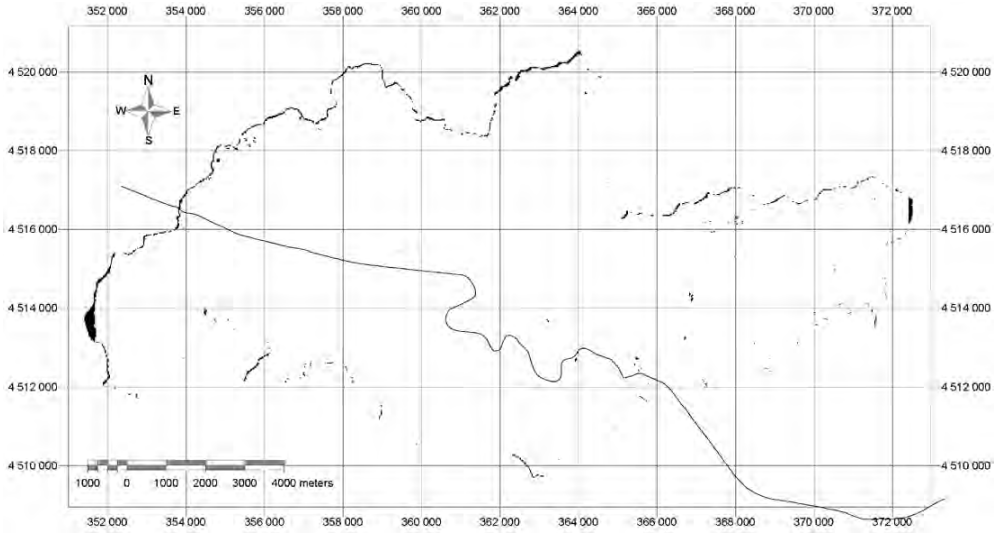


Figure 6.19. The locations of the misclassified pixels

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.

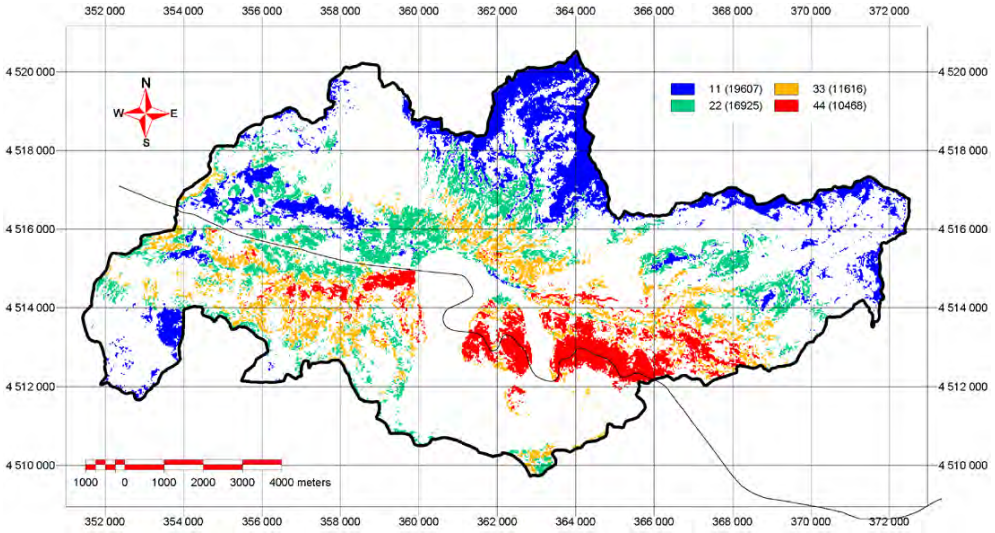


Figure 6.20. The locations of the correctly classified pixels (the numbers in the legend indicate the pixel counts).

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.

