ISSN 2249 –5460

INTERNATIONAL JOURNAL OF MATHEMATICAL SCIENCES, TECHNOLOGY AND HUMANITIES

www.internationalejournals.com

Available online at www.internationaleJournals.com

International eJournals

International Journal of Mathematical Sciences, Technology and Humanities 111 (2014) 1195–1208

Multivariate Approach to Earthquake Data

D.Vijaya Laxmi[1] , Jakkula Srinivas[2] & V.V. Hara Gopal[3]

[1 & 2] Dept. of Statistics, Kakatiya University, Warangal,

e-mail: dammavalam_vijaya@yahoo.co.in

[3] Department of Statistics, Osmania University, Hyderabad

ABSTRACT

The voluminous earth quake data sets, covering entire Northern part of India comprised of Himalayan Mountain belt is divided in to 3 parts i.e., North Western (NW), Central (C) and North Eastern (NE), and also the total data samples of entire Himalayan belt, constitute with Latitude, Longitude, Magnitude ($M \geq 4$) and Focal depth(Lt, Ln, M, Fd) are considered for the present study. An attempt is made to analyze the relation between the 4 parameters Lt, Ln, M and Fd using the cluster and Factor analyses. The analyses of monthly resolution earthquake frequency time series suggest that earthquake processes in all three regions evolve on a high dimensional chaotic plane. The significant distinction in the earthquake dynamical patterns seems to be associated with the underlying seism tectonics of these three regions.

Keywords: Multivariate Analysis, Cluster Analysis, Factor Analysis, Interactive Graphs.

AMS Classification: 62H30, 62H25, 62-09

1. Introduction

During the past several years, one of the main themes of the researchers has been the characterization of the nature of earthquakes and its interpretation. Earthquakes (Nicholson, Craig and Wesson, 1990) happen mostly where the earth's tectonic plates collide. Yet earthquakes cannot be predicted accurately enough to know, it is all the harder in the Himalayan region, with hidden underground faults that are poorly monitored by seismic instruments.

The Indian subcontinent has a history of devastating earthquakes. The major reason for the high frequency and intensity of the earthquakes is that India is driving into Asia at a rate of approximately 47mm/year.Geographical statistics of India show that almost 54% of the land is vulnerable to earthquakes.

2. Selection of Earthquake Data

The Himalayas (Tiwari and Sri Lakshmi, 2005) are tectonically one of the most complex and seismically active regions of the world. The occurrences of large and small frequent earthquakes here reflect long range interactions of mega tectonic units. The bends in the Himalayan tectonic zone, the Eastern and Western edges are the main locations of the complex tectonics. Evidence also shows localized vertical movement in this region and small earthquakes are most common. Many major earthquakes of differing size that have occurred during the past centuries dominate the seismicity of the Himalayan region.

The whole Himalayas covering 20-38^0 N and 70-98^{0}E has been approximately divided into three zones: (i) Central (28-38^{0}N Lat and 78-98^{0}E Long) (ii) North-East (20 - 28^0 N Lat and 88 - 98^0 E Long) and (iii) Western Himalayas (30-38^{0}N Lat and 70-78^{0}E Long) for the period 1973-2003.

3. Analysis of Himalayan Data

An attempt is made to analyze the relation between four parameters Lt, Ln, M and Fd using Factor and cluster analyses. Cluster analysis sorts out different objects into homogeneous groups in a way the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. In other words, cluster analysis identifies structures in data without explaining why they exist. Further, to substantiate this study the data sets are analyzed using Factor analysis technique which attempts to identify the underlying factors that explain pattern of correlations with a set of observed variables.

4. Interactive Graphs

Scatter Plots

We first plotted the original data as a scatter plot in two and three dimensions to visualize the relationships between the variables and found that the points fall near a line or a well defined curve. In three dimensions, the points may fall near a surface.

This study of the scatter plot can help us to develop a mathematical pattern of relationship. Points that do not fit the relationship stand out in the plot and cautions out that we should investigate them further.

Here, we have plotted 2D and 3D graphs with variables Latitude(Lt), Longitude(Ln), Focal depth(Fd) and Magnitude(M) of Central Himalaya, North-East Himalaya, Western Himalaya, and Himalayas (total). But here we are showing 2D and 3D graphs of Central Himalaya and 2D and 3D graphs of North-East Himalaya, Western Himalaya, and Himalayas (total) are drawn in the similar manner and how they are appearing is written below.

