CHAPTER 5

LANDSLIDE DATABASES

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

5.1. Topological, Morphometrical database

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

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

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

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

5.1.1. Polystats Database

5.1.1.1. 1952 Period

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

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

Kurtosis is "A measure of the extent to which observations cluster around a central point". For a normal distribution, the value of the kurtosis statistic is 0. Positive *kurtosis indicates that the observations cluster more and have longer tails than those in the normal distribution and negative kurtosis indicates the observations cluster less and have shorter tails.*

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

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

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

Table 5.3 The descriptive statistics of Polystats 1952.

Descriptive Statistics of Variables of Polystats Database (1952)

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

a. Multiple modes exist. The smallest value is shown

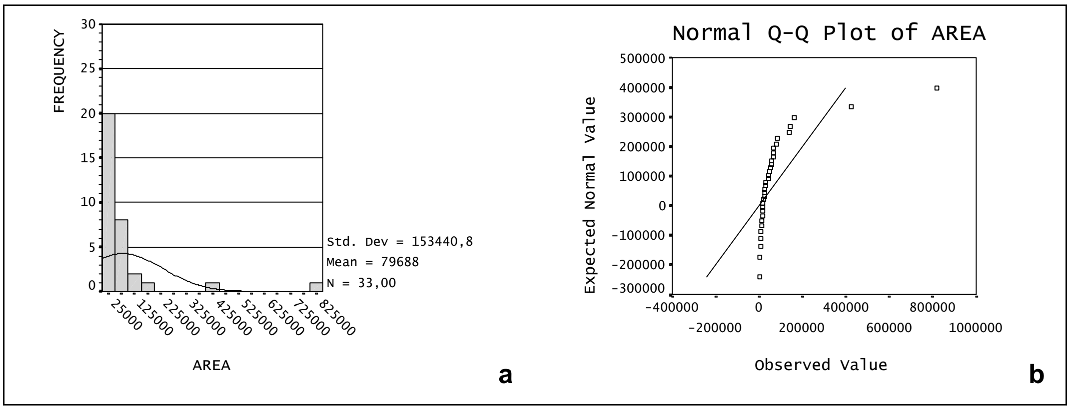


Figure 5.1. The distribution plots of area of 1952: a) Frequency distribution, b) Normal Q-Q Plot.

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

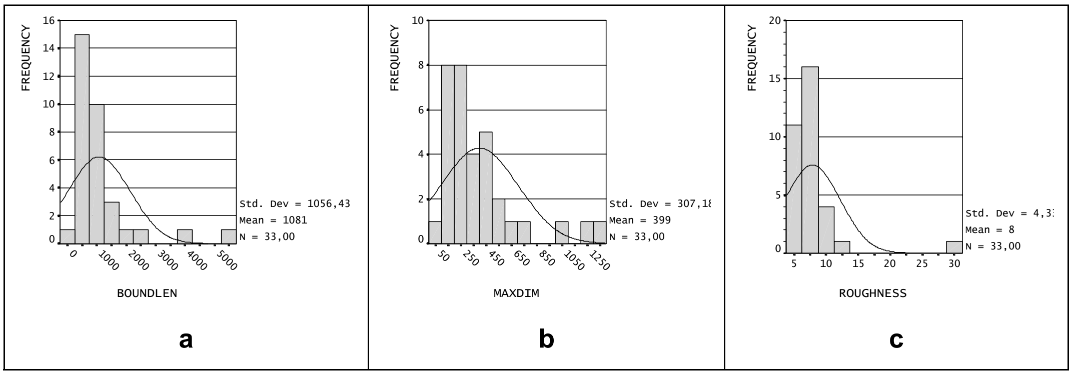


Figure 5.2. Histograms of Polystats database, a) “BoundLen”, b) “MaxDim”, c) “Roughness”.

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

Table 5.4. Correlations of Polystats variables 1952.

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

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

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

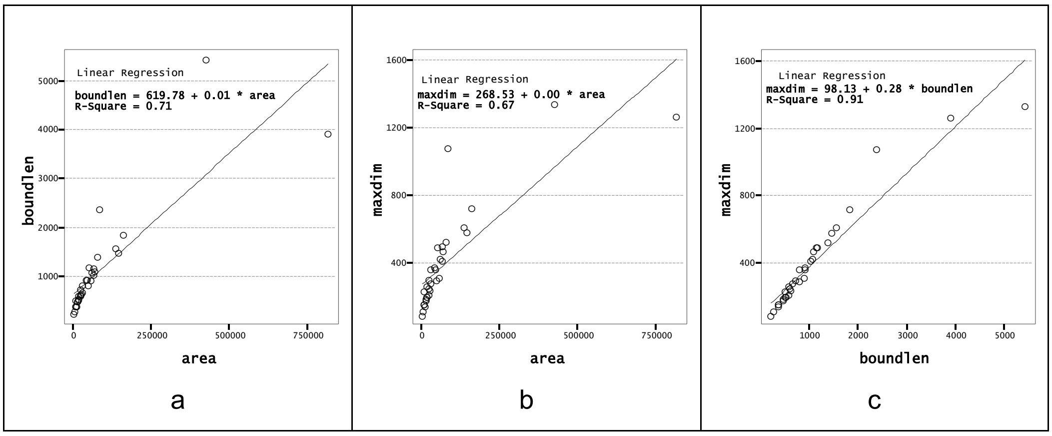


Figure 5.3. Scatter plots of highly correlated variable pairs of Polystats database, a) ”area-BoundLen”, b) “area-MaxDim”, c) “BoundLen MaxDim”.

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

5.1.1.2. 1972 Period

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

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

The “BoundLen” variable is ranging from 176.39 to 5502.14 meters with a mean of 964.23 and a standard deviation of 950.44 (Table 5.5). The kurtosis and skewness values suggest a long right tail with clustering around smaller values. The “MaxDim” variable has a minimum of 65.51 meters and the maximum of this variable is 1328.03 meters. The variables distribution suggests a less positively skewed and less clustered distribution rather than “area” and “BoundLen” variables with a mean of 357.15 and a standard deviation of 277.49 (Figure 5.4). For the “roughness” variable the minimum maximum range is defined as 4.9 and 30.68 with a mean of 8.11 and standard deviation of 4.24. The distribution has again a right long tail and clustering around small values.

Table 5.5. The descriptive statistics of Polystats 1972.

Descriptive Statistics of Variables of Polystats Database (1972)

		AREA	BOUNDLEN	MAXDIM	ROUGHNESS
N	Valid	45	45	45	45
	Missing	0	0	0	0
Mean		64906.0579833980	964.2372695904	357.1536664922	8.1114281029
Std. Error of Mean		19320.6706300324	141.6841362389	41.3659686927	.6332609391
Median		21517.4001464800	684.0998115000	264.8689724500	6.6746764100
Mode		2027.52832031[a]	176.39683459[a]	65.51217920[a]	4.90223315[a]
Std. Deviation		129606.9986989081	950.4460798904	277.4913538558	4.2480435223
Variance		16797974111,73878000	903347.7507790310	77001,4514647170	18.0458737673
Skewness		4.520	3.268	2.102	3.792
Std. Error of Skewness		.354	.354	.354	.354
Kurtosis		23.029	12.706	4.614	18.215
Std. Error of Kurtosis		.695	.695	.695	.695
Range		784151.61450196	5325.74997547	1262.52586739	25.77959581
Minimum		2027.52832031	176.39683459	65.51217920	4.90223315
Maximum		786179.14282227	5502.14681006	1328.03804659	30.68182896
Sum		2920772.60925291	43390.67713157	16071.91499215	365.01426463

a. Multiple modes exist. The smallest value is shown

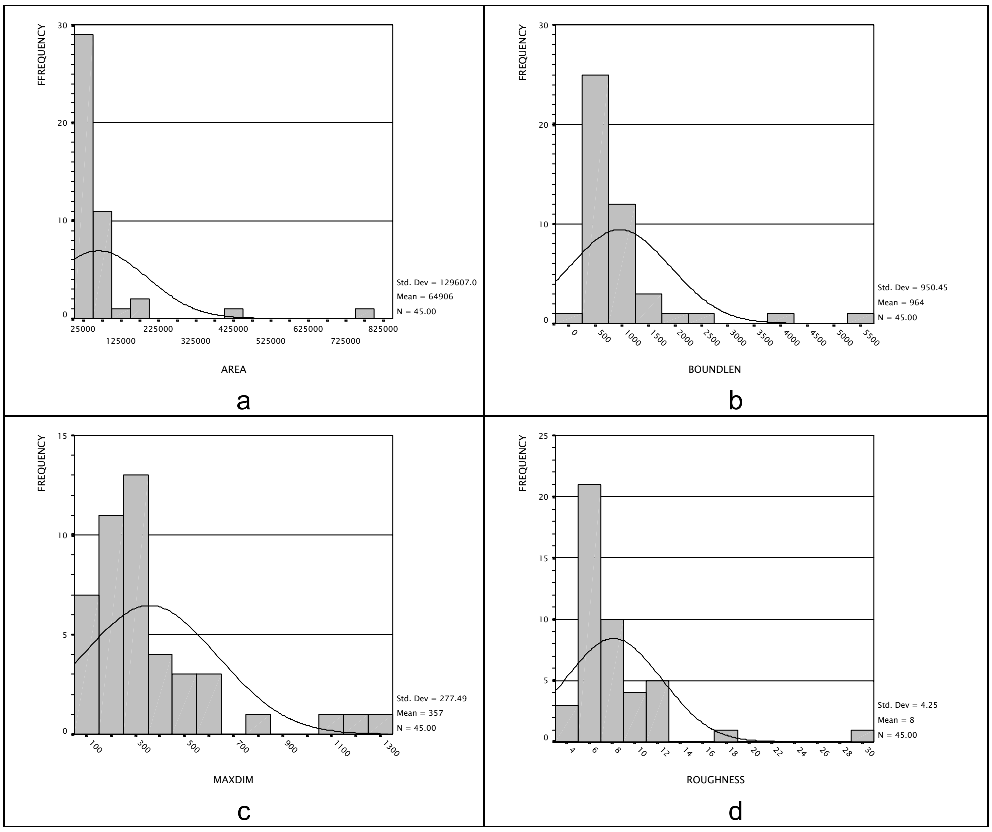


Figure 5.4. The frequency distributions of variables of 1972 Polystats database, a) area, b) BoundLen, c) MaxDim, d) roughness.

