Multidimensional scaling technique for analysis of magnetic storms at Indian observatories

M Sridharan[1] and A M S Ramasamy[2]

[1]*IIG Magnetic Observatory, Pondicherry 605 014, India.*
[2]*Professor of Mathematics, Pondicherry University, Pondicherry 605 014, India.*

Multidimensional scaling is a powerful technique for analysis of data. The latitudinal dependence of geomagnetic field variation in horizontal component (H) during magnetic storms is analysed in this paper by employing this technique.

1. Introduction

Multidimensional Scaling (MDS) comprises a set of models and associated methods for constructing a geometrical representation of proximity and dominance relationship between elements in one or more sets of entities. MDS can be applied to data that express two types of relationships: proximity relations and dominance relations. In proximity data, the data values indicate the proximity (similarity or dissimilarity) between the entities to which their indices refer. In dominance data, the data values indicate how strongly one entity dominates the other. In this paper, multidimensional reduction technique is applied to analyse magnetic storms recorded at the following 7 observatories from 1985 to 1993 where common data are available: Trivandrum, Kodaikanal, Annamalainagar, Hyderabad, Alibag, Ujjain and Sabhawala. It is worthwhile to recollect some of the previous findings of researchers, which will be relevant to the present study. They are as follows: The daily variation of horizontal magnetic field H at an equatorial station shows significant differences during low and high sunspot years (Rastogi and Patil 1992). It has been pointed out that the pattern of geomagnetic field variation at equatorial latitudes of Trivandrum, Kodaikanal and Annamalainagar differs significantly from that of higher latitudes (Alex and Rao 1995). The Storm Sudden Commencement (SSC) in X Y and Z (geomagnetic field variations with respect to geographic coordinates) components at a large number of stations round the world shows that the amplitude of storm sudden commencement in X and Y varies in a regular fashion with geomagnetic latitudes (Obayashi and Jacob 1957). Even during an intense solar flare effect (SFE) and storm sudden commencement event, the amplitude of H decreases progressively with increasing latitudes at the Indian chain of observatories (Rastogi *et al* 1997). The aim of this study is to apply the method of multidimensional scaling technique to examine the accuracy of results in comparison with the conventional method of correlation coefficients in the analysis of the latitudinal dependence of geomagnetic field variation (H) during the storm days at the 7 Indian observatories and to verify the results with the existing theories.

2. Locations of the observatories

The geographic and dipole coordinates of the observatories are provided in table 1 and figure 1. During the years 1985–1993, data for the 7 observatories were available only for 40 storm days. The relevant data pertaining to the H-variation during the storm periods in the 7 stations are provided in table 2.

Keywords. Multidimensional scaling; covering distance; storm time range.

Table 1. *Geographic and dipole co-ordinates of observatories.*

	Geographic		Dipole	
Station	Lat. (N)	Long. (E)	Lat.	Long.
Trivandrum	8.48 deg	76.97 deg	0.8 S	148.5 deg
Kodaikanal	10.23	77.47	0.9 N	149.1
Annamalainagar	11.37	79.68	1.4 N	149.4
Hyderabad	17.41	78.55	7.6 N	148.9
Alibag	18.61	72.87	9.7 N	145.6
Ujjain	23.18	47.78	14.0 N	148.8
Sabhawala	30.37	77.80	20.9 N	151.5