Figure (4a): 2D Interactive Graphs of Central Himalaya:

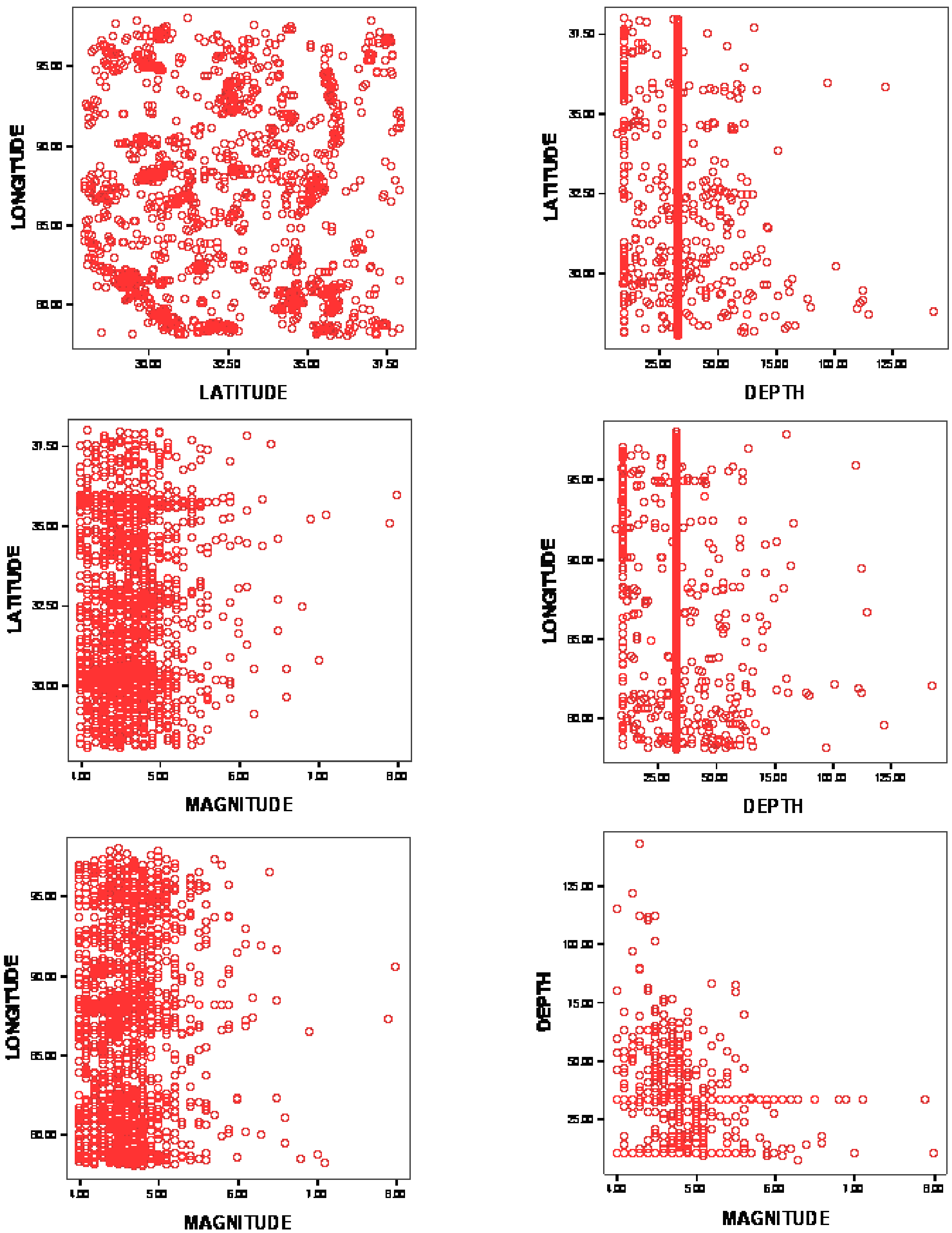
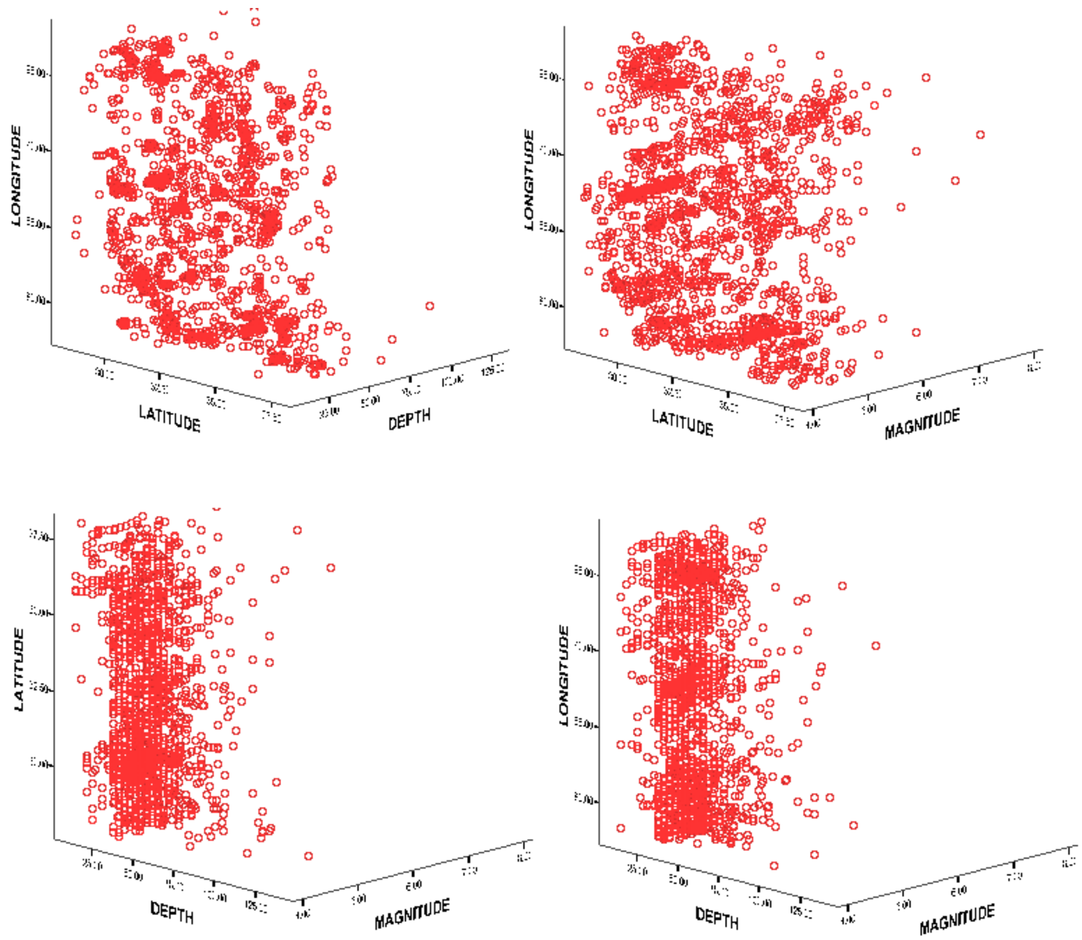