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

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

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

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

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

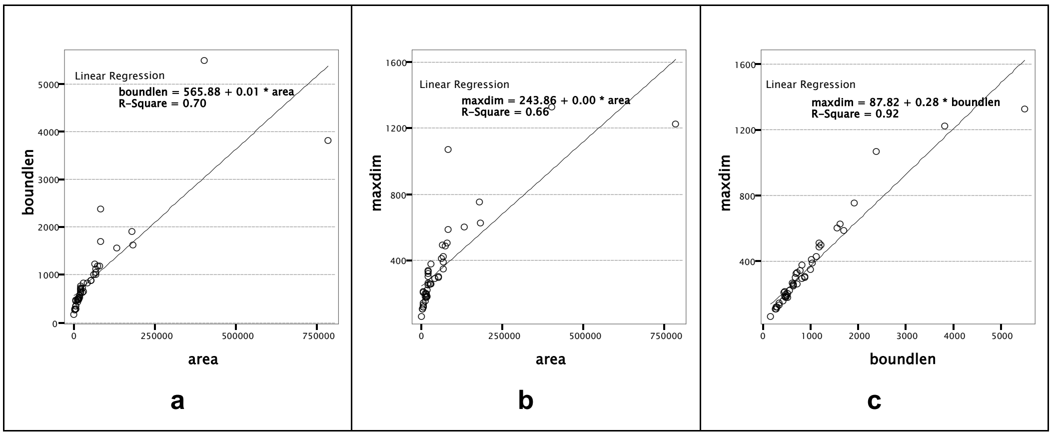


Figure 5.5. Scatter plots of highly correlated variable pairs of Polystats database (1972), a) ”area”-“BoundLen”, b) “area”-“MaxDim”, c) “BoundLen”-“MaxDim”.

5.1.1.3. 1984 Period

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

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

Table 5.7. The descriptive statistics of Polystats 1984

Descriptive Statistics of Variables of Polystats Database (1984)

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

a. Multiple modes exist. The smallest value is shown

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

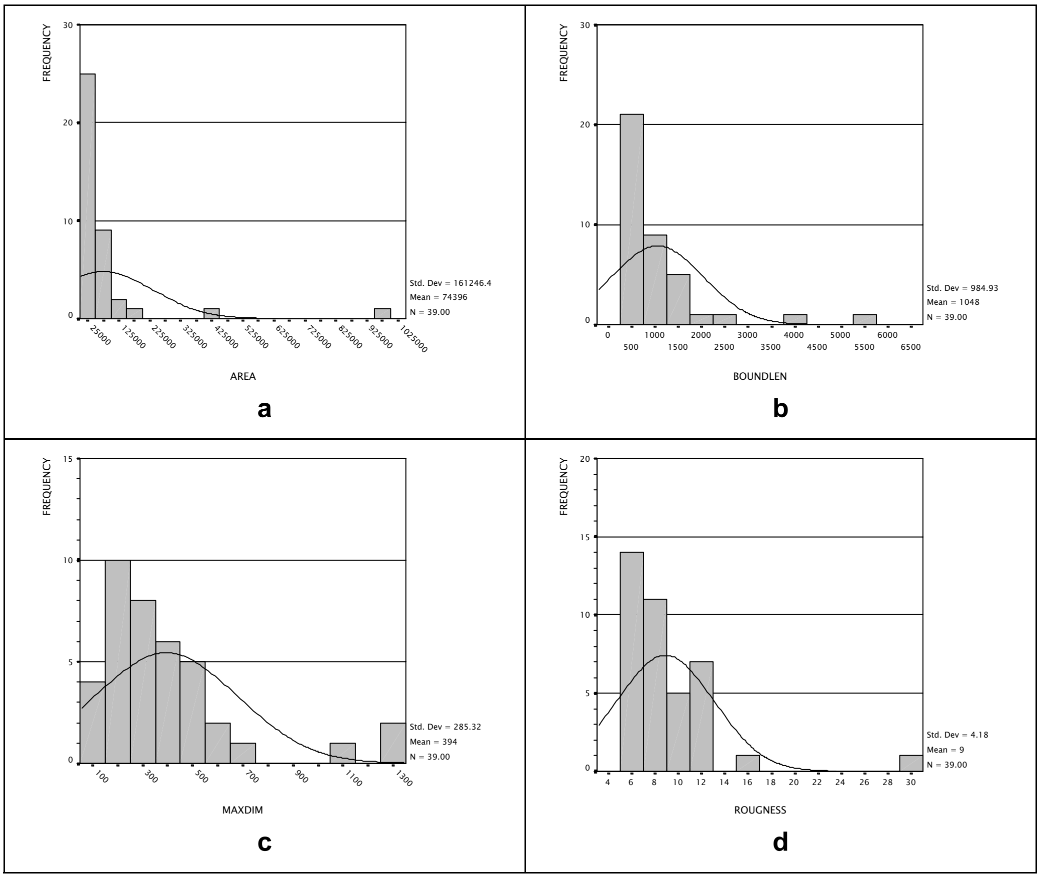


Figure 5.6. The frequency distributions of variables of 1972 Polystats database, a) area, b) BoundLen, c) MaxDim, d) roughness.

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

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

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

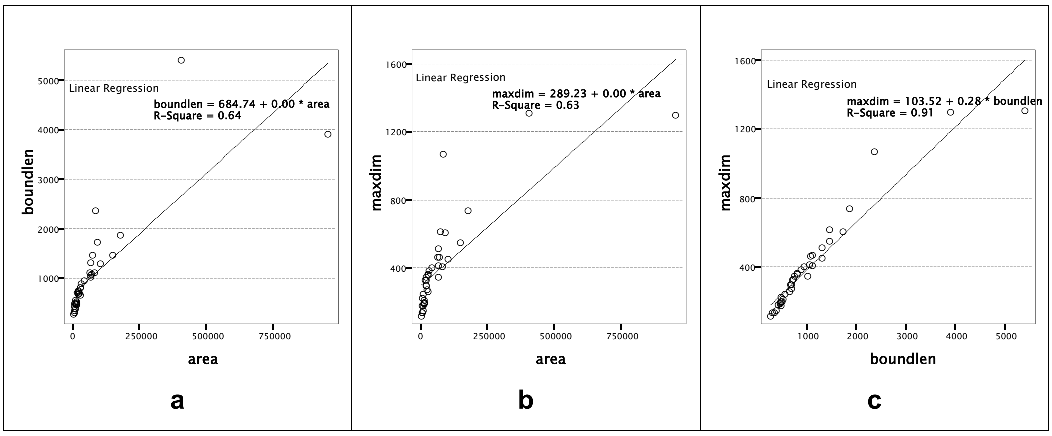


Figure 5.7. Scatter plots of highly correlated variable pairs of Polystats database 1984, a) "area"-"BoundLen", b) "area"-"MaxDim", c) "BoundLen"-"MaxDim".

5.1.1.4. 1994 Period

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

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

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

Table 5.9. The descriptive statistics of Polystats 1994

Descriptive Statistics of Variables of Polystats Database (1994)