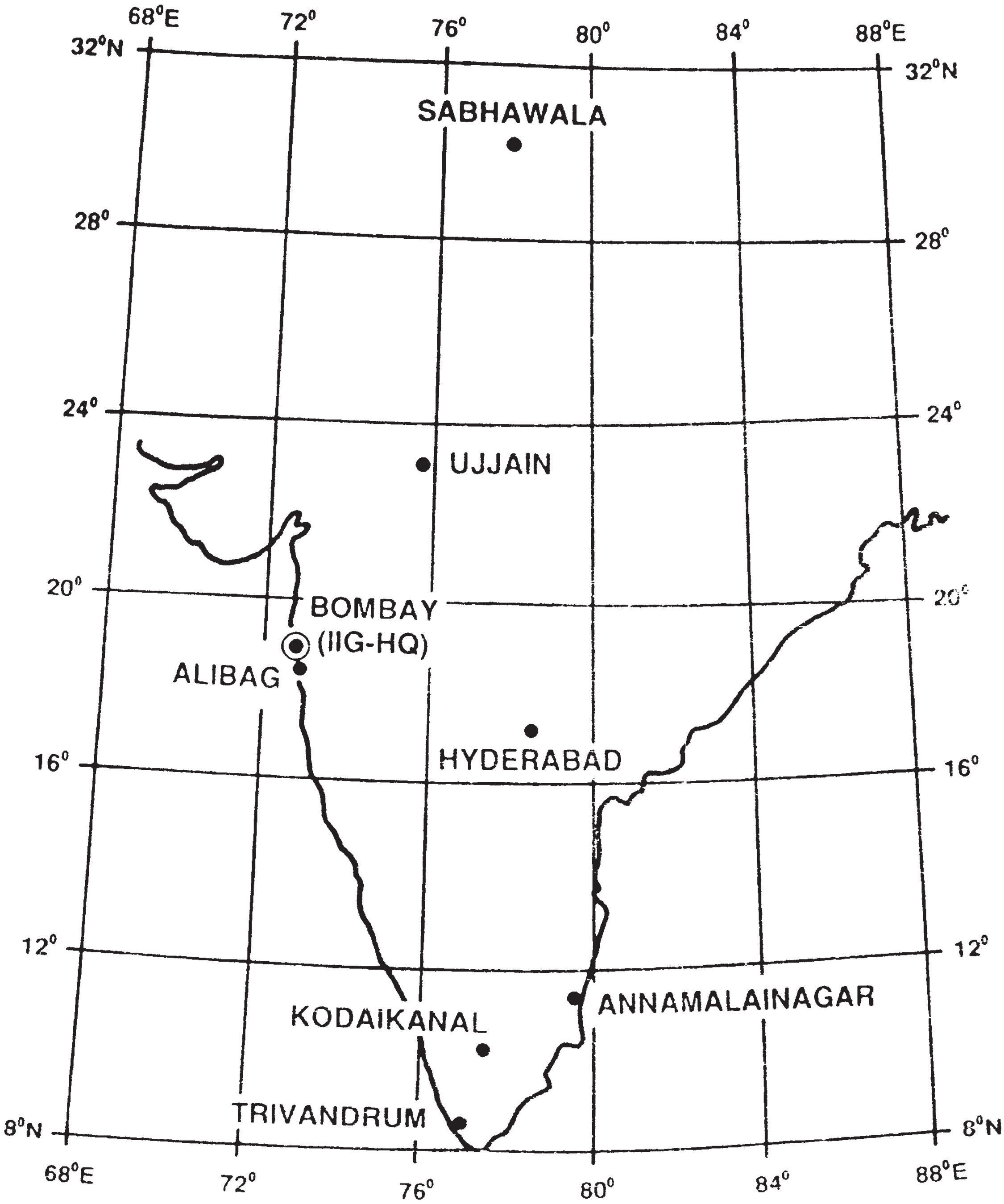

Figure 1. Location map of the geomagnetic observatories whose data are included in this paper.