Figure (4b): 3D Interactive Graphs of Central Himalaya:



In figure (4a) 2D interactive graphs of central Himalaya we observe that the plot of Lt against Ln is well scattered, in the plot of M against Lt all the points appear to be very close at M=5 and this is similar to the plot of M against Ln.

For the plots of Fd against Lt and Fd against Ln the points formed a straight line nearly at Fd 25 and appear to be similar. And for plot M against Fd the points are very close at M=6 and Fd=75.

In figure (4b) 3D interactive graphs of central Himalaya we observe that the plots of Lt, Ln, Fd and Lt, Ln, M appear to be similar. While the plots of Fd, Lt, M and Fd, Ln, M are similar.

In 2D interactive graphs of North East Himalaya we observed that the plot of Lt against Ln the points are dense at Ln=95, in the plot of M against Lt all the points appear to be very close at M=5 and this is similar to the plot of M against Ln.

For the plots of Fd against Lt and Fd against Ln the points formed a straight line nearly at Fd 40 and appear to be similar. And for plot Fd against M the points are very close at M=6.

In 3D interactive graphs of North East Himalaya we observed that the plots of Lt, Ln, Fd and Lt, Ln, M appear to be similar. While the plots of Fd, Lt, M and Fd, Ln, M are similar.

In 2D interactive graphs of Western Himalayan we observed that the plot of Lt against Ln the points are highly concentrated at Lt=35, in the plot of M against Lt the points are very close at Lt between 36-38 and it is similar in opposite direction for the graph M against Ln.

For the plots of Fd against Lt and Fd against Ln the points formed a straight line nearly at Fd 25 and appear to be similar in opposite directions. And for plot Fd against M the points are very close at M=5 and Fd less than 200.

In 3D interactive graphs of Western Himalayan we observed that the plots of Lt, Ln, Fd and Lt, Ln, M appear to be similar. While the plots of Fd, Lt, M and Fd, Ln, M are similar in opposite direction.

In 2D interactive graphs of Himalayas (Total) we observed that the plot of Lt against Ln are highly concentrated at Lt =35 and Ln=90, in the plot of M against Lt All the points appear to be very close at M=5 and this is similar to the plot of M against Ln.

For the plots of Fd against Lt and Fd against Ln the points formed a straight line and appear to be similar in opposite directions. And for plot M against Fd the points are very close at M=6 and Fd=275.

In 3D interactive graphs of Himalayas (Total) we observed that the plots of Lt, Ln, Fd and Lt, Ln, M appear to be similar. While the plots of Fd, Lt, M and Fd, Ln, M are similar.

5. Cluster Analysis for Himalayan Data

Cluster analysis (Hair, Anderson, Tatham and Black, 1998) can be used to discover structures in data without providing an explanation or interpretation. In other words cluster analysis simply discovers structures in data without explaining why they exist.

Most commonly used clustering algorithms can be classified into two general categories: (1) Hierarchical and (2) nonhierarchical. Here we consider nonhierarchical clustering for the Himalayas. Nonhierarchical procedures assign objects into clusters once the number of clusters to be formed is specified. These procedures are frequently referred to as K-means clustering.

K-Means Clustering

K-means methodology is a commonly used clustering technique. In this analysis the user starts with a collection of samples and attempts to group them into k number of clusters based on certain specific distance measurements.

The k-means algorithm:

The k-means algorithm assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster i.e. its coordinates are the arithmetic mean of each dimension separately over all the points in the cluster.

The algorithm steps are:

1. Choose the number of clusters, k.
2. Randomly generate k clusters and determine the cluster centers, or directly generate k random points as cluster centers.
3. Assign each point to the nearest cluster center.
4. Recompute the new cluster centers.
5. Repeat the two previous steps until some convergence criterion is met.

The main advantage of this algorithm is its simplicity and speed which allows it to run on large data sets.

Let us briefly go through the different stages of k-means cluster analysis for Central Himalaya, North-East Himalaya, Western Himalaya, and Himalayas (total) data. Firstly we determine number of clusters to be 20 and the initial cluster centers are evaluated.

The initial cluster centers are given in Table 5(a). They are vectors with their values based on the four variables Latitude, Longitude, Depth and Magnitude. In Table 5(b) we can see the number of iterations and changes in the cluster centers. In Table 5(c) we can see the final cluster centers. Table 5(d) presents data for the number of units in each cluster as well as their total number and missing units (if there are any).

The following table is the Cluster Analysis of Central Himalaya and the same procedure is carried for North-East Himalaya, Western Himalaya, and Himalayas (total) data.

Cluster Analysis of Central Himalaya

Table- 5(a): Initial Cluster Centers

Clusters	Initial Cluster Centers			
	LATITUDE	**LONGITUDE**	**DEPTH**	**MAGNITUDE**
1	28.85	95.91	110	4.4
2	31	78.07	21	4.5
3	36.69	83.44	10	4.3
4	32.13	95.45	62	4.5
5	37.06	96.48	20	5.5
6	35.81	79.5	122	4.2
7	30.33	79.13	71	4.5
8	29.63	95.65	09	5.1
9	29.27	80.28	58	4.6
10	30.16	82.14	101	4.5
11	28.66	86.58	115	4.0
12	28.76	81.95	143	4.3
13	30.47	79.2	44	4.9
14	28.01	87.73	33	4.1
15	29.66	97.8	80	4.4
16	30.3	94.88	44	4.8
17	35.84	88.17	53	4.0
18	28.9	81.36	90	4.3
19	37.04	97.84	33	5.0
20	37.19	78.08	33	4.6