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

a. Multiple modes exist. The smallest value is shown

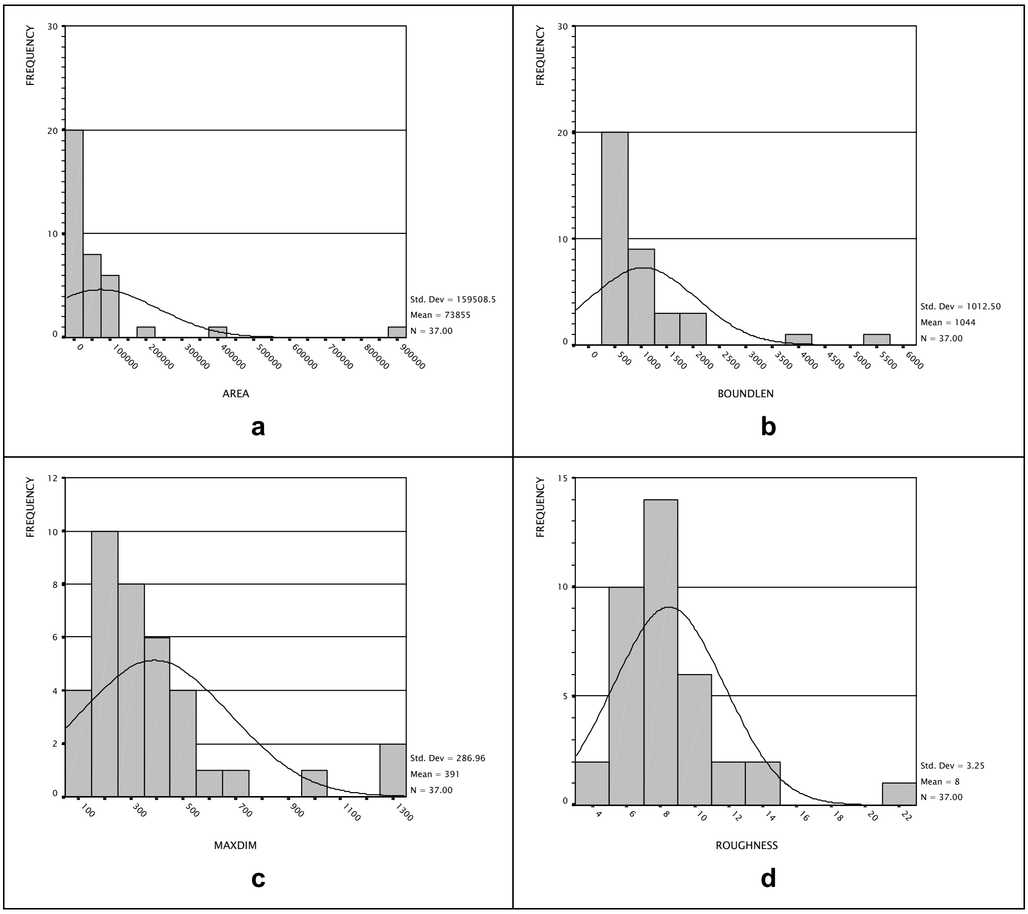


Figure 5.8. The frequency distributions of variables of 1994 Polystats database, a) area, b) BoundLen, c) MaxDim, d) roughness.

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

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

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

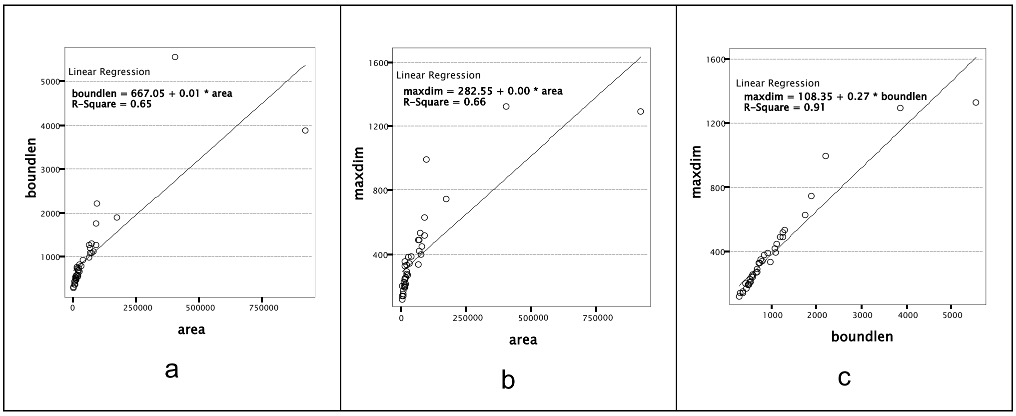


Figure 5.9. Scatter plots of highly correlated variable pairs of Polystats database (1994), a) ”area-BoundLen”, b) “area-MaxDim”, c) “BoundLen MaxDim”.

5.1.1.5. The Comparison of Four Periods

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

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

Table 5.11. The results of Paired Samples t test.

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	A_52 - A_72	3133.2352943428	225138.3699281057	39799.2170203623	-78037.8029747261	84304.2735634117	.079	31	.938
Pair 2	A_72 - A_84	-2274.5032497102	226608.6488145063	37254.2323797402	-77829.5884501831	73280.5819507627	-.061	36	.952
Pair 3	A_84 - A_94	-22118.4051513676	176248.2498681428	28975.0337926642	-80882.4973644682	36645.6870617330	-.763	36	.450
Pair 4	B_52 - B_72	6.8854877403	1551.3268994336	274.2384426067	-552.4275034747	566.1984789553	.025	31	.980
Pair 5	B_72 - B_84	-11.4809617414	1476.2485762631	242.6937709486	-503.6867426775	480.7248191948	-.047	36	.963
Pair 6	B_84 - B_94	-69.1765357922	1375.9556543506	226.2057161523	-527.9429917081	389.5899201238	-.306	36	.762
Pair 7	M_52 - M_72	9.1444219503	444.0748629193	78.5020867312	-150.9616395081	169.2504834087	.116	31	.908
Pair 8	M_72 - M_84	-7.4818299246	414.7648201123	68.1869163963	-145.7713060134	130.8076461642	-.110	36	.913
Pair 9	M_84 - M_94	-21.4812258095	378.1292478052	62.1640654097	-147.5557939434	104.5933423245	-.346	36	.732
Pair 10	R_52 - R_72	-2.8773702812E-02	5.0865587919	.8991850537	-1.8626737106	1.8051263049	-.032	31	.975
Pair 11	R_72 - R_84	-.2011592211	6.0301049122	.9913431409	-2.2116962981	1.8093778559	-.203	36	.840
Pair 12	R_84 - R_94	.4110326595	5.5999202025	.9206212103	-1.4560736943	2.2781390132	.446	36	.658
Pair 13	A_52 - A_84	27069.3210870145	187088.1695223885	34741.4009083855	-44095.2126496878	98233.8548237167	.779	28	.442
Pair 14	A_52 - A_94	-8360.2636152007	261756.4594128563	49467.3211193707	-109858.8226564464	93138.2954260449	-.169	27	.867
Pair 15	A_72 - A_94	1592.8329171311	225210.0548926682	38067.4472200363	-75769.5276795204	78955.1935137826	.042	34	.967
Pair 16	B_52 - B_84	76.4396940662	1552.4303504508	288.2790789233	-514.0732300303	666.9526181627	.265	28	.793
Pair 17	B_52 - B_94	-78.3791612904	1749.3247757647	330.5913084969	-756.6964965475	599.9381739667	-.237	27	.814
Pair 18	B_72 - B_94	29.5674699980	1454.6627507336	245.8828825955	-470.1266680919	529.2616080879	.120	34	.905
Pair 19	M_52 - M_84	19.4612184045	434.1497285018	80.6195806536	-145.6805063749	184.6029431839	.241	28	.811
Pair 20	M_52 - M_94	-33.3368587636	494.8430475471	93.5165458442	-225.2169613225	158.5432437954	-.356	27	.724
Pair 21	M_72 - M_94	7.5023579734	412.3526156054	69.7003135055	-134.1457214458	149.1504373926	.108	34	.915
Pair 22	R_52 - R_84	-.8524458559	6.7143750055	1.2468281373	-3.4064575169	1.7015658052	-.684	28	.500
Pair 23	R_52 - R_94	-.7490438543	5.6189166838	1.0618754416	-2.9278322901	1.4297445816	-.705	27	.487
Pair 24	R_72 - R_94	.2064057986	5.9090675511	.9988147165	-1.8234299248	2.2362415219	.207	34	.838

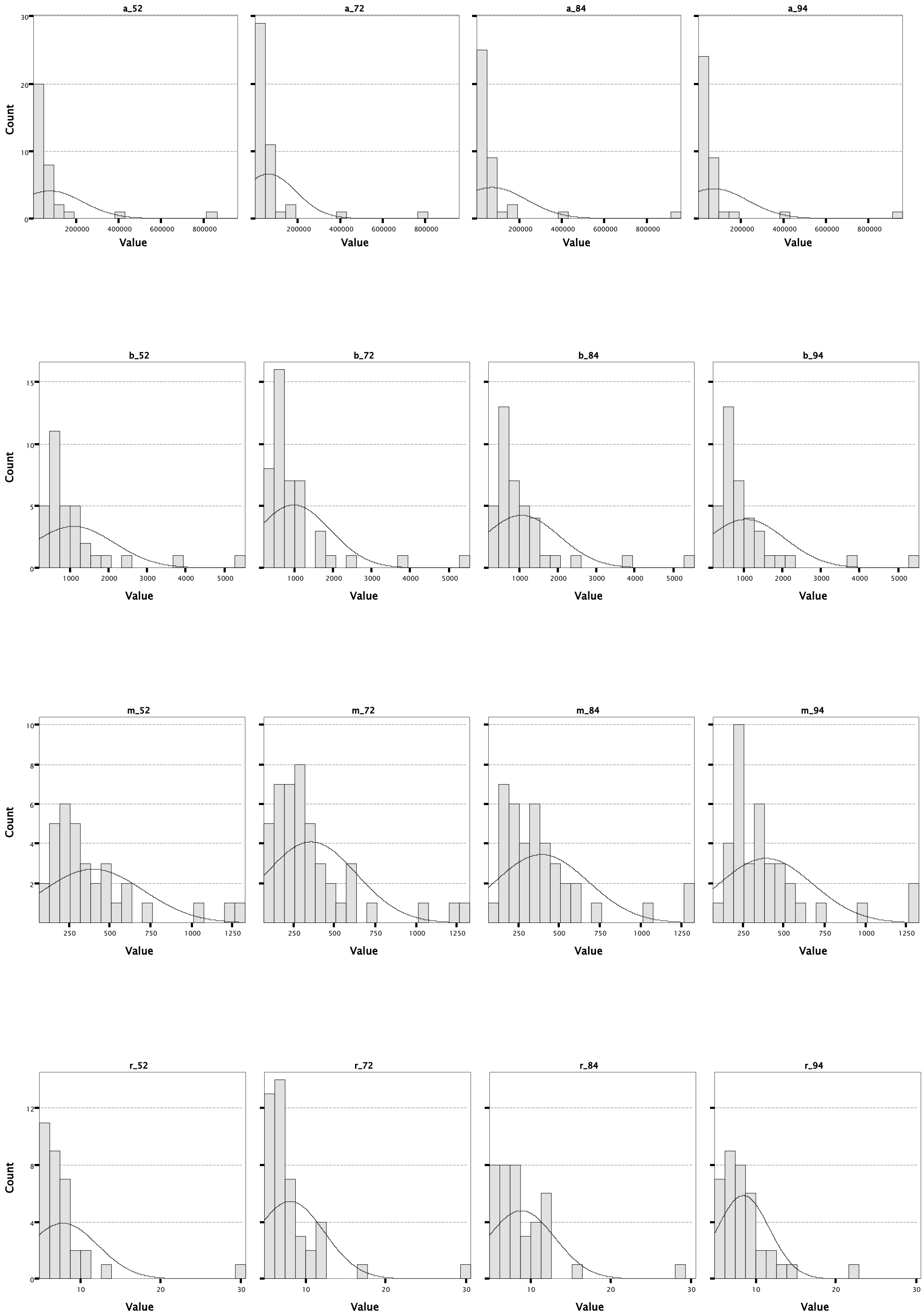


Figure 5.10. The distributions of Polystats variables in 4 periods. (A_52: “Area” in 52 period, b_72 “BoundLen” in 1972, m_84: “MaxDim” in 1984, r_94: “roughness” in 1994), each column represents single period and each row represents single parameter.

