J. Ind. Geophys. Union (January 2013)
Vol.17, No.1, pp. 49-58

Geomagnetic field variations in India – an analysis by Fuzzy c-means clustering

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

[1]Magnetic Observatory, Indian Institute of Geo-magnetism, Pondicherry University Campus, Puducherry- 605 014, India. e-mail:sri26357@gmail.com

[2]Ramanujan School of Mathematics, Pondicherry University, Puducherry- 605 014, India, e-mail: amsramasamy@gmail.com

ABSTRACT

The aim of this paper is to identify the pattern of geomagnetic field variations between the internationally declared quiet and disturbed days in the Indian region. The significance of this study is due to the fact that India has a unique set of geomagnetic observatories spanning the magnetic equator and the Sq focus in the region of the world where geomagnetic and geographic planes are least separated. The technique of fuzzy *c-means* clustering which is powerful for pattern recognition is employed in this study. The distance between the cluster centers can be taken as a measure of geomagnetic activity. The abnormal variations at Ujjain due to conductivity anomalies, at Trivandrum due to equatorial electro jet current system have been identified. The pattern of variation differs between the geomagnetically quiet and the disturbed periods. The difference in pattern shows anomalous variations at Ujjain. A major geological feature i.e. Narmada-Son lineament, inspite of being tectonically active is not affecting the variations in the vertical Z-component in magnetic records of Ujjain. However, it is found to be significant in the horizontal H component. Analytical technique and the results of the analysis are presented here.

INTRODUCTION

The anomalous geomagnetic behavior observed at some of the places can also be linked to the geological structure present there. The correlation of the magnetic fields at low latitudes and that in the vicinity of the dip equator has been studied earlier (Rangarajan and Dhar, 1992; Jadhav et.al, 2002). Geomagnetic variations of different origins exhibit remarkable enhancement surrounding the dip equator. However, the extent and magnitude of enhancement for events of varying origin and factors responsible for its variability are still not clear (Arora and Bhardwaj, 2003). In addition to the ionospheric dynamo currents, parameters relating to magnetosphere, interplanetary magnetic field and solar wind also partially modify the regular variations of the magnetic field. The procedure of dividing the geomagnetic variations according to the seasons gives a useful indication of the average daily variation of the field for each of the seasons. But it does not always represent the typical variation in a particular epoch. Considerable changes in phase or form as well as amplitude of the variation may occur within the period defining a season, which may be obscured by taking an average. So, it is desirable to examine the daily ranges for quiet days and disturbed days separately in order to obtain the general pattern of variations in a particular region. Though many mathematical techniques are applied for the analysis of geomagnetic data, the fuzzy mathematical model (Sridharan, 2009 and Sridharan et.al, 2005) is found to be the most suitable and powerful for pattern recognition.

The purpose of clustering is to distill natural groupings of data from a large data set and producing concise representation of the behavior of a system. Fuzzy *c-means* clustering enables one to lump together the data points that populate some multidimensional space into a specific number of different clusters. For a given finite set X of data, the problem of clustering in X is to find several cluster centers (Klir and Yuvan, 1997) that can properly characterize relevant classes of X. In classical cluster analysis, these classes are required to form a partition of X, such that the degree of association is strong for data within blocks of the partition and weak for data across the classes. However, this requirement is too severe in many applications when the data is chaotic in nature and so it is desirable to replace it with a weaker requirement of a fuzzy pseudo partition or fuzzy c-partition, where c designates the number of fuzzy classes in the partition. The technique of fuzzy *c-means* clustering provides a quantitative estimation for geomagnetic quiet and disturbed day ranges and anomalies at the Indian region.

DATA ANALYSIS

The hourly variations of the horizontal H component of the six geomagnetic observatories situated at Alibag (ABG), Pondicherry (PON), Nagpur (NGP), Ujjain (UJJ), Visakapatnam (VSK) and Trivandrum (TRD) for each of the five international quiet and disturbed days of every month from 1996 to 1999 have been considered for this analysis. The daily range used in the present study is the daily maximum of hourly means – daily minimum of hourly means of the horizontal field in a Greenwich day. The geographic and dipole coordinates of the six observatories and their location are provided in table 1 and Fig. 1 respectively. There were 240 days of internationally declared quiet and disturbed geomagnetic conditions during this period. Trivandrum Observatory, which was nearer to the dip equator, was closed in September 1999. Hence, the available simultaneous data for 225 days were taken up for the present study.

Analytical technique:

Let $X = \{x_1, x_2, .. \quad , x_n\}$ be a set of given data. Let c denotes a positive integer. A fuzzy pseudo partition or fuzzy c-partition of X is a family of fuzzy subsets of X, denoted by $P = \{A_1, A_2, \ldots, A_c\}$, such that

$$\sum_{i=1}^{c} A_i(x_k) = 1 \qquad \text{..........(1)}$$

for all $k \in N_n$ and

$$0 < \sum_{k=1}^{n} A_I(x_k) < n \qquad \text{..........(2)}$$

for all $i \in N_c$.