Table- 5(b): Iteration History

Clusters	Change in Cluster Centers Iterations									
	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	3.004	0.496	0.185	0.405	0	0	0	0	0	0
3	3.685	1.778	0.174	0	0	0	0	0	0	0
4	3.027	0.884	0	0.688	0	0	0	0	0	0
5	4.171	1.012	0.834	0.82	0.772	0.463	0.316	0.2	0.331	0
6	0	0	0	0	0	0	0	0	0	0
7	3.862	1.36	0.953	1.046	0.821	0	0	0	0	0
8	5.059	0.554	0.114	0	0.094	0	0.062	0	0.065	0
9	2.924	0.768	0.696	0.506	0	0	0	0	0	0
10	4.072	0	0	0	0	0	0	0	0	0
11	3.076	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0
13	2.155	0.373	0.191	0.215	0	0	0	0	0	0
14	3.115	0.422	0.171	0.065	0.045	0.019	0	0	0	0
15	6.082	1.121	0	1.199	0.909	0	0	0	0	0
16	2.485	1.192	0.274	0	0	0	0	0	0	0
17	4.112	0.801	0.394	0	0	0	0	0	0	0
18	3.347	1.697	0	0	1.802	0	0	0	0	0
19	5.494	0.633	0.036	0.06	0.032	0.03	0	0.015	0	0
20	3.726	1.072	0.161	0.034	0.021	0	0	0	0	0

Table- 5(c): Final Cluster Centers

Clusters	Final Cluster Centers			
	LATITUDE	LONGITUDE	DEPTH	MAGNITUDE
1	28.85	95.91	110	4.4
2	32.91	80.49	21.2	4.97
3	31.98	82.69	11.49	4.8
4	30.67	93.03	60.4	4.55
5	30.81	93.43	23.64	4.87
6	35.81	79.5	122.0	4.2
7	30.66	82.2	66.06	4.64
8	34.62	93.24	10.47	4.82
9	33.18	79.73	56.69	4.6
10	33.07	80.11	99.0	4.35
11	29.07	84.86	112.5	4.3
12	28.76	81.95	143.0	4.3
13	32.16	80.15	45.79	4.74
14	31.67	87.62	33.06	4.56
15	29.94	90.79	77.13	4.86
16	31.34	91.58	44.35	4.66
17	31.15	86.57	52.57	4.54
18	29.01	81.76	83.2	4.32
19	31.82	94.63	33.05	4.67
20	32.84	80.5	33.0	4.61

Table- 5(d): Number of cases in each cluster

Clusters	Number of Cases
1	1.00
2	35.00
3	63.00
4	10.00
5	25.00
6	1.00
7	18.00
8	126.00
9	26.00
10	2.00
11	4.00
12	1.00
13	34.00
14	481.00
15	8.00
16	20.00
17	14.00
18	5.00
19	364.00
20	453.00
Valid	1691
Missing	6.00

In Central Himalaya, a convergence achieved due to no or small change in the cluster centers. The maximum absolute in coordinate change for any center is 0.000. The current iteration is 10. The minimum distance between initial centers is 10.595

In North-East Himalaya, a convergence achieved due to no or small change in the cluster centers. The maximum absolute in coordinate change for any center is 0.000. The current iteration is 10. The minimum distance between initial centers is 10.286.

In Western Himalaya, iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is 0.860. The current iteration is 10. The minimum distance between initial centers is 16.138.

In Himalayas (total), iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is 5.500. The current iteration is 10. The minimum distance between initial centers is 26.023.

Thus, in the iteration history tables of Central Himalaya and North East Himalaya, it is observed that the cluster centers seem to be same indicating a similar pattern with respect to cluster centers whereas, in the tables of Western Himalaya and Himalayas(total) it does not seem to have any pattern with respect to cluster centers. Thus, the same data is analyzed through Factor analysis to know the patterns of earthquakes with respect to four variables considered.

6. Factor Analysis for Himalayan Data

Factor analysis (Lawley and Maxwell, 1962) is a statistical method used to describe variability among observed variables in terms of fewer unobserved variables called factors. The observed variables are modeled as linear combinations of the factors and error terms. The information gained about the interdependencies can be used later to reduce the set of variables in a data set of earthquake data.