Table 5.12. The ANOVA table of Polystats variable.

ANOVA

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

5.1.2. The Fuzzystats Database

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

5.1.2.1. Form Ratio

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

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

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

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

a. Multiple modes exist. The smallest value is shown

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

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

a. Test distribution is Normal.

b. Calculated from data.

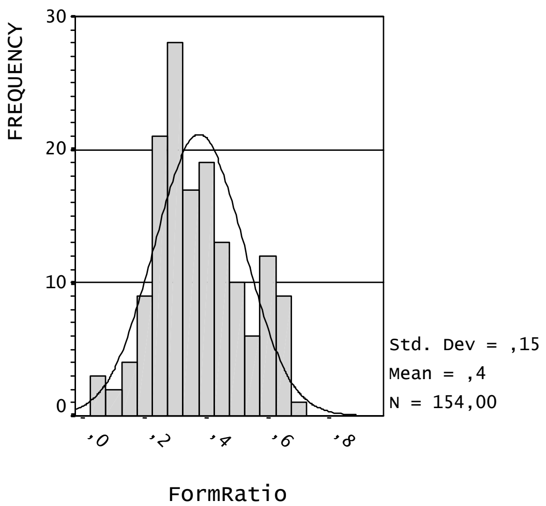


Figure 5.11. The frequency distribution of cumulative “Form ratio”.

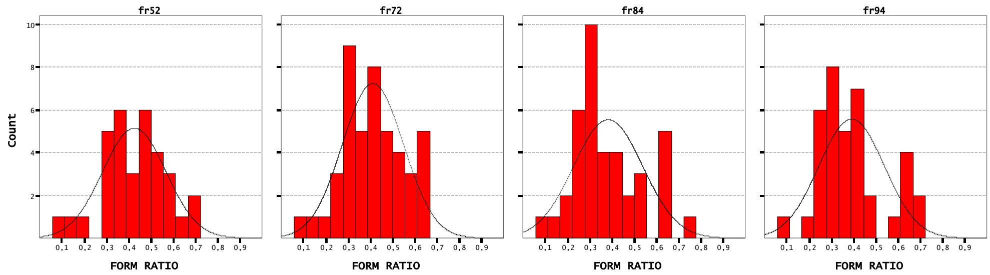


Figure 5.12. The frequency distributions of “form ratio” in all periods.

5.1.2.2. Grain Shape Index

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

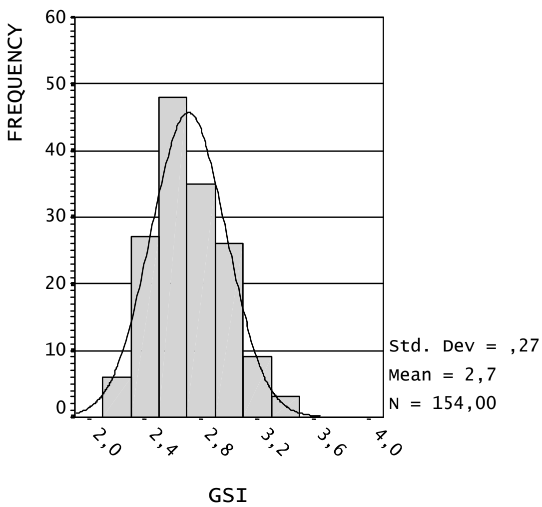


Figure 5.13. The frequency distribution of cumulative “Grain Shape Index”.

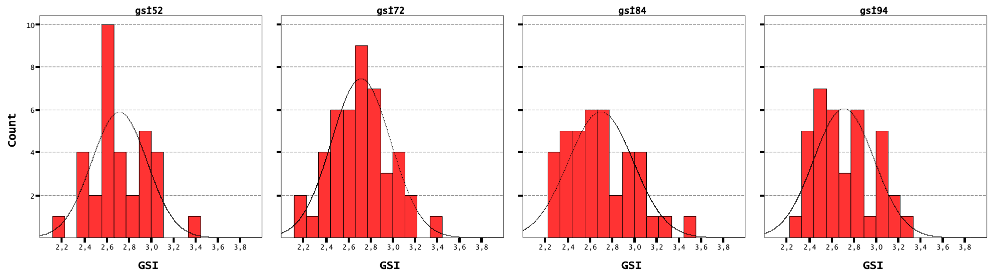


Figure 5.14. The frequency distributions of “form ratio” in all periods.

5.1.2.3. Compactness

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

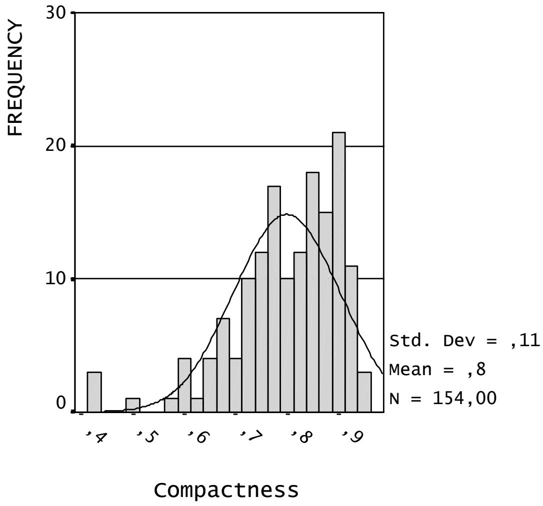


Figure 5.15. The frequency distribution of cumulative “Compactness”.

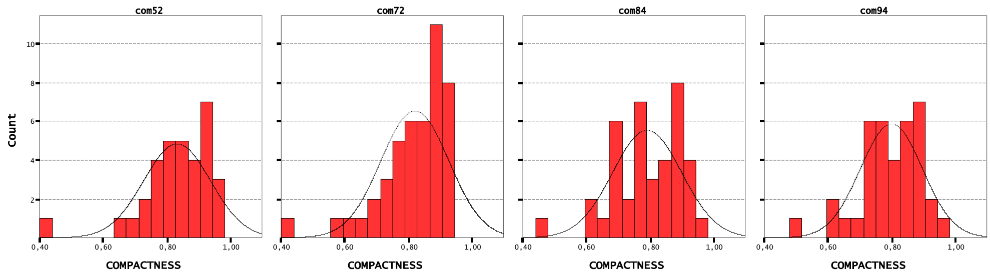


Figure 5.16. The frequency distributions of “compactness” in all periods.

5.1.2.4 Circularity 1

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

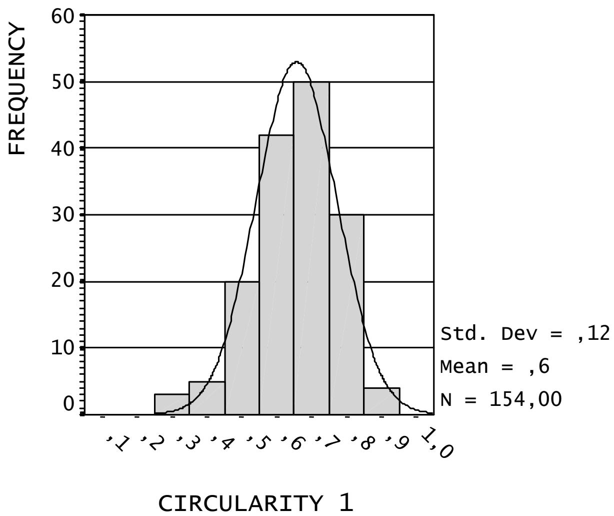


Figure 5.17. The frequency distribution of cumulative “Circularity 1”.

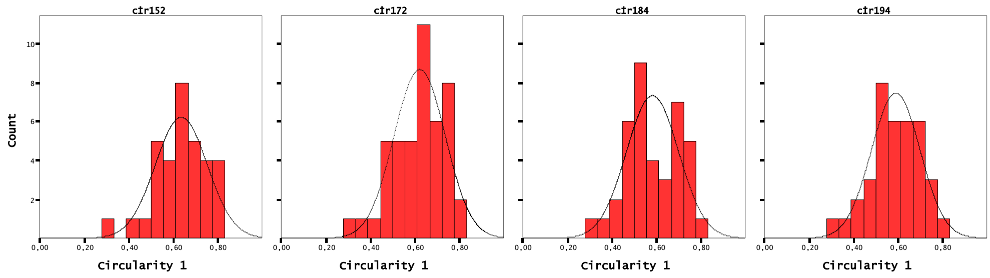


Figure 5.18. The frequency distributions of “Circularity 1” in all periods.