For all $k \in N_n$, the problem of fuzzy clustering is to find a fuzzy pseudo partition and the associated cluster centers by which the structure of the data is represented as best as possible. This requires some criterion expressing the general data that the associations in the sense described by the criterion are strong within the clusters and weak between the clusters. To solve the problem of fuzzy clustering, one has to formulate this criterion in terms of performance index. Usually, the performance index is based upon cluster centers. Given a pseudo partition $P = \{A_1, A_2, \ldots A_c\}$, the c cluster centers, $v_1, v_2, \ldots, v_c$, associated with the partition are calculated by the formula

$$v_i = \frac{\sum_{k=1}^{n} \left[A_i(x_k)\right]^m x_k}{\sum_{k=1}^{n} \left[A_i(x_k)\right]^m} \qquad \text{.......... (3)}$$

for all $i \in N$, where $m > 1$ is a real number that governs the influence of membership grades. It

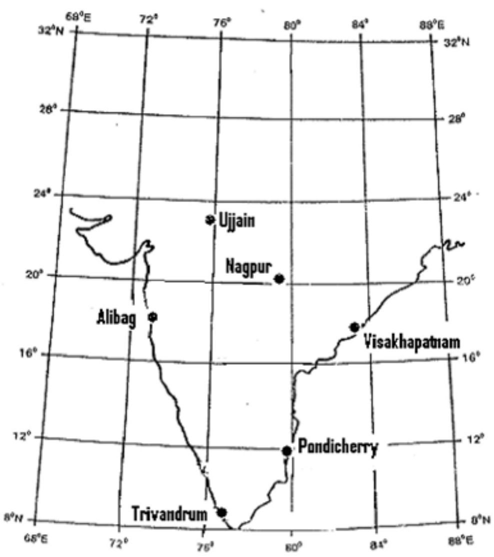

Figure 1. Location map of geomagnetic observatories

Table 1. Location of Observatories

Serial No.	Station	Geographic coordinate	Dipole coordinate
1	Alibag (ABG)	Latitude 18° 37'N Longitude 72° 52'E	10.0°N 145.9°
2	Pondicherry (PON)	Latitude 11° 55'N Longitude 79° 55'E	2.7°N 152.1°
3	Nagpur (NGP)	Latitude 21° 09'N Longitude 79° 05'E	11.96N 152.1°
4	Trivandrum (TRD)	Latitude 08° 29'N Longitude 76° 58'E	0.8°S 148.5°
5	Ujjain (UJJ)	Latitude 23° 11'N Longitude 75° 47'E	14.3°N 149.2°
6	Visakhapatnam (VSK)	Latitude 17° 41'N Longitude 83° 19E	08.2°N 155.9°

is to be noted that the vector v_i calculated by (3), which is viewed as the cluster center of fuzzy class A_i, is actually the weighted average of data in A_i. The weight of datum $x_k$is the m^{th} power of the membership grade of $x_k$in the fuzzy set A_i.

The performance index of a fuzzy pseudo partition P, J_m (P), is then defined in terms of the cluster centers by the formula

$$J_m(P) = \sum_{k=1}^{n}\sum_{k=1}^{c}[A_i(x_k)]^m \,||x_k-v_i||^2 \qquad ....(4)$$

where ||.|| is some inner product- induced norm in space R^p and $|| x_k - v_i ||^2$ represents the distance between x_k and v_i. This performance index measures the weighted sum of distances between cluster centers and elements in the corresponding fuzzy clusters. Clearly, the smaller the value of J_m (P), the better the fuzzy pseudo partition. Therefore, the goal of the fuzzy *c-means* clustering method is to find a fuzzy pseudo partition P that minimizes the performance index J_m (P). Thus the clustering problem turns out to be an optimization problem.

The process consists of the following three steps:

1. Partition the items into K initial clusters.
2. Proceed through the list of items, assigning an item to the cluster whose centroid is the nearest. Compute the Euclidean distance. Recalculate the centroid for the cluster receiving the new item and for the cluster losing the item
3. Repeat Step 2 until no more reassignment is possible.

The following fuzzy *c-means* algorithm developed by Bezdek (1986) was implemented for solving the optimization problem.

Fuzzy c-Means Algorithm

The algorithm is based on the assumption that the desired number of clusters c is given and, in addition, a particular distance, a real number $m \in (1, \infty)$, and a small positive number ε, serving as a stopping criterion, are chosen.

Step1. Let t = 0. Select an initial fuzzy pseudo partition $P^{(0)}$.

Step2. Calculate the c cluster centers $v_1^{(t)}$ …, $v_c^{(t)}$ for $P^{(t)}$ using the relation (3) and the chosen value of m.

Step3. Update $P^{(t+1)}$ by the following procedure:

For each $x_k \in X$, if $||x_k-v_i^{(t)}||^2 > 0$ for all $i \in N_c$, then define

$$A_i^{(t+1)}(x_k) = \frac{[(\| x_k - v_i(t) \|^2)^{1/m-1}]^{-1}}{\left\| x_k - v_j(t) \right\|^2}$$

if $\| x_k - v_i(t) \|^2 = 0$ for some $i \in N_c$, then define $A_i^{(t+1)}(x_k)$ for $i \in 1$ by any nonnegative real numbers satisfying

$$\sum_{i\in 1} A_i^{(t+1)}(x_k) = 1$$ and define $A_i^{(t+1)}(x_k) = 0$ for $i \in N_{c-1}$.

Step4. Compare $P^{(t)}$ and $P^{(t+1)}$. If $| P^{(t+1)} - P^{(t)} | \le \varepsilon$, then stop; otherwise, increase t by one and return to step 2.