Table 2. *Principal magnetic storms from 1985 to 1993 amplitude in nT for H-variations.*

Serial no.	Date	TRD	KOD	ANN	HYB	ABG	UJJ	SAB
1	08.01.85	180	160	166	134	123	122	134
2	05.02.85	197	194	175	147	133	117	144
3	19.04.85	268	278	288	275	256	255	270
4	30.04.85	250	253	259	265	237	258	269
5	03.07.85	190	185	186	168	154	151	150
6	12.08.85	151	130	115	108	101	102	126
7	14.09.85	188	177	172	147	140	117	121
8	29.11.85	186	177	178	177	170	159	176
9	12.12.85	202	197	160	175	162	100	159
10	18.12.85	225	216	227	137	162	132	77
11	06.01.86	126	124	130	131	129	129	139
12	25.01.86	140	139	154	137	122	113	110
13	20.08.86	187	183	155	144	129	131	68
14	11.09.86	263	251	215	228	217	228	273
15	30.11.88	305	286	291	264	250	241	271
16	17.12.88	177	101	120	104	98	84	115
17	20.01.89	361	381	378	374	359	378	376
18	18.09.89	329	320	349	330	304	197	318
19	29.12.89	257	167	225	183	172	157	159
20	13.02.90	294	231	212	83	80	96	118
21	12.06.90	322	282	322	236	250	225	196
22	01.02.91	167	138	125	107	103	100	106
23	31.05.91	293	294	292	224	208	186	183
24	17.06.91	192	173	191	136	139	162	220
25	01.10.91	226	201	184	170	175	186	231
26	08.11.91	539	532	383	446	412	395	421
27	01.01.92	127	96	109	99	95	100	117
28	02.02.92	277	231	220	175	163	159	154
29	20.02.92	422	576	501	329	304	271	268
30	29.02.92	250	236	232	203	189	175	179
31	22.05.92	341	312	286	249	227	203	99
32	10.06.92	143	239	155	132	120	104	100
33	18.06.92	151	138	107	130	103	101	117
34	04.08.92	181	155	142	129	125	137	149
35	09.09.92	429	402	390	347	303	290	319
36	08.10.92	355	327	296	261	238	219	226
37	17.02.93	350	335	291	333	308	283	295
38	11.03.93	308	284	231	246	230	208	201
39	23.03.93	214	220	173	116	105	107	130
40	04.04.93	225	254	240	244	227	219	247

Abbreviations used for the observatories:
TRD–Trivandrum, KOD–Kodaikanal, ANN–Annamalainagar, HYB– Hyderabad, ABG–Alibag, UJJ–Ujjain, SAB–Sabhawala.

3. Multidimensional scaling – analytic technique

For the purpose of analysis of data, spanning distances between one station and the remaining stations in respect of the ranges of H variations during magnetic storms are computed.

There are many practical situations where various independent factors come into play and influence the behavior of a dependent factor. If the values assumed by these independent factors at a particular period of time for a specific situation are put as an array, we get a point in a space of several dimensions. A drawback in a space of several dimensions is that the points in such a space and the relationship between them cannot be represented geometrically in a form enabling a person to visualize. To obviate this difficulty, one may use multidimensional reduction technique. With the help of this technique, one can reduce the dimensions and finally represent the data in a tree structure so that this structure represents the relationship between the data-points in a two-dimensional sense. A spanning sub graph of a directed (or undirected) graph is said to be a tree if and only if it is connected and contains no circuits.

Each magnetic storm recorded at Indian magnetic observatories from 1985 to 1993 was represented as a point in a multidimensional space of which one dimension was allocated for each of the 7 stations. It is not possible to visualize geometrically such a higher dimensional space for the purpose of understanding the concept of representation of storms by means of points in that space. However it is worth commenting that just as we can compute the distance between two points in a two-dimensional plane or in a three-dimensional space, the distance in an n-dimensional Euclidean space between two points with coordinates $(X_{11}, X_{12}, \ldots, X_{1n})$ and $(X_{21}, X_{22}, \ldots, X_{2n})$ can be calculated from the following well known formula:

$$D = \text{square root of } \{(X_{11} - X_{21})^2 + (X_{12} - X_{22})^2 + \ldots + (X_{1n} - X_{2n})^2\}. \quad (1)$$

This distance is a number, which tends to be small if the values of the corresponding coordinates of the two points are similar in magnitude and large if the points differ significantly in several coordinates. Therefore the computed distance between two points can be used as a measure of similarity or dissimilarity between the ranges of two storms at two given points at a time. If a series of points represent the sequence of storm intensity of a single station in time, the path or trajectory connecting the points is indicative of the station's storm-time intensity.