5.1.2.5. Circularity 2

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

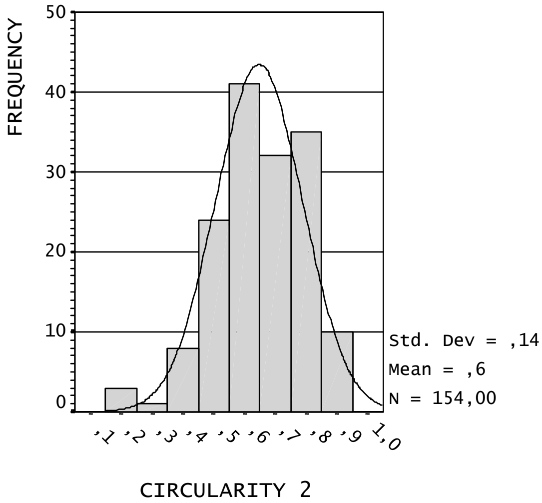


Figure 5.19. The frequency distribution of cumulative “Circularity 2”.

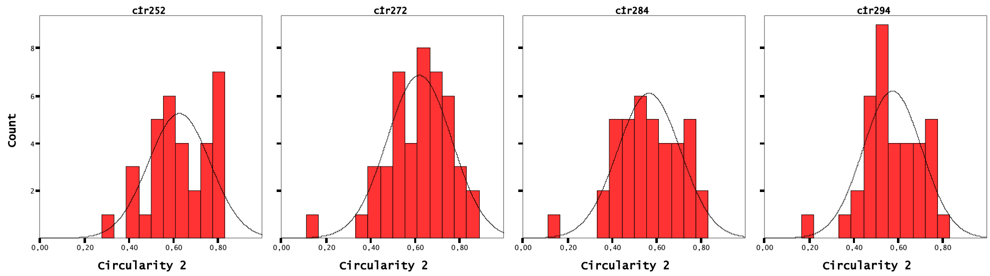


Figure 5.20. The frequency distributions of "Circularity 2" in all periods.

5.1.2.6. Elongation

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

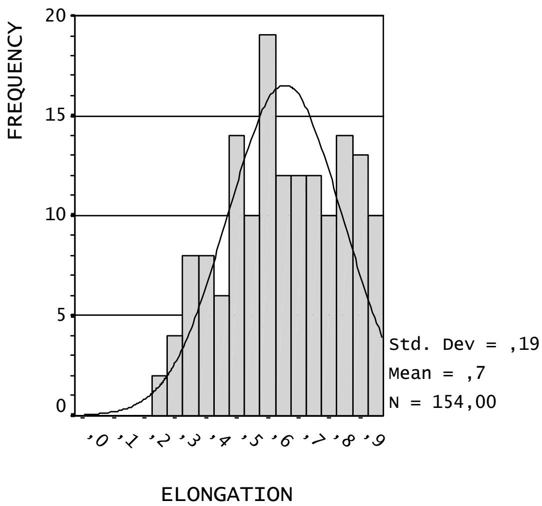


Figure 5.21. The frequency distribution of cumulative "Elongation".

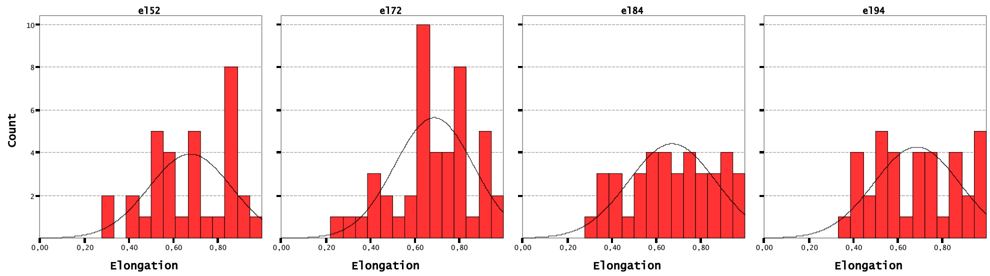


Figure 5.22. The frequency distributions of “Elongation” in all periods.

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

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

Table 5.15. The changes through time in Polystats database

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

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

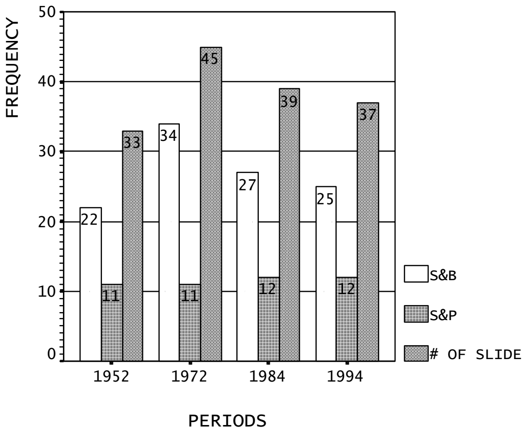
The effect of re-forestation activities is seen in 1984 and 1994 periods cumulatively. In the images of 1984 reforestation is started few years ago and yet the land cover is still not mature, in 1994 the land cover is very similar to that of 1952, mainly covered with dense forests. These are reflected as a decrease from both sides of the “Form Ratio” graph, clustering into mean values from 1984 to 1994, giving rise to more ellipsoidal shape. Namely the increase in elongation in 1984 is balanced with the road cuts and dense forests in 1994. The “Grain Shape Index” responds to these changes as cutting off the extreme values and resulting in more homogenous distribution. This is probably due to the decrease in long axis and so the perimeter. Although the reforestation seems to be efficient, some outlier peaks are observed in the 2.4-2.5 field of “Grain Shape Index” 1994, which could be attributed to the landslides occurring by the road cuts. Also for the compactness variable the increase in 1972 period is trimmed in both 1984 and 1994 periods giving rise to a decrease in perimeter and area. Both of the circularity variables show slight increments in the circular side and decreases in the elongated side. A very sharp increase in the “Circularity 2” 1994 could easily be attributed to newly generated landslides caused by road cuts. The “elongation” graph of 1972 is trimmed down to mean values in the 1984 period especially in the 0,6-0,8 range and a very slight increment in 1994 period in 0,5 group. As a result in 1984 and 1994 periods, the newly generated flow type instabilities are either vanished or stabilized but other new or regenerated larger slides (larger than that of de-forestation) occur due to the road activities in the area.

5.2. Photo-characteristics Database

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

5.2.1. Massinfo

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

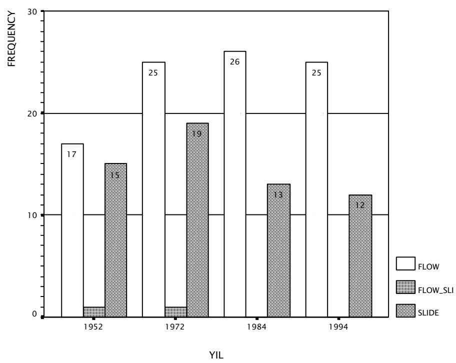


	1952	1972	1984	1994
s&b	22	34	27	25
s&p	11	11	12	12
# of slide	33	45	39	37

Figure 5.23. The frequencies of the mass info variable through time and data table

5.2.2. Type

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

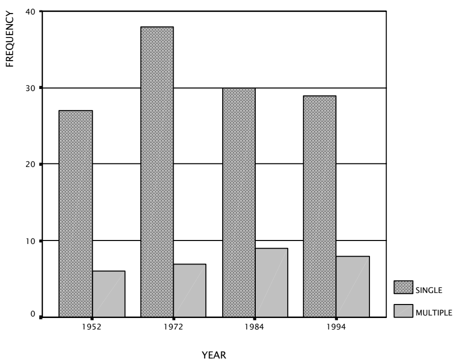


	1952	1972	1984	1994
Flow	17	25	26	25
Slide	15	19	13	12
Flow/slide	1	1	0	0

Figure 5.24. The frequencies of the type variable through time and data table

5.2.3. Style

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

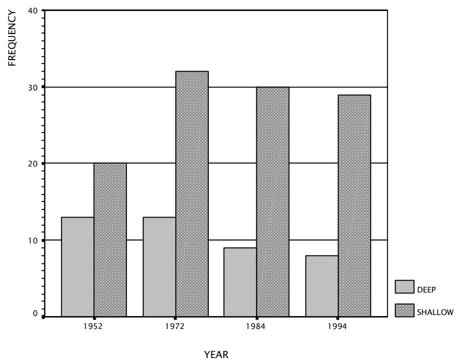


	1952	1972	1984	1994
single	27	38	30	29
multiple	6	7	9	8

Figure 5.25. The frequencies of the style variable through time and data table

5.2.4. Depth

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



	1952	1972	1984	1994
deep	13	13	9	8
shallow	20	32	30	29

Figure 5.26. The frequencies of the depth variable through time and data table

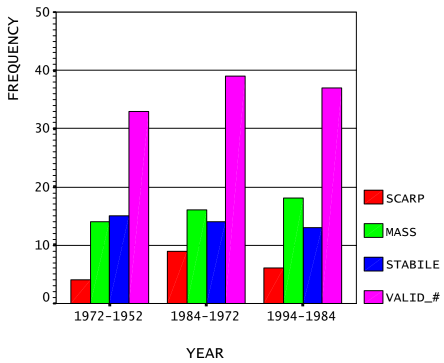
5.2.5. Distribution of Activity

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

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

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

As this analysis deals with the two periods of time the number of the slides used in this analysis is taken as the lowest amount of slide of the two years. Furthermore if the slide does not have an attribute as either scarp related or mass related the stable attribute is given. The striking fact in the graph presented in Figure 5.27 is the increase in scarp in 1972-1984 periods that might be attributed to the new re-routing of E-5 highway. The mass related variable has an increasing trend with a very small slope through time and this might again be attributed to the construction of highway in the study area. When stayed out of the between variable changes, the starting period has a significant discrepancy within the variables themselves as the difference of mass related activity is started as 18, which means there are 18 cases which were not active in the 1952 period and activated in 1972. This is one of the clear evidence of de-forestation.