In Step 4, $|P^{(t+1)} - P^{(t)}|$ denotes a distance between $P^{(t+1)}$ and $P^{(t)}$ in the space $R^{nxc.}$ An example of this distance is

$$| P^{(t+1)} - P^{(t)} | = \max_{i \in Nc, k \in Nn} | A_i^{(t+1)}(x_k) - A_i^{(t)}(x_k) | \quad ..(5)$$

In the algorithm, the parameter 'm' is selected according to the problem under consideration. When $m \to 1$, the fuzzy c-means converges to a "generalized" classical c means. When $m \to \infty$, all cluster centers tend towards the centroid of the data set X. That is, the partition becomes fuzzier with increasing m. Currently, there is no theoretical basis for an optimal choice for the value of m. However, it was established that the algorithm converges for any $m \in (1, \infty)$. The value m = 1.25 suggested by Klir & Yuan (1997) for finding 2 cluster centers are applied in this analysis.

For the present analysis a fuzzy pseudo partition is determined with two clusters c 2, m = 1.25 and ε = 0.5.Searches were carried out to find clusters in input-output training data. Initially the cluster centers are very inaccurately placed. Additionally, every data point has a membership grade for each cluster. By iteratively updating the cluster centers and membership grades for each data point, the cluster centers move to the "right" location. This iteration is based on minimizing an objective function that represents the distance from any given data point to a cluster weighted by the membership grade of that data point. The final output of fuzzy c-means is a list of cluster centers. The spanning distance between the cluster centers are determined (table 2 and 3) for quiet and disturbed day ranges. The number of iterations and the co-ordinates of the cluster centers are provided in table 4. Fig.s 2 to 7 represent the fuzzy clustering of data and the cluster centers (C1,

Table 2. Spanning distances for quiet day ranges

	PON	NGP	TRD	UJJ	VSK
ABG	15.2	14.2	52.7	10.2	4.5
PON		6.0	37.6	25.3	10.8
NGP			39.6	24.1	10.4
TRD				62.9	48.3
UJJ					14.6

Table 3. Spanning distances for disturbed day ranges

	PON	NGP	TRD	UJJ	VSK
ABG	19.6	8.3	75.0	15.7	2.8
PON		23.1	56.8	35.2	17.9
NGP			74.8	17.1	7.1
TRD				90.6	72.6
UJJ					18.0

Table - 4. Co-ordinates of cluster centers and distances during iterations

Station	Number of iterations	Coordinates of cluster centers for Q days	Coordinates of cluster centers for D days	Latitudinal separation from Ujjain	Spanning distances from Ujjain Q days D days
UJJ	32	32.62 34.92	49.42 134.6	…	…
NGP	31	53.66 59.47	60.99 138.9	2° 02′	24.1 17.1
ABG	35	40.52 46.33	55.81 145.4	4° 34′	10.2 15.7
VSK	37	43.52 49.14	59.13 145.5	5° 30′	14.6 18..0
PON	30	51.28 58.19	66.45 161.0	11° 16′	25.3 35.2
TRD	30	80.76 108.5	89.75 187.4	14° 42′	62.9 90.6