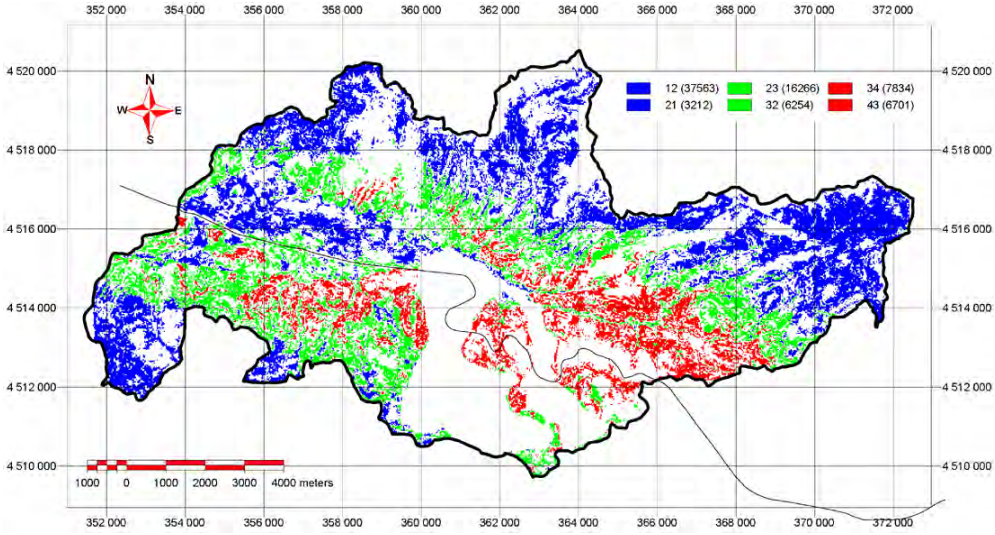


Figure 6.21. The locations of the acceptable pixels (the numbers in the legend indicate the pixel counts).

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.

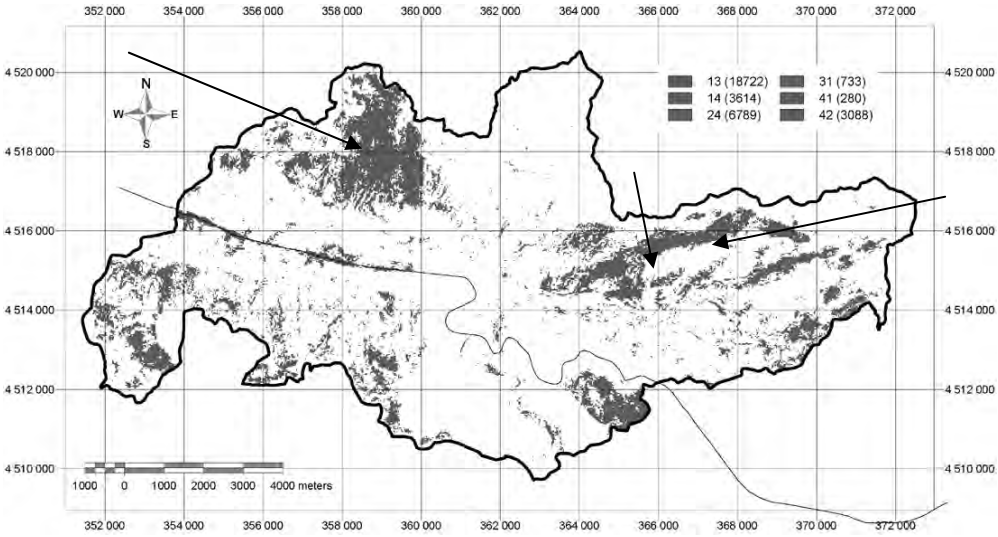


Figure 6.22. The locations of the not acceptable pixels (the numbers in the legend indicate the pixel counts).

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.

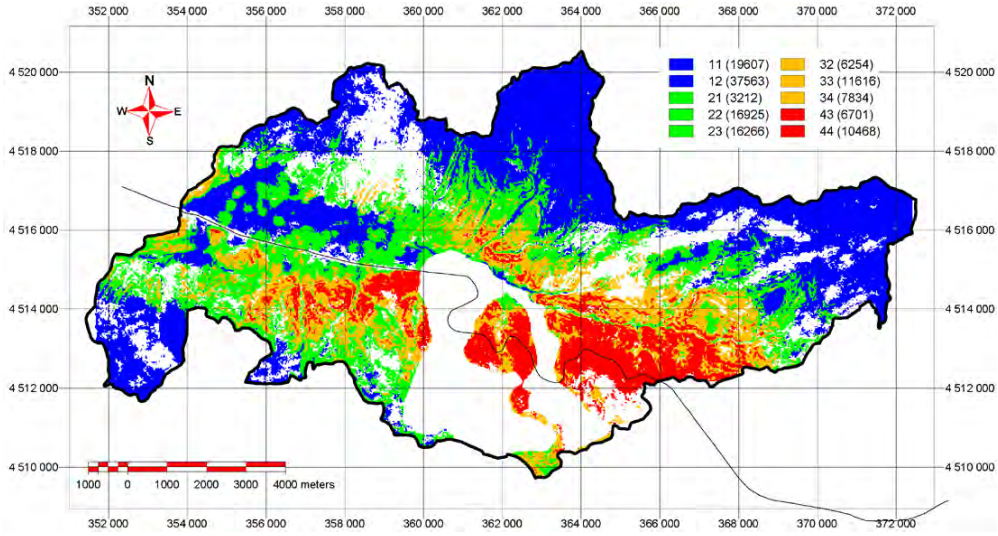


Figure 6.23. The locations of the correctly classified and the acceptable pixels united (the numbers in the legend indicate the pixel counts).