Using formula (1) distances can be computed between one point and all other variable points in the n-dimensional space. Next the distances are ranked in ascending order. This enables one to find out which point is closest to which other point, the second closest point and so on. A minimum coverage algorithm is employed to connect each point to at least one nearest neighbour. This enables one to construct a branching network that ties together all of the available points. Since a tree has no cycles, one has to avoid cycles in the construction of a network. If only points are connected which have not in some previous steps been connected to the same network; the final network will be a tree. It is not really possible to display accurately an n-dimensional network on a two-dimensional plane. However, by stretching, bending and twisting the arcs connecting adjacent points, it is usually possible to locate the points on a plane, so that most of the near neighbours of each point are closer to it, while points which are not its neighbours tend to be farther away. The resulting "road map" is at best an approximation of the real situation but still it may contain a surprising amount of useful information in a highly compressed form. The minimum coverage algorithm is utilised to indicate the nearest neighbor of each point and then to identify successively the shortest possible connections between pairs of points until all points are incorporated in a single network. Figure 2 is drawn with minimum coverage connections between two points, that are fairly neighbours of each other. The distances between the pairs of points and the sum of the distances is calculated. This sum is called the **covering distance** of the tree because the tree covers all the points with minimum possible distances between the pairs of points.

4. Dimension reduction technique

The way in which the points in the n-dimensional space have been connected by a tree will be maintained for the further process described below.

Now, instead of n coordinates, let us choose any two coordinates and determine the distances between the points as restricted to these two coordinates. These distances will be marked on the lines joining the points in the tree structure. The sum of the distances is calculated. This sum gives the covering distance for all the points in the graph, in the two-dimensional sense. This procedure is repeated for each pair of coordinates in the n-dimensional space. For each pair of coordinates, we consider the tree structure and the corresponding covering distance in the two-dimensional sense. We will have $n(n-1)/2$ values of such covering distance. The covering distances for all possible trees are put in ascending order. With the help of this ordering, we can find out which two factors are approximately nearby, compared to the remaining factors, in the two-dimensional sense. Nearness of two points in the two-dimensional sense does not mean that they are geographically nearby. But it indicates that they possess almost similar characteristics with respect to a certain feature. The covering distances for the 21 possible trees are provided in table 3. Among all possible trees, the one with the minimum covering distance is chosen with the help of table 3 and represented by figure 3. The coefficients of correlation between pairs of stations for H variation on the storm days are also furnished in table 3.

5. Findings of the study

For the purpose of analysis of data, spanning distances between one station and the remaining stations in respect of the ranges of H variations during magnetic storms are computed. It is represented in table 3.

From table 1 and figure 3, certain inferences are drawn as follows. In the case of Trivandrum and Annamalainagar, they are geographically close to each other. However, according to multidimensional scaling technique, they are found to be at a large distance with respect to their storm time range.

The spanning distances of the 3 observatories Annamalainagar, Kodaikanal and Trivandrum