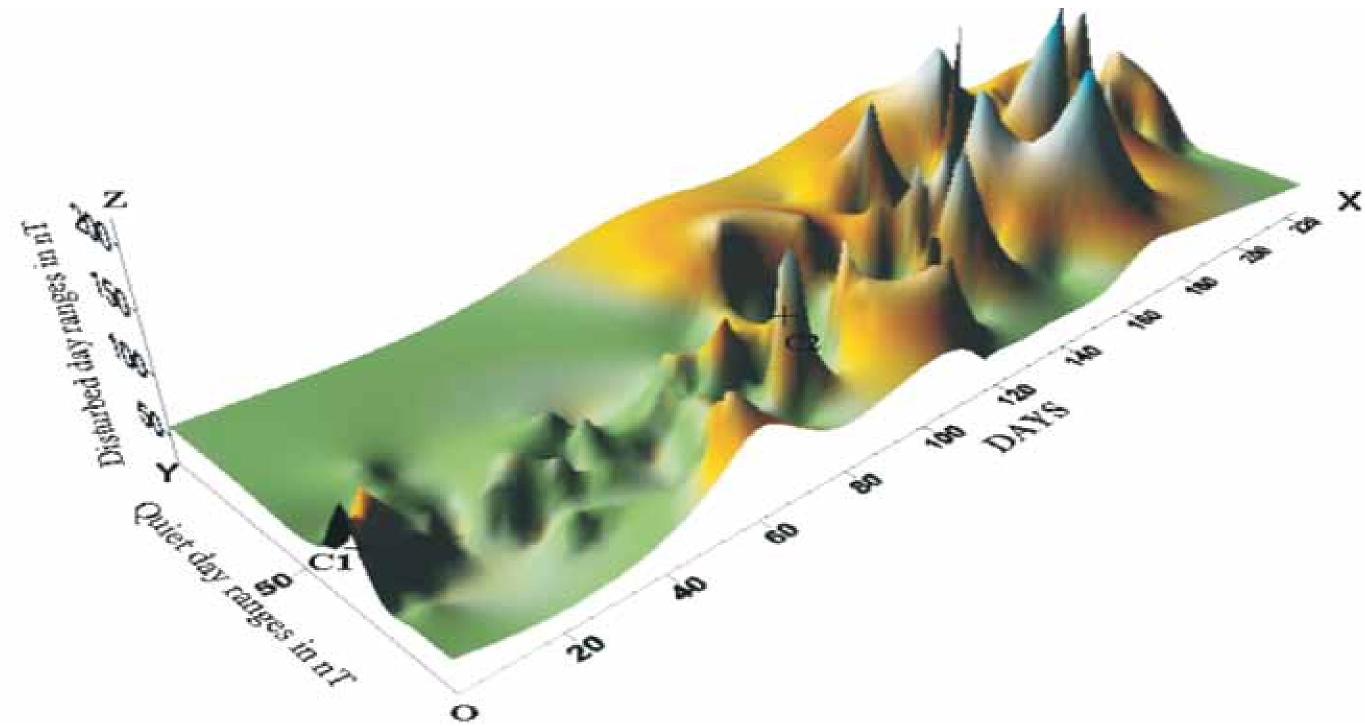

Figure 2. Quiet and Disturbed day ranges of Alibag

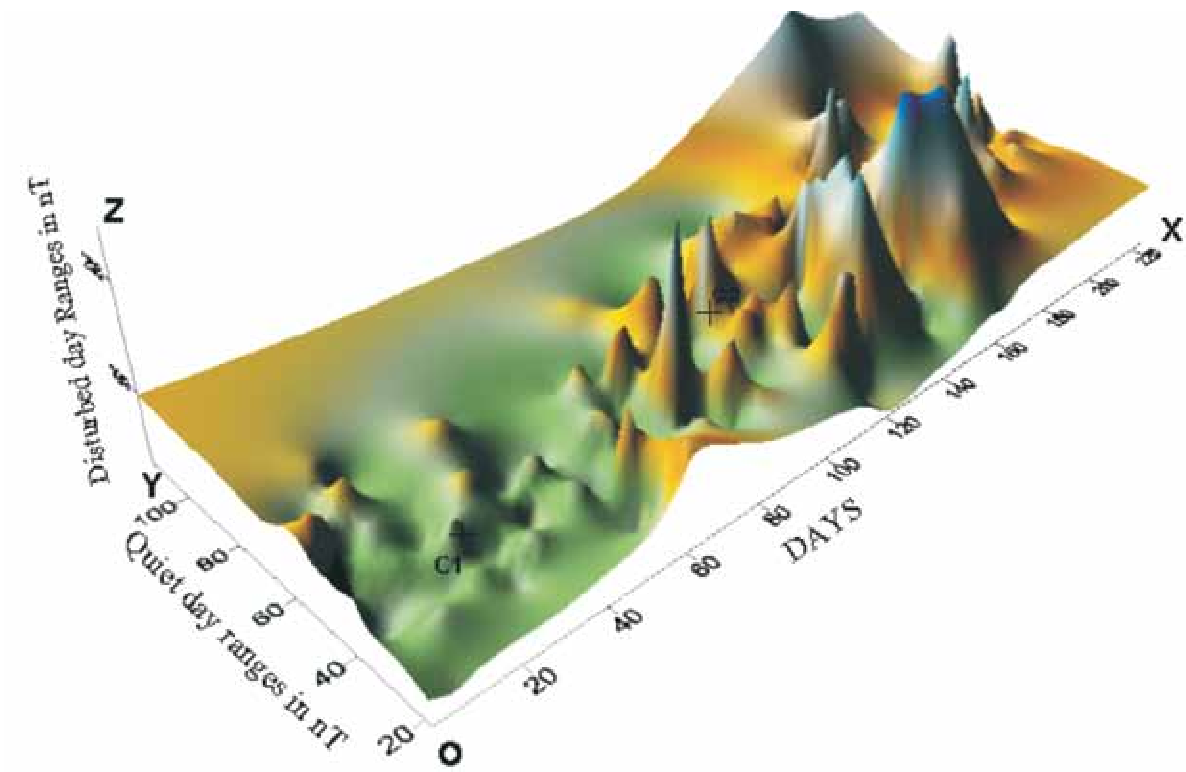

Figure 3. Quiet and Disturbed day ranges of Pondicherry

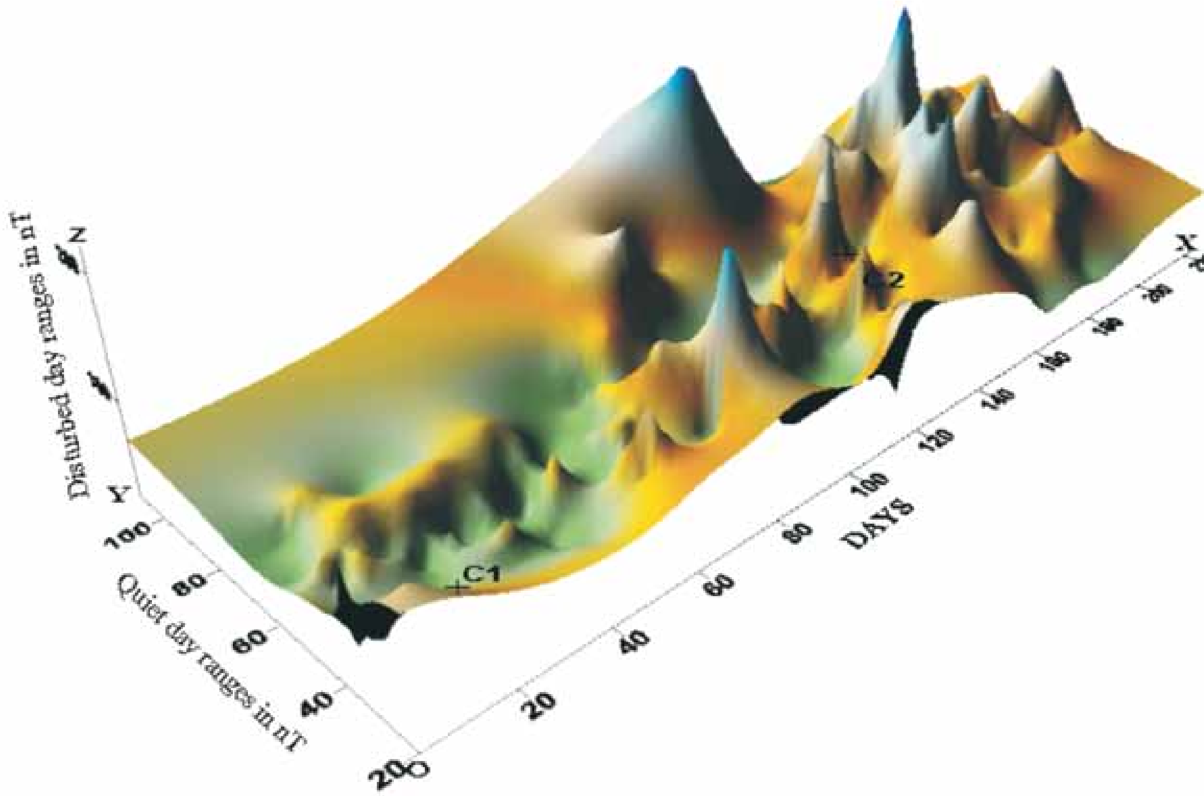

Figure 4. Quiet and Disturbed day ranges of Nagpur

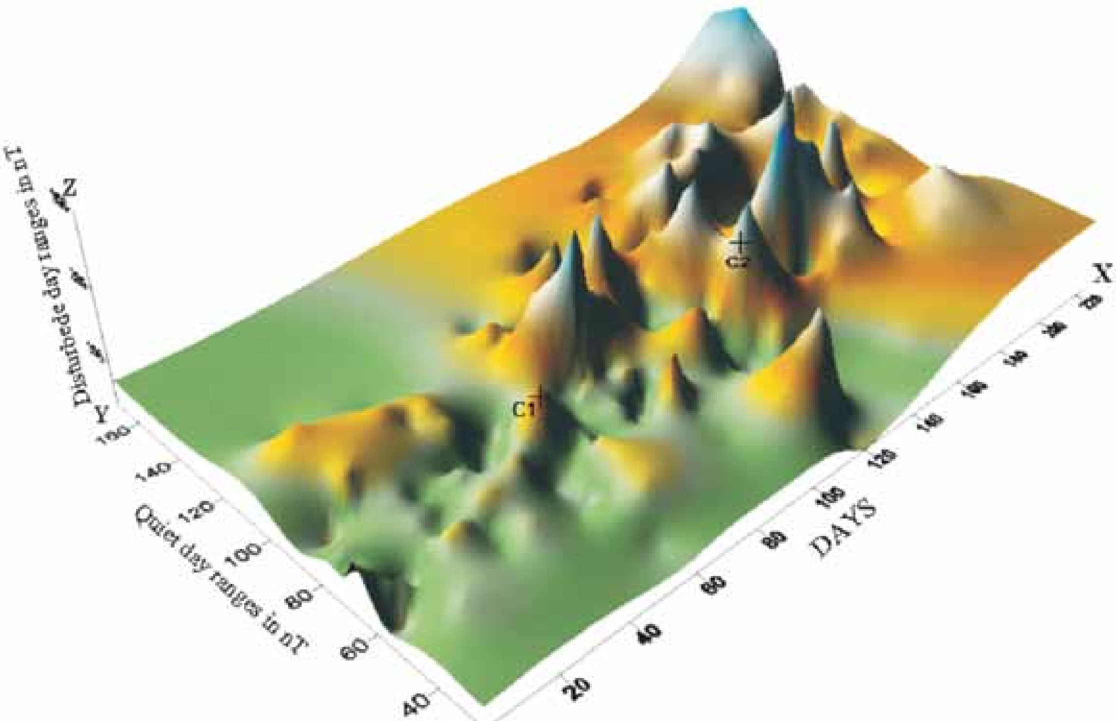

Figure 5. Quiet and Disturbed day ranges of Trivandrum

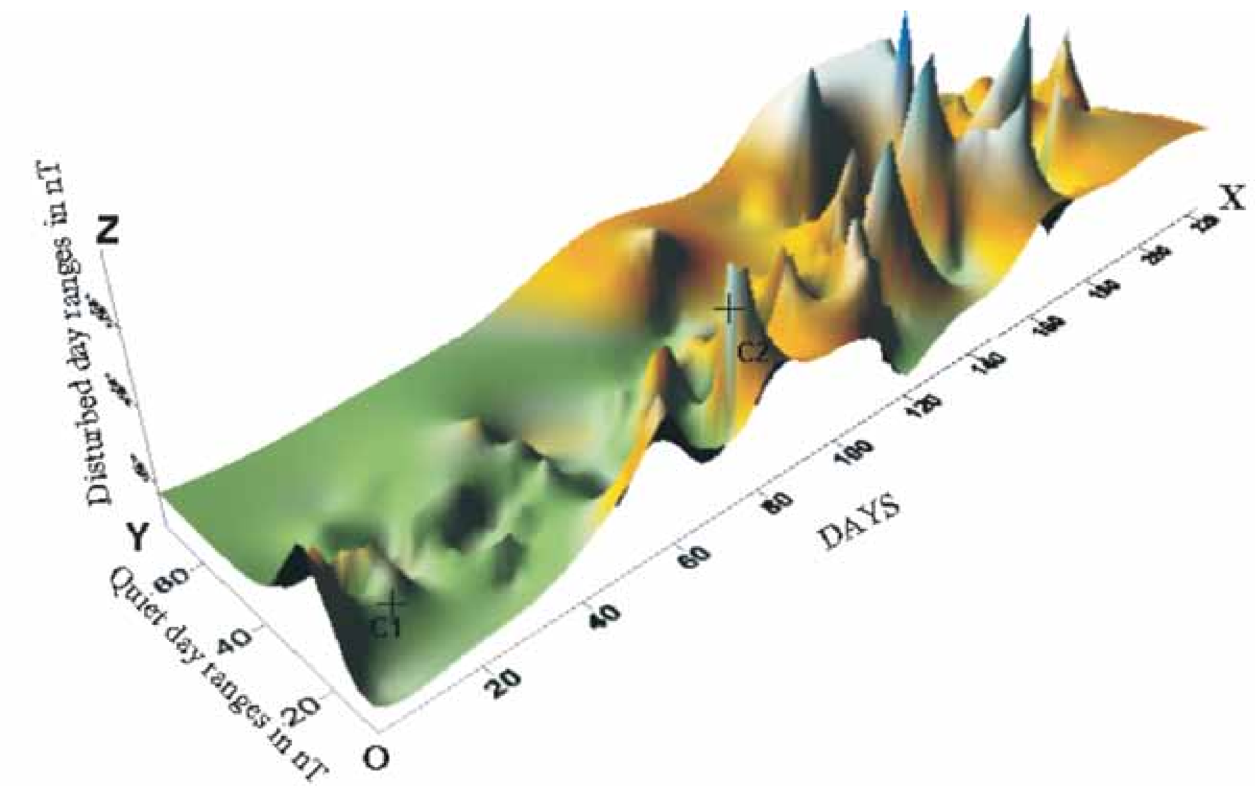

Figure 6. Quiet and Disturbed day ranges of Ujjain

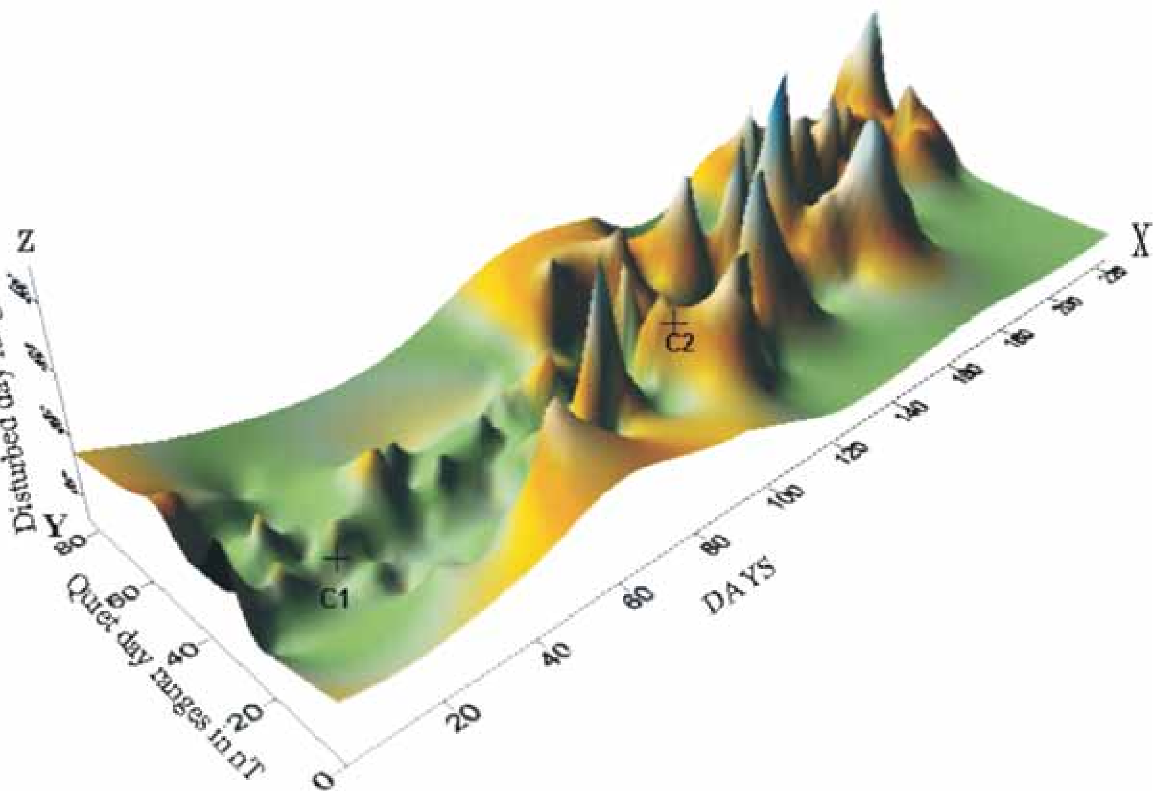

Figure 7. Quiet and Disturbed day ranges of Visakhapatnam

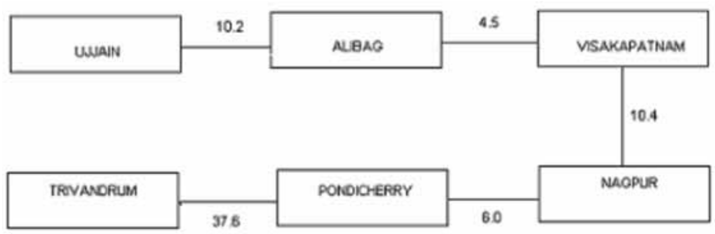

Figure 8. Pattern of quiet day variation and spanning distances

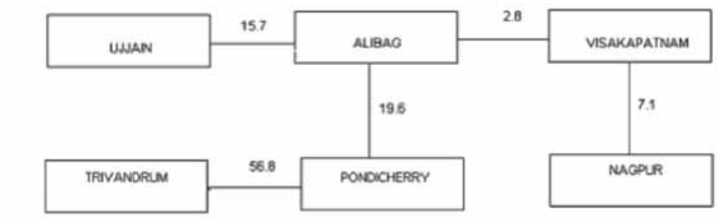

Figure 9. Pattern of disturbed day variation and spanning distances

C2) for the quiet and disturbed days respectively. The concept of multi dimensional scaling technique (Sridharan and Ramasamy, 2002 and Sridharan and Selvaraj, 2009) is applied and the spanning sub graph is obtained with minimum covering distance and it is provided in Fig.s 8 and 9. The spanning distances given in table 2and 3 are the distances between the two clusters in the planes of various pairs of stations. Spanning distances are unitless measurements. It is a distant measure to establish proximity and dominant relationship between different sets of data. For geomagnetic variations, three seasons namely Summer - (Nov, Dec, Jan, Feb); Winter – (May, Jun, Jul, Aug) and Equinox – (Mar, Apr, Sep, Oct) are followed. Simultaneous data of six observatories are considered in this analysis. The seasonal variation in one place will have the same impact in other places also. As such the present pattern of variations will not change.