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.

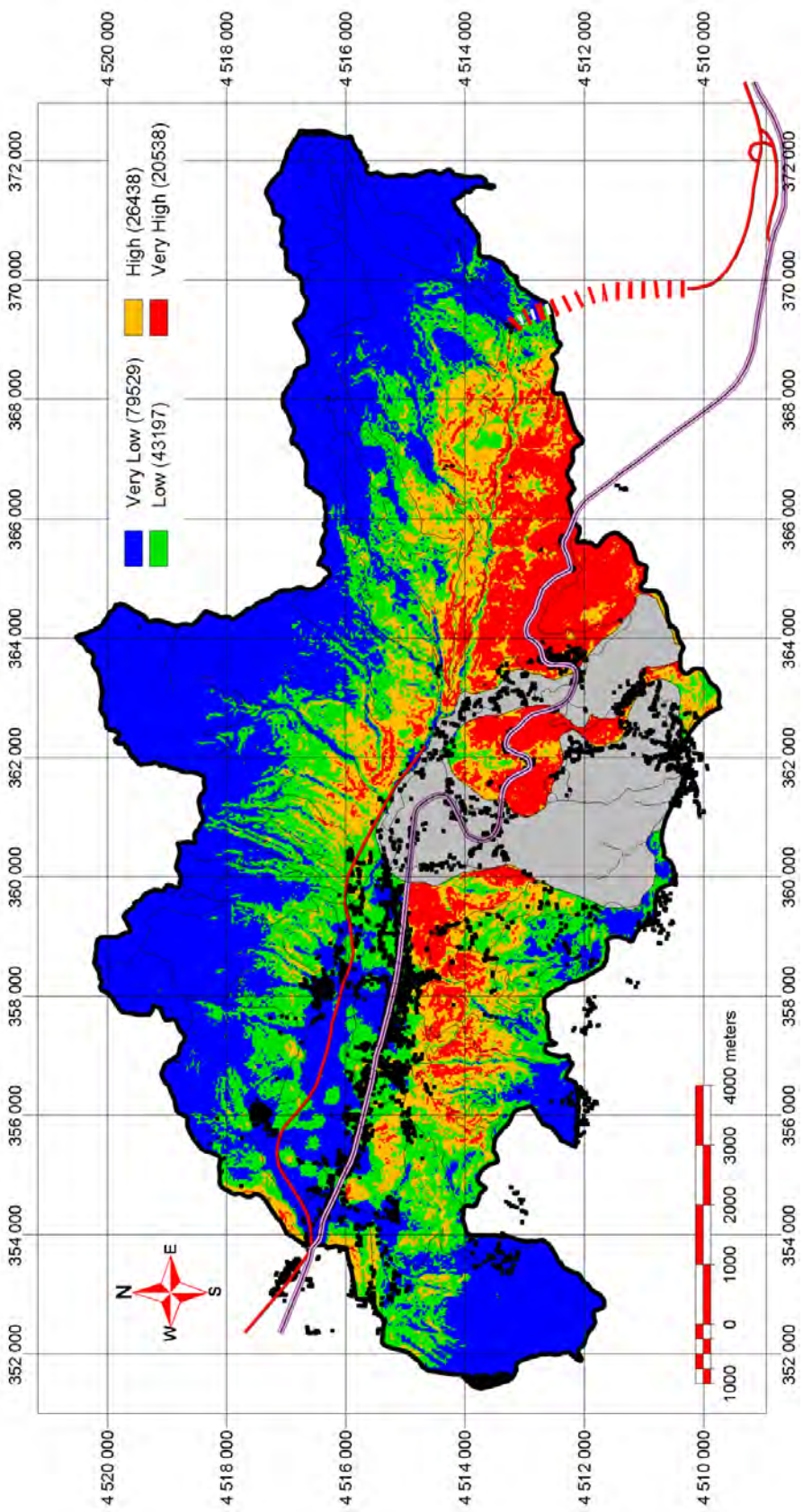


Figure 6.24. The final hazard map and the infrastructure of Asarsuyu catchment.