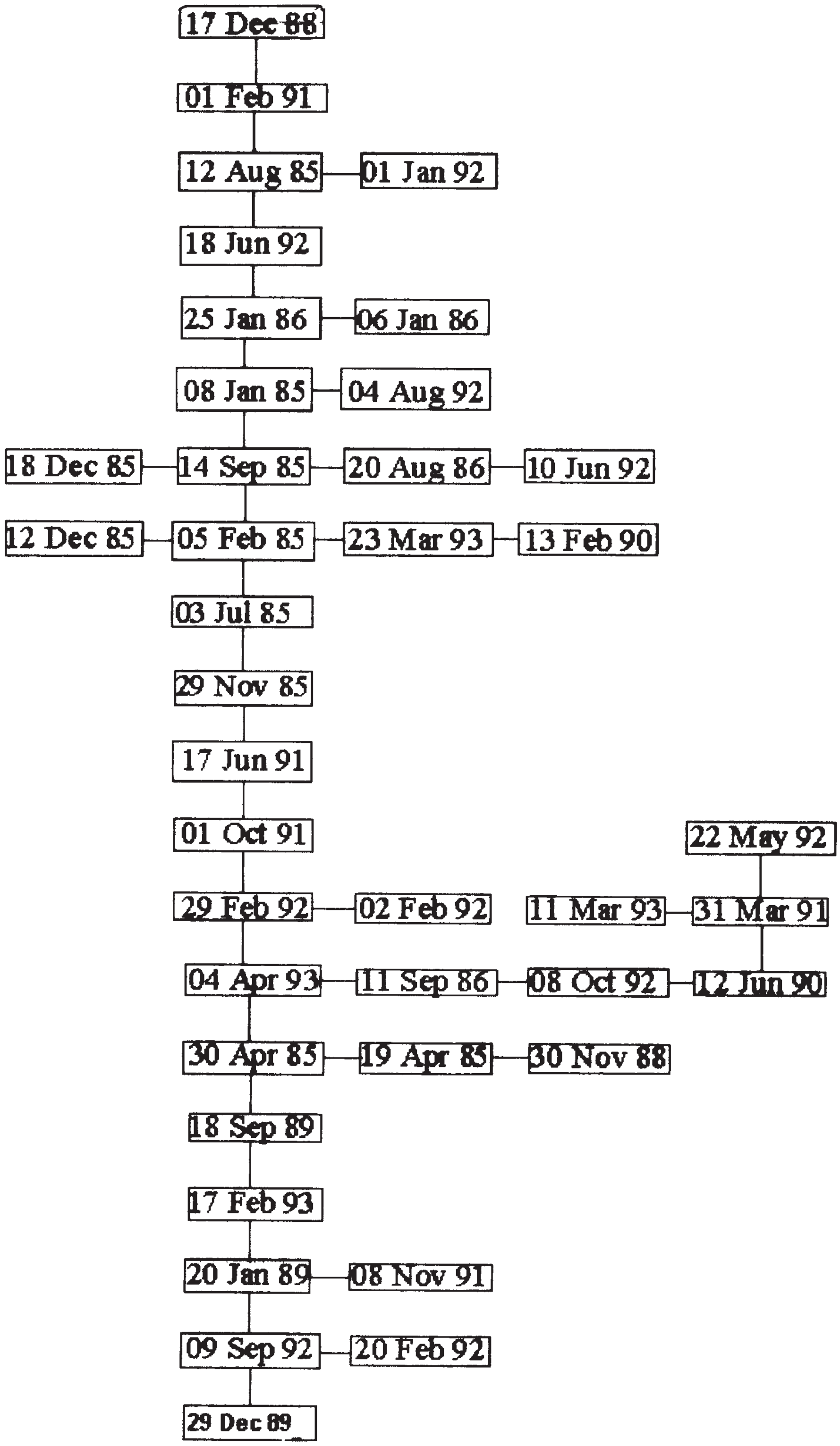

Figure 2. Network with minimum coverage connections in 7 dimensional space.