RESULTS OF THE ANALYSIS

The following are the observations from the table 4 and Fig.s 8 & 9.

1. For both quiet and disturbed day ranges, ABG-VSK maintains the shortest spanning distance.
2. For both quiet and disturbed day ranges, UJJ-TRD maintains the maximum spanning distance.
3. For geomagnetically quiet and disturbed days, when the latitudinal separation is increased from UJJ to TRD (table 4), the corresponding distances between the cluster centers are also increasing except for UJJ-NGP. Though UJJ and NGP are geographically very close to each other, they maintain a larger spanning distance whereas, comparatively far away places ABG and VSK maintain a smaller distance with UJJ.
4. The variations at Trivandrum maintain a maximum spanning distance with all other places.

SUMMARY OF THE ANALYSIS

As the latitudinal separation between the two places, ABG and VSK is the least, one can see the minimum distance between the cluster centers during the quiet and disturbed days. The similarity and nearness in pattern of variations at ABG and VSK during geomagnetically quiet days for the horizontal H component (Sridharan et.al, 2005) revealed in the geomagnetic coastal studies is maintained for the disturbed days also. When the latitudinal separation is increasing from UJJ to TRD, the corresponding distances between the cluster centers are also increasing for both the disturbed and quiet days. As suggested by Alex and Jadhav (2007), anomalous changes observed in the day-to-day variation in the latitudinal pattern of horizontal *H* component