Communality: The proportion of variance of a particular item that is due to common factors (shared with other items) is called communality.

Now Principal component analysis is used as a tool in exploratory data analysis and for making predictive models. And it is closely related to factor analysis.

Table-1 is the table of communalities before and after extraction. Principal component analysis works on the initial assumption that all variance is common, therefore, before extraction the communalities are all 1. The communalities in the column labeled Extraction reflect the common variance in the data structure. The amount of variance in each variable that can be explained by the retained factors is represented by the communalities after extraction.

Table-2 lists the eigen values associated with each linear component (factor) before extraction, after extraction. We have identified four linear components with in the data set. The eigen values associated with each factor represent the variance explained by that particular linear component. And the eigen value in terms of the percentage of variance. Now extract all factors with eigen values greater than 1, which leaves us with two factors.

The eigen values associated with these factors are displayed in the columns labeled Extraction Sums of Squared Loadings. The values in this part of the table are the same as the values before extraction, except that the values for the discarded factors are ignored.

Table-3 is the component matrix. This matrix contains the loadings of each variable onto each factor.

Factor Analysis of Central Himalaya

Table 6(a): With four components: Latitude, Longitude, Depth and Magnitude

TABLE-1: Communalities

	Initial	Extraction
LATITUDE	1.000	.707
LONGITUDE	1.000	.795
DEPTH	1.000	.606
MAGNITUDE	1.000	.266

Extraction Method: Principal Component Analysis.

TABLE-2

Total Variance Explained

Component	Initial Eigen values			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.304	32.593	32.593	1.304	32.593	32.593
2	1.070	26.749	59.342	1.070	26.749	59.342
3	.924	23.092	82.434			
4	.703	17.566	100.000			

Extraction Method: Principal Component Analysis.

TABLE-3

Component Matrix (a)

	Component	
	1	2
LATITUDE	.546	-.639
LONGITUDE	.372	.810
DEPTH	-.775	-.072
MAGNITUDE	.515	-.017

Extraction Method: Principal Component Analysis.

a 2 components extracted.

Table 6(b): <u>with three components</u>: Latitude, Longitude and Depth

TABLE-1
Communalities

	Initial	Extraction
LATITUDE	1.000	.778
LONGITUDE	1.000	.822
DEPTH	1.000	.696

Extraction Method: Principal Component Analysis.

TABLE-2

Total Variance Explained

Component	Initial Eigen values			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.226	40.871	40.871	1.226	40.871	40.871
2	1.070	35.664	76.535	1.070	35.664	76.535
3	.704	23.465	100.000			

Extraction Method: Principal Component Analysis.

TABLE-3

Component Matrix (a)

	Component	
	1	2
LATITUDE	.598	-.648
LONGITUDE	.420	.804
DEPTH	-.832	-.060

Extraction Method: Principal Component Analysis.

a 2 components extracted.

Table 6(c): <u>with two components</u>: Latitude and Longitude

TABLE-1: Communalities

	Initial	Extraction
LATITUDE	1.000	.536
LONGITUDE	1.000	.536

Extraction Method: Principal Component Analysis.

TABLE-2

Total Variance Explained

Component	Initial Eigen values			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.071	53.565	53.565	1.071	53.565	53.565
2	.929	46.435	100.000			

Extraction Method: Principal Component Analysis.

TABLE-3

Component Matrix (a)

	Component
	1
LATITUDE	.732
LONGITUDE	-.732

Extraction Method: Principal Component Analysis.

a 1 components extracted.

From the above tables we observe that Factor analysis of Central Himalaya with four components Latitude (Lt), Longitude (Ln), Focal depth (Fd) and Magnitude (M) are reduced to two components. And with three component combinations Lt, Ln, Fd and Lt,Ln, M are reduced to two components, while the combinations Lt, Fd, M and Ln, Fd, M are reduced to only one component. And for different combinations of two components are reduced to one component.

It is observed that Factor analysis of North East Himalaya with four components Latitude (Lt), Longitude (Ln), Focal depth (Fd) and Magnitude (M) are reduced to two components. And with three component combinations Ln, Fd, M and Lt,Ln, M are reduced to two components, while the combinations Lt, Ln, Fd and Lt, Fd, M are reduced to only one component. And for different combinations of two components are reduced to one component.