Table 3. *Spanning distances in 2-D and correlation coefficients.*

Serial no.	Reference stations as two-dimensional coordinates	Spanning distance	Correlation coefficient
1.	Alibag – Hyderabad	1178.27	0.9927
2.	Alibag – Ujjain	1323.66	0.9583
3.	Hyderabad – Ujjain	1329.25	0.9500
4.	Alibag – Sabhawala	1499.27	0.8967
5.	Alibag – Annamalainagar	1555.36	0.7777
6.	Hyderabad – Sabhawala	1560.63	0.8964
7.	Annamalainagar – Hyderabad	1575.22	0.7705
8.	Hyderabad – Kodaikanal	1599.62	0.8822
9.	Alibag – Kodaikanal	1610.06	0.8775
10.	Hyderabad – Trivandrum	1630.71	0.8816
11.	Annamalainagar – Ujjain	1637.23	0.7274
12.	Alibag – Trivandrum	1650.41	0.8813
13.	Sabhawala – Ujjain	1664.73	0.9090
14.	Kodaikanal – Ujjain	1746.04	0.8359
15.	Annamalainagar – Sabhawala	1776.61	0.6472
16.	Trivandrum – Ujjain	1782.00	0.8451
17.	Kodaikanal – Trivandrum	1782.62	0.9269
18.	Kodaikanal – Annamalainagar	1812.85	0.7826
19.	Kodaikanal – Sabhawala	1879.21	0.7454
20.	Annamalainagar – Trivandrum	1901.61	0.8131
21.	Trivandrum – Sabhawala	1937.45	0.7677

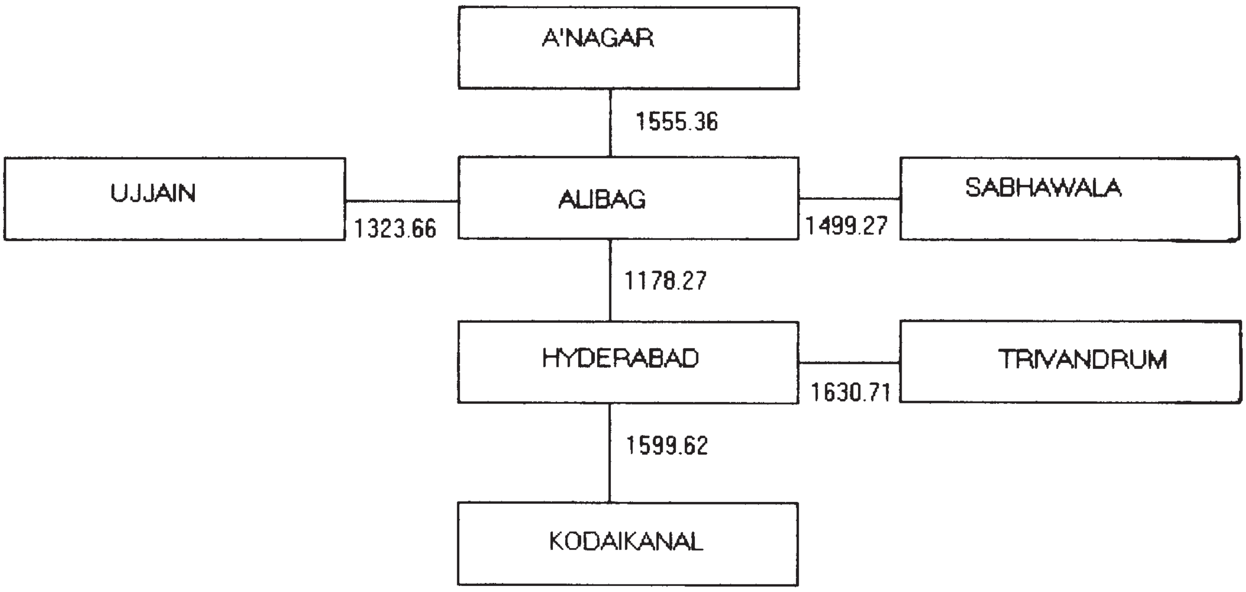

Figure 3. Tree with minimum covering distance in two-dimensional sense.

from Alibag and their corresponding correlation coefficients with respect to Alibag are provided in table 4.

It is observed from table 4 that, in general, the geographical latitudes match with their spanning distances. However, this is not the case with Annamalainagar, Kodaikanal and Trivandrum observatories. From this table it is seen that variations arising from application of the technique of correlation may not correspond with those in the case of latitudinal dependence as Trivandrum and Alibag show close correlation and larger distance when compared to Annamalainagar and Alibag.