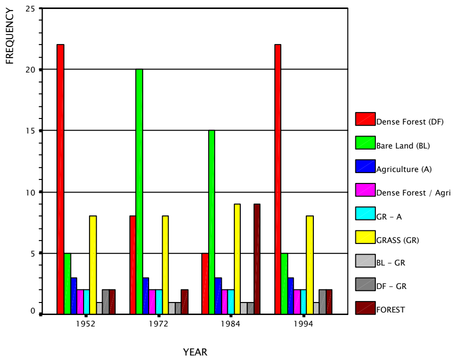
scarp	mass	Stabile	valid count	year
6	18	13	37	1994-1984
9	16	14	39	1984-1972
4	14	15	33	1972-1952

Figure 5.27. The frequencies of the distribution of activity variable through time and data table

5.2.6. Land cover

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

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



denseforest	bare	agri	DF/Agri	GR/Agri	Grass	Bare/GR	DF/GR	Forest	Year
22	5	3	2	2	8	1	2	2	1952
8	20	3	2	2	8	1	1	2	1972
5	15	3	2	2	9	1	1	9	1984
22	5	3	2	2	8	1	2	2	1994

Figure 5.28. The frequencies of the land cover variable through time and data table

5.3. Landslide Attribute Databases

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

Table 5.17. The nature and ranges of transferred attribute database

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

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

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

5.3.1. LITHOMAP

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

In the LITHOMAP of seed cells, the situation slightly changes, though the preferred conditions are still the same. The cumulative total of the preferred first three lithologies are 89,5 % which is 3,4% lower than that of slided mass. Moreover the distributions in the preferred lithologies are also changing. While percentage of Aksudere formation is increasing the percentage of Yedigöller (21,4 % to 18,8 %) and Fındıklıdere (43.4 % to 32,6 %) formations are decreasing. Also in the remaining 10% of the data some slight changes have been observed as the percentage of Quaternary is decreased as expected and Çaycuma formation is increased.

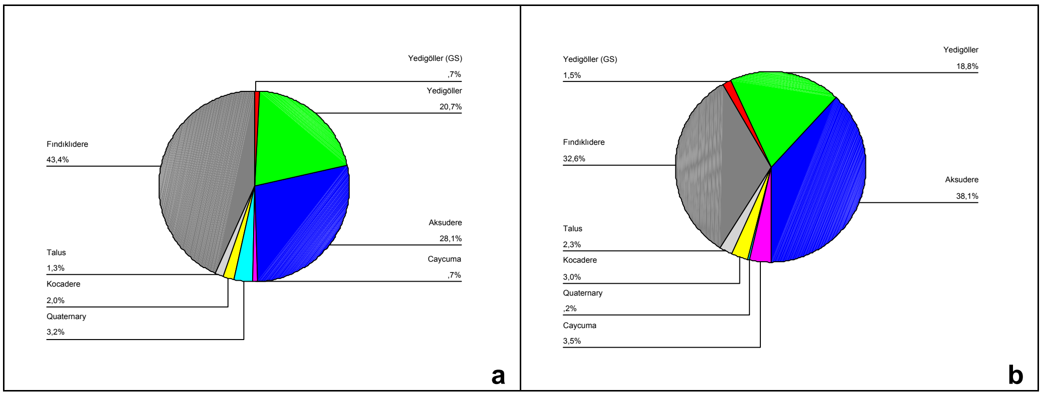


Figure 5.29. The preferred lithologies and their percentages: a) slided Mass, b) the seed cells.

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

5.3.2. DISTFAULT and FAULTDENS

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

It is seen that the maximum value of the landslide data can go up to 1441 and the seed cells have a maximum of 1517 that is nearly one third of the maximum of the whole data (Figure 4.5), which means there is redundant information in the whole data, not relevant to landslides. On the other hand the maximum of the FAULTDENS are nearly the same both in observed and in whole data (Figure 5.30). This is because of the spatial orientation of both the landslides and the fault lines. A polynomial distribution is seen in the frequency histogram of FAULTDENS as having peaks around 50, 225 and 600. These could be attributed to landslides very distant to faults, landslides near to one set of faults and landslides near to two or more sets of faults. This could easily be seen when the thematic maps of the fault density is investigated (Figure 4.6). Hence, it can be said that landslides are not much far away than 1500 meters to the fault lines and the preference order related to density is: near to one set of fault, very far away from faults and the least preferred as crisscrossing of two or more sets of fault lines, which is valid also for the seed cells.

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

Correlations

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

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

a

Correlations

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

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

b

Table 5.19. The comparison of whole data and landslide data

		Slided mass data	Seed cells	Whole Data
DISTANCE TO FAULT				
	Min	1	0	0
	Max	1411	1517	4791
	Mean	361,47	408,52	725,887
	St.Dev	286,44	339,66	827,725
FAULT DENSITY				
	Min	0	0	0
	Max	671	705	768
	Mean	229,1	235,56	195,193
	St.Dev	163,58	161,03	154,61

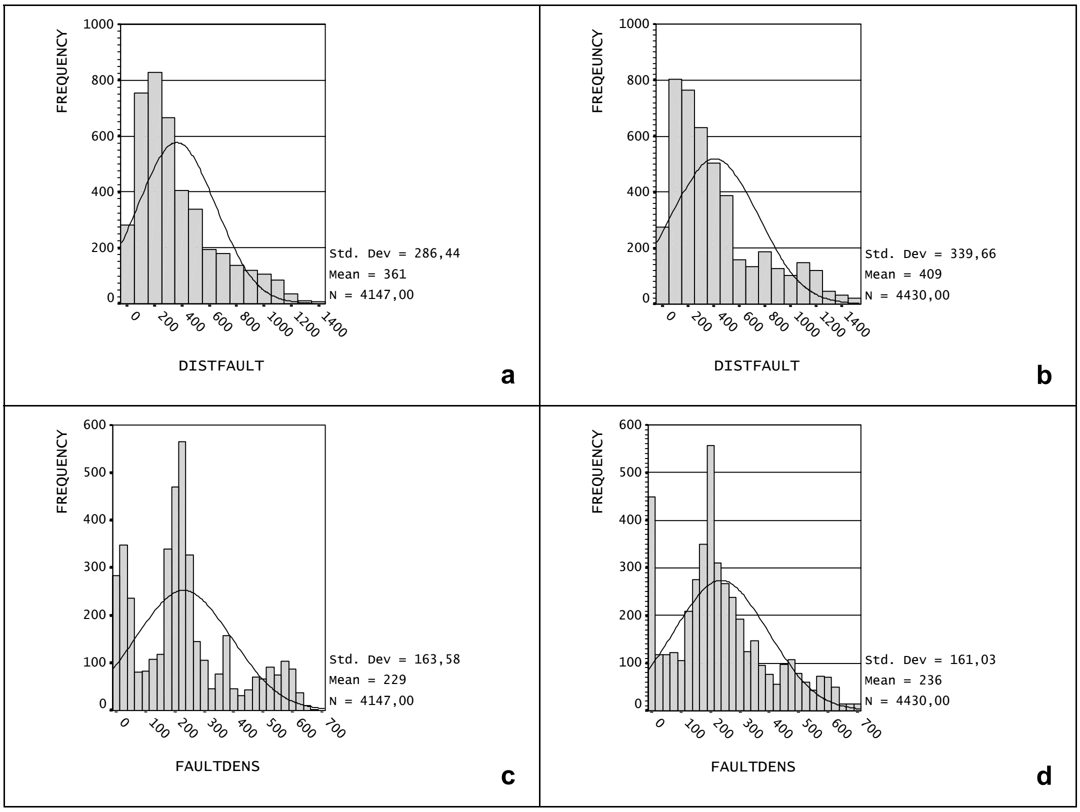


Figure 5.30. The frequency distributions of DISTFAULT (a. slided mass, b. seed cells) and FAULTDENS (c. slided mass, b. seed cells).

5.3.3. ELEVMAP

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

Table 5.20. The descriptive statistics of the ELEVMAP.