are the influence of high-latitude magnetospheric current systems on certain quiet days. When the variation of *H* is examined, southward turning of the *H* field during the pre-noon hours seems to have started from the latitude of Ujjain, suggesting that the focus must have its location to the south of this latitude. The variation ranges from NGP to UJJ does not fit into the trend of gradual increase in H range towards north. The variations in Sq and other long period variations (Chandra and Arora, 1992) will help to provide better insight into the electrical conductivity-depth distribution at upper mantle depth. The electrical conductivity of the upper mantle can be determined by comparing the measured response of the Earth to magnetic field variations of all frequencies with the theoretical response of particular conductivity distributions (Banks, 1969). Electrical conductivity is a promising geophysical parameter that has the potential for predicting earthquakes. Electrical conductivity is sensitive to very small changes in the structure of rocks and is a complementary parameter to other geophysical techniques, such as potential field methods and most seismic methods, which are sensitive to the bulk properties of the medium.

The effect of induction currents at Ujjain for the studies on conductivity anomalies signifies a major geological feature i.e. Narmada-Son lineament, which is known to be tectonically active. However, its significance is not reflected in magnetic records of Ujjain for the variations in the vertical Z component (Agarwal, 1981). But, significance is seen for the horizontal H component in the present study during the geomagnetically quiet and disturbed days. The abnormalities of H at TRD and ABG measured in terms of solar daily variation of ΔH (TRD–ABG) averaged over the International Quiet and Disturbed days during different solar epochs and very active solar epoch has been studied. From table 2 one can observe that solar daily variations are consistently smaller on quiet days than on disturbed days. The abnormal variations of H at the equatorial region Trivandrum is due to the effect of Petersen conductivity that supports the east-west dynamo currents augmented due to the inhibition of the vertical Hall current (Baker and Martyn, 1953). The quantity of disturbance level is considered to be an index of the net disturbance field at the geomagnetic equator brought out by the symmetric ring current, the tail current in the night side and the current flowing at the magnetopause in the dayside. As reported by Yacob (1975), analysis of data from observatories in the vicinity of dip equator has shown that the ranges there and at station outside the influence of the equatorial electro jet are not well established. The equatorial enhancement of transient variation in the geomagnetic field is well known and has been attributed to increased conductivity of the S_q current layer in the equatorial region. There is a detectable and well-defined enhanced solar wind pressure measured by the magnitude of the disturbed level in the equatorial, low latitude and mid-latitude horizontal field. The asymmetric component of the elcetrojet field is complex in nature in contrast to that observed at low latitudes outside the jet influence.