As latitudinal difference between two stations increases, it is expected that their correlation coefficients would decrease. But a contrary situation is observed from table 4 in the case of Annamalainagar, Kodaikanal and Trivandrum by taking Alibag as reference station. As latitudinal difference increases, the general expectation is that the spanning distance would increase. The three stations in relation to Alibag as seen from table 4 fulfill this expectation. Thus, while the spanning distance follows the expected pattern, the correlation coefficient does not follow the expected fashion.

Table 4. *Comparison of latitudinal difference with correlation coefficient and spanning distance.*

Pairs of stations	Difference in geographic latitude	Correlation coefficient	Spanning distance
ABG – ANN	7.24	0.7777	1555.36
ABG – KOD	8.38	0.8775	2777.89
ABG – TRD	10.13	0.8813	2808.98

From this study, it is concluded that Hyderabad and Alibag observatories are close to each other while Trivandrum and Annamalainagar are at extreme ends in respect of the activity ranges of storms (H-variations).

The amplitude of geomagnetic field variations (oscillations) is a function of latitude and local time. Close to the dip equator just south of Trivandrum in the equatorial zone, the total field vector is entirely horizontal and the day time electrical conductivity in the ionosphere (approximately 110 kms height) is enormously increased leading to tremendous enhancement of the magnetic field fluctuations as compared to neighboring low latitudes. The maximum distance between Trivandrum and Annamalainagar observatories with respect to their storm time ranges is attributed to the influence of enhanced electrical currents, i.e., 'Equatorial Electrojet'.

Thus the present study not only confirms the existing theories on latitudinal dependence of geomagnetic field (H) variation but also provides evidence to the conclusion that the technique of multidimensional scaling is preferable to the conventional technique of correlation.

6. Conclusion

The approach described here differs significantly from the approach of classical univariate or multivariate statistics. In statistics individual measurements are lost sight of in favour of estimated means and variances of samples of population whereas in the present approach the identity of each point, each observation vector and its relationship to all other points are preserved and are central to the analysis. This provides a powerful tool for examining interrelationships among the individual observatories. As a result of this study, it is expected that this technique will yield in the future a means of forecasting the trend in storm time variation of a particular observatory.

References

Armitage P and Cotton T 1998 Encyclopaedia of Biostatistics, Vol. 5 (New York: John Wiley and Sons)

Alex S and Rao D R K 1995 Magnetospheric contribution to low latitude sq variations; *J. Geomag. Geoelectr.* **47** 599–606

Chen W K 1976 *Applied Graph Theory* (Amsterdam: North Holland Publishing Company)

Kruskal J B and Wish M 1978 Multidimensional Scaling (Newsbury Park: Sage)

Matsushita S and Camphell W H 1967 Physics of Geomagnetic Phenomena (New York and London: Academic Press)

Obayashi T and Jacobs J A 1957 Sudden commencement of magnetic storms and atmospheric dynamo action; *J. Geophys. Res.* **62** 589–616

Parkinson W D 1983 Introduction to Geomagnetism (Edinburgh and London: Scottish Academic Press)

Rastogi R G, Rao D R K, Alex S, Pathan B M and Sastry T S 1997 An intense SFE and SSC event in geomagnetic H, Y and Z fields at the Indian chain of observatories; *Annales Geophysicae* **15** 1301–1308

Rastogi R G and Patil A R 1992 On certain aspects of daily variation of geomagnetic field at low latitudes; *J. Geomag. Geoelectr.* **44** 495–503

Thompson H K Jr and Woodbury M A 1970 Clinical data representation in multidimensional space, computers and biomedical research, **3** 58–73

Indian Magnetic Data 1985–1993 (Mumbai: Indian Institute of Geomagnetism)

MS received 2 August 2001