		Slided mass data	Seed cells	Whole Data
ELEVATION				
	Min	295	300	220
	Max	1050	1130	1580
	Mean	594,62	661,84	679,96
	St.Dev	177,191	179,21	265,06

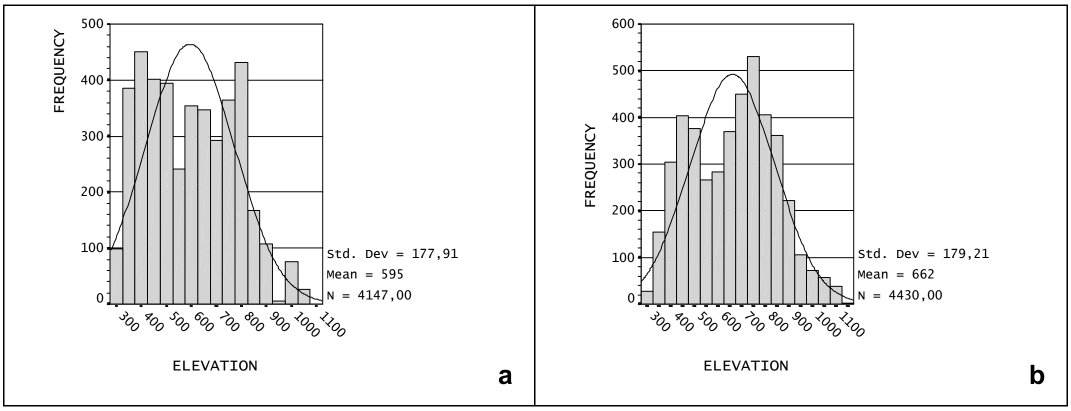


Figure 5.31. The frequency distributions of ELEVMAP, a. Slided mass, b. Seed cells

5.3.4. Distance to drainage

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

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

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

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

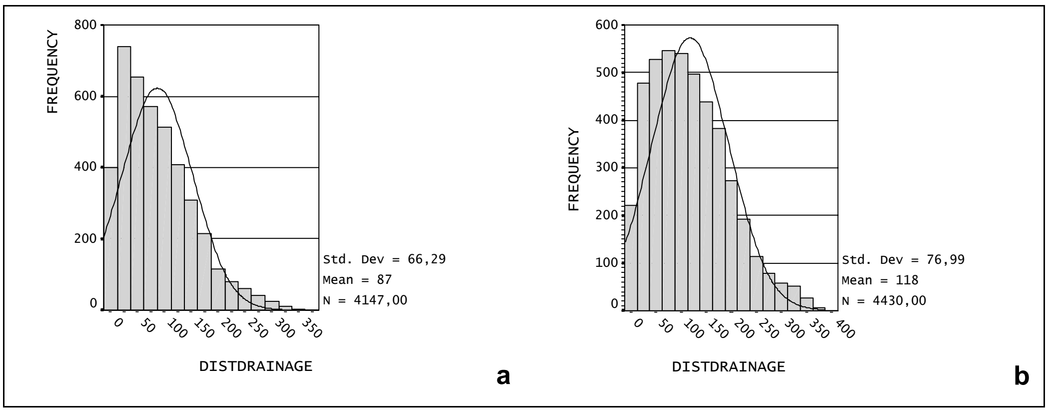


Figure 5.32. The frequency distributions of Distance to Stream map, a. Slided Mass, b. Seed Cells.

5.3.5. Drainage Density

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

Table 5.22. The descriptive statistics of the DRAINAGE DENSITY.

		Slided mass data	Seed cells	Whole Data
DRAINAGE DENSITY				
	Min	58	55	12
	Max	292	287	352
	Mean	176.33	160	160.12
	St.Dev	42.53	37	50.22

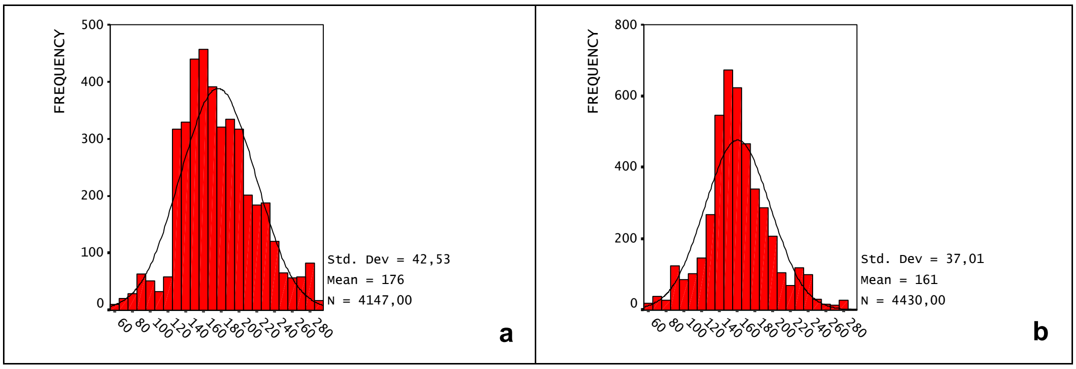


Figure 5.33. The frequency distributions of Drainage Density, a. Slided mass, b. Seed cells.

5.3.6 DISTRIDGE

All of the distance to ridge parameter distributions are similar to each other (whole data & landslide related data), which is an indication of no significant distribution free preference is present. However, a genetic preference is seen in the distributions as both the whole and landslide data exhibit clustering around smaller values. This situation can be summarized as both the area and the landslides do tend to prefer nearer distances to ridges, namely to the microcatchment divide lines. They tend to be located in the upper parts of the slopes. Although the distributions are similar, here are slight changes like, the Distance to Ridge of seed cells seems to have a more sharp peak around 25-100 meters of distance than that of the slided mass data as expected (Figure 5.34). The descriptive statistics and the comparison of the data are shown in Table 5.23.

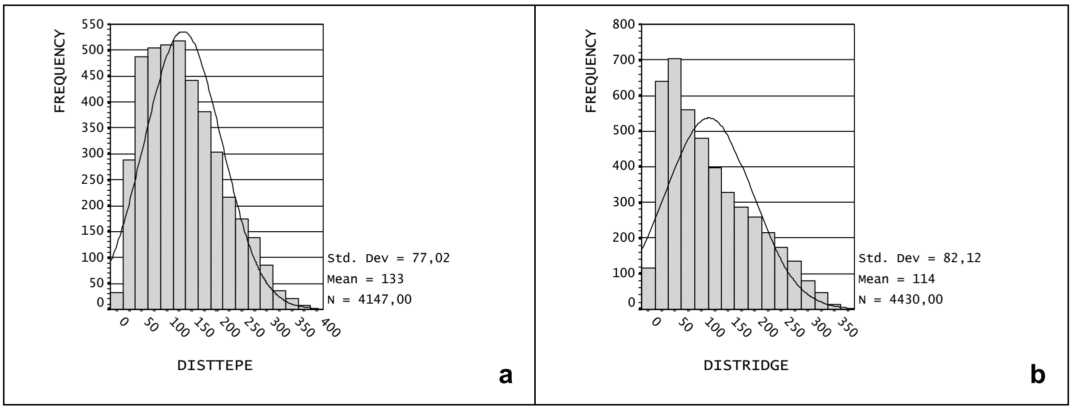


Figure 5.34. The frequency distributions of Distridge, a. Slided mass, b. Seed cells.

Table 5.23. The descriptive stats of distridge

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

5.3.7. Aspect

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

Table 5.24. The descriptive stats of aspect

		Slided mass data	Seed cells	Whole Data
ASPECT				
	Min	-1	-1	-1
	Max	358	358	359
	Mean	156,36	168,37	182,45
	St.Dev	137,07	126,25	110,08

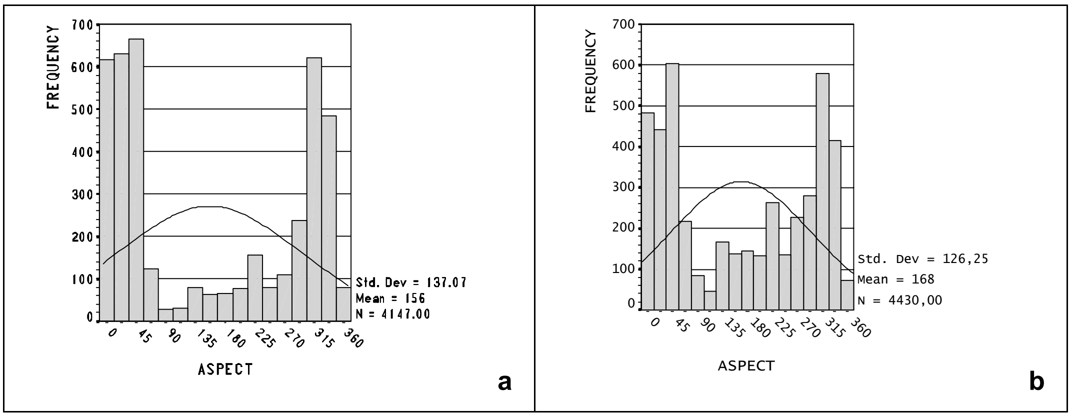


Figure 5.35. The frequency distributions of Aspect, a. Slided mass, b. Seed cells.

5.3.8.Slope

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

Table 5.25. The descriptive stats of Slope

		Slided mass data	Seed cells	Whole Data
SLOPE				
	Min	0	1	0
	Max	52	51	56
	Mean	21,44	22,41	16,97
	St.Dev	8,88	9,12	10,46

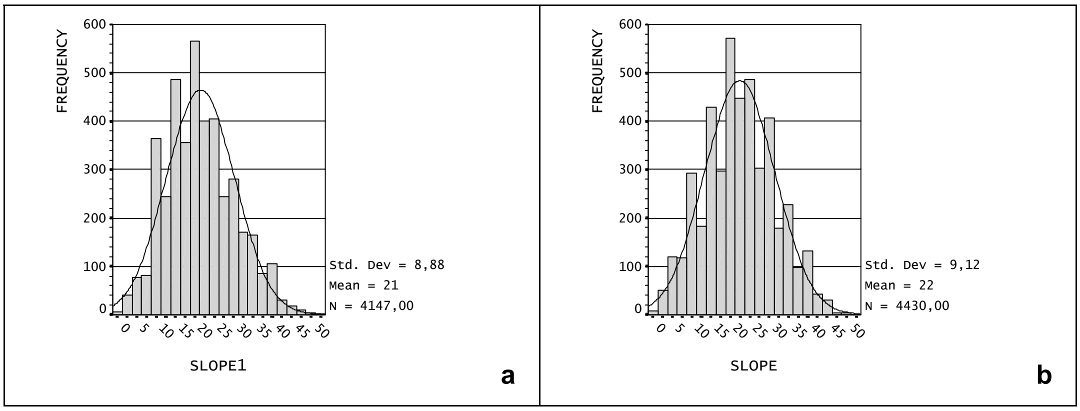


Figure 5.36. The frequency distributions of Slope: a. Slided mass, b. Seed cells.