It is observed that Factor analysis of Western Himalaya with four components Latitude (Lt), Longitude (Ln), Focal depth (Fd) and Magnitude (M) are reduced to two components. And with all the combinations of three components are reduced to only one component. And also for different combinations of two components are reduced to one component.

It is observed that Factor analysis of Himalayas (Total) with four components Latitude (Lt), Longitude (Ln), Focal depth (Fd) and Magnitude (M) are reduced to two components. And with all the combinations of three components are reduced to only one component. And also for different combinations of two components are reduced to one component.

The present analysis indicate that there is a pattern among the communalities with respect to the four parameters considered that is all the three parts and total Himalayas formulate into the same pattern of communalities.

7. Conclusions

From the analysis of the data we determine the relation between the four parameters Latitude(Lt), Longitude(Ln), Magnitude(M), Focal depth(Fd), for which the cluster and factor analysis techniques are used. And it is determined that the total number of clusters is 20 in each zone (NW, C, and NE) and also for the entire area, respectively. It is observed that the distance from the centre point to each cluster (within a group of 20 clusters) varies for all the 3 zones and as well for the entire Himalayas for all the parameters Lt, Ln, M and Fd. From Factor analysis (for the 4 parameters) it is noticed that NE and NW follow the same pattern of Commonalities M,Ln,Fd and Lt in descending order, whereas the Central Himalayas and total Himalayas follow the similar pattern of commonalities in Ln,Lt,Fd and M .It is evident that from Factor analysis the Longitude (Ln) is a Common factor, which is dominant, extracted from the data sets of Central and Total Himalayas though independently. However, the Magnitude parameter is the Common factor which is dominant, extracted from data sets of Western and North Eastern Himalayas though independently. The most significant inference that is drawn from this analysis is that it could be possible only when the stress is applied on opposite directions that is, North western and North Eastern Parts of Himalayan zones are compressing against each other as a result the distribution could take place along the Longitudinal direction. Further it also correlates with the crustal thicknesses in the central portion of the Himalayas is more compared to the neighboring North Eastern and North Western blocks. It is like by when the stress is applied on opposite directions of a clay ball that would not only make the clay ball elongate in perpendicular directions of forces but also the thickness would increase in the central portion.

REFERENCES:

1. **Nicholson, Craig and Wesson, R.L. (1990)**, “Earthquake Hazard Associated with Deep Well Injection”, A Report to the U.S. Environmental Protection Agency, U.S. Geological Survey Bulletin, 1951, 74 p.

2. **Tiwari, R.K. and Sri Lakshmi, S. (Feb 2005),** “Some Common and Contrasting Features of Earthquake Dynamics in Major Tectonic Zones of Himalayas using Nonlinear Forecasting Approach”, Current Science, vol. 88, No. 4, pp. 640-647.

3. **Joseph, F. Hair, JR. Rolph E. Anderson, Ronald L. Tatham, Wlliam C. Black. (1998),** “Multivariate Data Analysis”, 5th edition by Pearson Education India Pvt. Ltd. New Delhi, ISBN 81-7758-573-8.

4. **Lawley, D. N and Maxwell, A. E. (1962),** “ Factor Analysis as a Statistical Method”, Journal of the Royal Statistical Society, Series D, vol.12, No. 3, pp. 209-229.

5. **Tiwari, R.K and Rao, K.N.N. (2001),** J.Geol. Soc. India, 57, 369-376.

6. **Pandey, M, Tandukar, R.P, Avovac, J.P, Lave, J, Massot, J.P. (1995),** Geophys. Res. Lett., 22,751.

7. **Gupta, H.K, Rajendran, K, and Singh, H.N. (1986),** J. Geol. Soc. India,28, 345-365.

8. **Saraf.A.K, Mishra, P, Mitra, S, Sarma, B, and Mukhopadyay, D.K. (2002),** Remote Sensing and GIS Technologies for improvements in Geological structures interpretation and mapping. International Journal of Remote Sensing.

9. **Marzocchi, W, Mulargia, F, and Gonzato, G. J. (1997),** Geophys, Res., 102, 3195-3209.

10. **Khattri, K, Wyss, M, Gaur, V.K, Saha, S.H. and Bansal, V.K. (1983),** Bull.Seism.Soc.Am, 73,459-469.