As reported by Agarwal (1981), observatories located in the regions of high-conductivity anomalies (Rangarajan and Dhar,1992) show anomalous transient geomagnetic variations. But, one can attempt only a qualitative interpretation and any quantitative interpretation should be supported by other controls. However, it could not be resolved unequivocally the reasons for abnormal variations. There is a probability that to certain regional differences in the conductivity distribution could be responsible. Agarwal (1981) reported that mean direction of horizontal variations do not fit in latitudinal trend of gradual shift towards east, indicating an anomalous behavior at Ujjain. It cannot be ascribed to latitudinal variations. The anomalous horizontal variations during the internationally declared quiet days in the present study confirm the conductivity contrast on the northern side of Ujjain. This needs further investigation. As the geomagnetic observatory at Ujjain is closed, more data is not available and, hence, we are unable to perform further analysis on this aspect.

CONCLUSION

Patterns of geomagnetic variations during the internationally declared quiet and disturbed days at certain selected places in the Indian region are explained in this paper. The abnormal variations at the equatorial station Trivandrum due to day time electrojet current system, and the regional inconsistency in variations at the Ujjain due to conductivity anomalies, have been confirmed with the updated data by applying fuzzy model. An important result of this analysis is that the anomalous variations not found in the magnetic records of Ujjain for the vertical Z component are found for the horizontal H component during the geomagnetically quiet and the disturbed periods. In fuzzy c-means clustering, the compactness and separation validity is more accurate and can produce a better segmentation result. The distance of an input sample to the center of the cluster to which the input sample belongs, is used as a criterion to measure the cluster compactness. Starting with given data points, the fuzzy clustering method provides a collection of cluster centers and unknown parameters. Based on this the inter-relationship between one observatory and the others, in terms of spanning distances, can be estimated by a fuzzy model in a naturally consistent way. It is hoped that the methodology adopted in the present study will open up a new avenue to identify the local anomalies and pattern(s) of variations for furthering research in geomagnetism.

ACKNOWLEDGEMENTS:

The authors are very grateful to the referee and the Editorial Advisor for comments and corrections made in the manuscript. They greatly improved the method and quality of presentation. The first author is very grateful to the Director and Prof.B.M.Pathan of Indian Institute of geomagnetism for their kind support in carrying out this research.

REFERENCES

Alex, S and Jadhav. 2007. Day-to-day variability in the occurrence characteristics of Sq focusduring d-months and its association with diurnal changesin the Declination component, Earth Planer Sci., Vol 59, 1197-1203.

Agarwal, A., 1981. Possible explanation of conductivity anomalies near Ujjain and Jaipur, Proc. Indian Acad. Sci. (Earth Planet.Sci.), , Vol.90 (3), 237-244.

Arora, B.R and Bhardwaj, S.K., 2003. Spatial and frequency characteristics of equatorial enhancement of geomagnetic field variations, Journal of Atmospheric and Solar-Terrestrial Physics 65, 1283-1292.

Baker, W.G., Martyn, D.F., 1953. Electric currents in the ionosphere, Philosophical Transactions of Royal society of London Series, a 246, 281-294.

Banks, R. J. 1969, Geomagnetic Variations and the Electrical Conductivity of the Upper Mantle. Geophysical Journal of the Royal Astronomical Society, 17: 457–487. doi: 10.1111/j.1365-246X.1969.tb00252.x

Bezdek, J.C, Biswas, G. and Huang, L.Y., 1986. Transitive closures of fuzzy thesauri for information and retrieval systems, Int. J. Man-Machine Studies., 25, 343-356.

Bhargava, B.N. and Yacob, A., 1974. Local time and solar cycle features of day-to-day variability in geomagnetic horizontal intensity, Ann. Geophys., 30, 487-492.

Chandra sekar, E., and Arora, B.R., 1992, Upper mantle electrical conductivity distribution beneath the Indian sub continent using geomagnetic storm time variations. Memoir Geological Society of India, (24), 149-157.

Geeta Jadhav, Mita Rajaram and Rajaram R., 2002, A detailed study of equatorial electrojet phenomenon using orsted satellite observations, J.Geophysical Research.107, 1029-1040.

Klir, G.J., and Yuvan, B, 1997, Fuzzy Sets and Fuzzy Logic., Theory and Applications, Prentice Hall of India Private Ltd, New Delhi, India.

Rangarajan, G.K and Ajay Dhar., 1992, Response of the Sq and Equatorial electrojet and variations to the north-south asymmetry in geomagnetic activity, J. Geomag. Geoelectr. 44, 899-908.

Sridharan, M and Ramasamy, A.M.S., 2002, Multidimensional scaling technique for analysis of magnetic storms at Indian Observatories, J. Earth System Science., 111, 459-465.

Sridharan, M, Gururajan, N and Ramasamy, A.M.S., 2005, Fuzzy clustering analysis to study geomagnetic coastal effects, Annals Geophysicae, 23, 1157-1163.

Sridharan, M and Samuel Selvaraj, R., 2009, Multi dimensional scaling of geomagnetic Sq (H) variations, Indian Journal of Radio & Space Physics., 938,165-173.

Sridharan, M, 2009, Fuzzy mathematical model for the analysis of geomagnetic field data, Earth, Planets Space, 61 (10), p.1169-1177.

Yacob, A., 1975, Latitudinal profile in India of Sq (H) range and of its prominent periodicities, Pageoph, 113, 601-609.