5.3.9. Distsettlement

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

Table 5.26. The descriptive stats of Distsettlement

		Slided mass data	Seed cells	Whole Data
DISTANCE TO SETTLEMENT				
	Min	2	4	0
	Max	2218	2629	6093,46
	Mean	700,04	699,45	1258,46
	St.Dev	639,12	651,25	1370,20

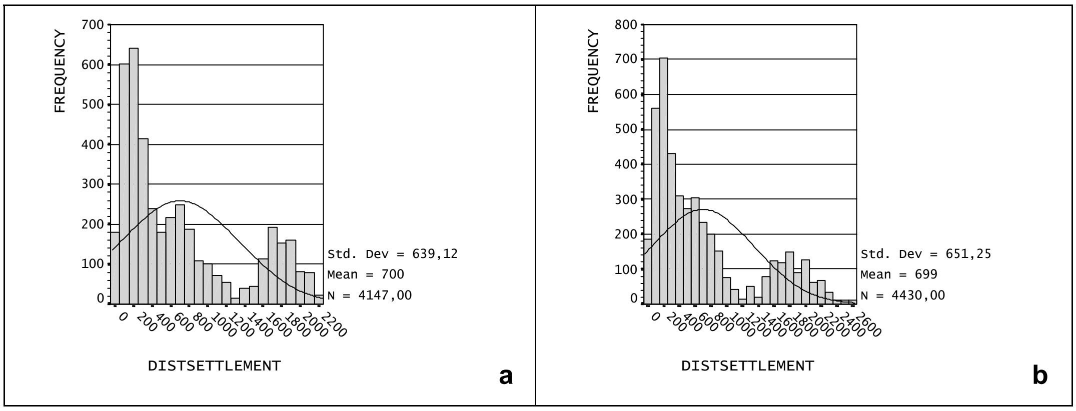


Figure 5.37. The frequency distributions of Distsettlement, a. Slided mass, b. Seed cells.

5.3.10. Distpower & Dist-Roadnetwork

The distance to power lines and road network has also significant low frequency noise in the histogram shown in Figure 4.18.h. The means also suggest a preference in smaller values, which are nearer to power lines and road network. Furthermore, the difference in maximums are quite striking as the whole data range up to 5566 meters but the slided data and seed cells range only up to 700 meters approximately (Table 5.27). The nearly 5000 meters of difference is the source of background noise in the whole data histogram. On the other hand no significant preference is seen in the frequency histograms of the slide related data, only a general clustering is present in the smaller values, which are nearer to power lines and road network. This could be attributed to the fact that the land cover of the region is disturbed during the construction of power lines and the road network. The descriptive stats of the distance to power lines and road network parameter map are presented in (Table 5.27). The frequencies of merged vector of power lines and road network are presented in Figure .5.38.

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

		Slided mass data	Seed cells	Whole Data
DISTANCE TO POWER & ROAD NETWORK				
	Min	0	0	0
	Max	547	683	2312,68
	Mean	184,53	183,78	246,121
	St.Dev	119,31	136,88	302,099

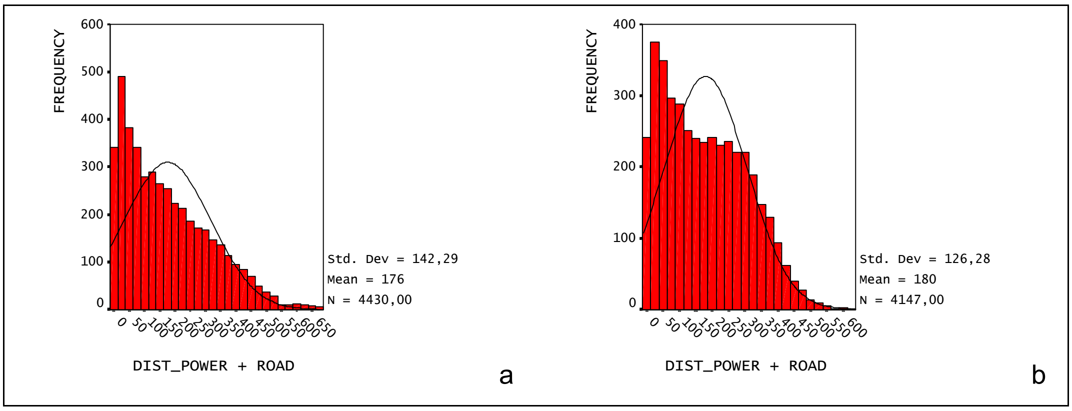


Figure 5.38. The frequency distributions of Dist-power+road a. Slided mass, b seed cells.

5.3.11 Dist E-5 Highway

The E-5 highway is probably the most important and active highway among the Turkey's national highways, which connects İstanbul and Ankara. The high traffic activity of this road creates more vibration and local load than the other roads in the area, based on this the road itself is extracted from the cumulative road network in the area as a separate parameter. Both of the slided mass and seed cell frequency histogram of the distance to E-5 highway reveals important information. The bimodal distribution of these histograms suggests that the landslides prefer two conditions. One group is caused by the presence of E-5 highway as the frequency of seed cells show a large peak in the values smaller than 400 meters. The other group is very distant to the highway, in the order of kilometers, at which the effect of vehicle vibration would be minimal, resulting in the conclusion that the landslides should be caused by other parameters.

Table 5.28. The descriptive stats of Dist_E-5 Road

		Slided mass data	Seed cells	Whole Data
DISTANCE TO E-5				
	Min	0	0	0
	Max	3101	3112	8366,8
	Mean	1254,62	1217,83	2467,49
	St.Dev	737,07	899,59	1771,85

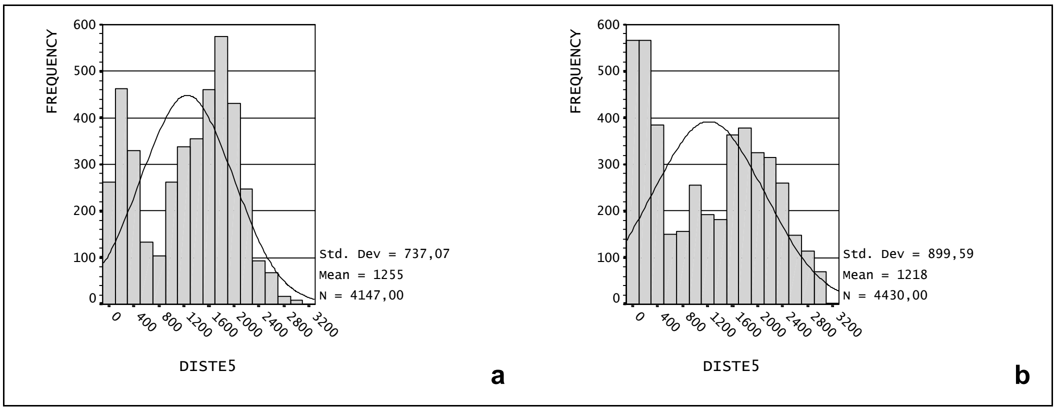


Figure 5.39. The frequency distributions of Distance to E-5 highway, a. Distroad, b. Diste-5

5.3.12. Land cover

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

Table 5.29. The % change of Landcover units

	Whole data	Slided mass	Seed cells
Mixed	33%	42,5%	38,2%
Dense forest	44%	26%	28,1%
Forest	16%	25,4%	26,8%
Total	93%	93,9%	95,1%

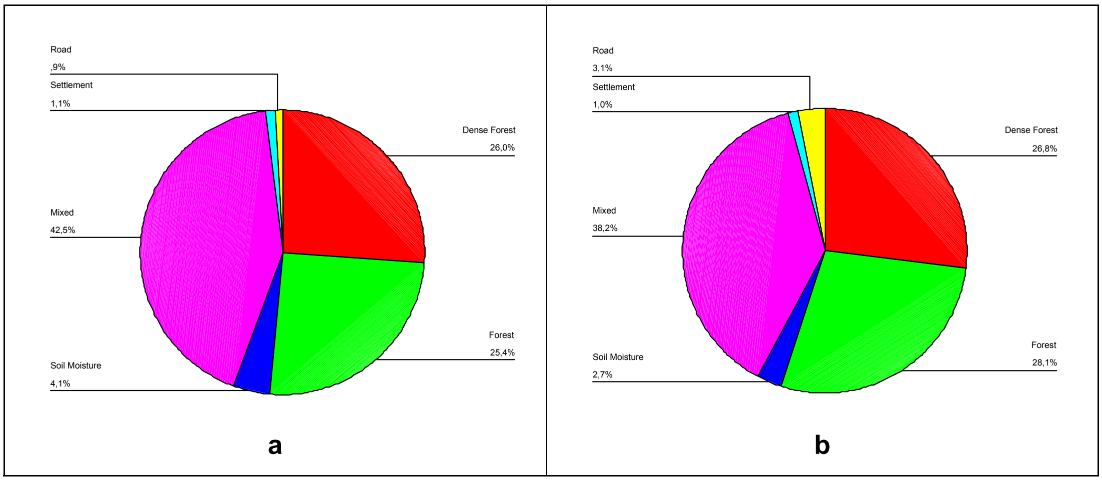


Figure 5.40. The percentage distribution of Landcover units, a.) slided mass, b) seed cells.