Main Patterns of the geomagnetic field: A Preliminary case study using principal component analysis.

Virgínia Klausner, Observatório Nacional/ON, Brazil
Odim Mendes Jr, CEA/DGE/INPE, Brazil
Andrés R. R. Papa, Observatório Nacional/ON, Brazil
Margarete Oliveira Domingues, CTE/LAC/INPE, Brazil

This paper was prepared for presentation during the 11th International Congress of the Brazilian Geophysical Society held in Salvador, Brazil, August 24-28, 2009.

Abstract

Principal Component analysis is used to study a time-series of the total geomagnetic field for different magnetic stations to determine the dominant patterns of the geomagnetic field variance. A five-day interval from 13 to 17 of July, 2000 was used. The purpose of this work is to describe the dominant patterns of variance in the time series of the Vassouras and the most used six stations that compose the asymmetric and symmetric indice.

Introduction

By international agreement, the Earth's field components are described by the "right-hand system". It means that the x direction would be indicated by our thumb, the y direction by our pointing finger and the z direction by the remaining finger. However, the Earth's field can be described in two ways: (1) three orthogonal component field called the X, Y and Z representation or (2) the horizontal magnitude, the eastward angular direction of the horizontal component from geographic northward and the downward component called, respectively, the H (horizontal), D (declination) and Z (vertical) representation. Figure 1 illustrates these nomenclatures for a location in the Northern Hemisphere where the total field vector points into the Earth (Campbell, 1997).

The geomagnetic field is a complicated function of space and time. Ground based magnetic measurements show a repetitive diurnal variation on geomagnetically quiet days (Tascione, 1988). But there is a great variety of irregular variations that occur from time to time, the "disturbance fields". Periods of great disturbance are called, by analogy with the weather, "magnetic storms" (Parkinson, 1983).

The primary causes of geomagnetic storms at Earth are strong dawn-to-dusk electric field associated with the passage of southward directed interplanetary fields, Bs, passing the Earth for sufficiently long intervals of time (more than 3 hours). The solar wind energy transfer mechanism is magnetic reconnection between the interplanetary magnetic field and the Earth's magnetic field (Gonzalez et al., 1994).

The magnetic field measured at mid-to-low latitudes can be affected significantly by variations of the solar wind ram pressure, which produces changes in the magnetopause current. This process gives place to a storm sudden commencement (SSC), when an increase in the horizontal magnetic field is observed at mid-to-low latitudes (Mendes et al., 2005). The characteristic signature of a magnetic storm is a depression in the horizontal component of the Earth's magnetic field due to the changes of the ring current (Gonzalez et al., 1994).

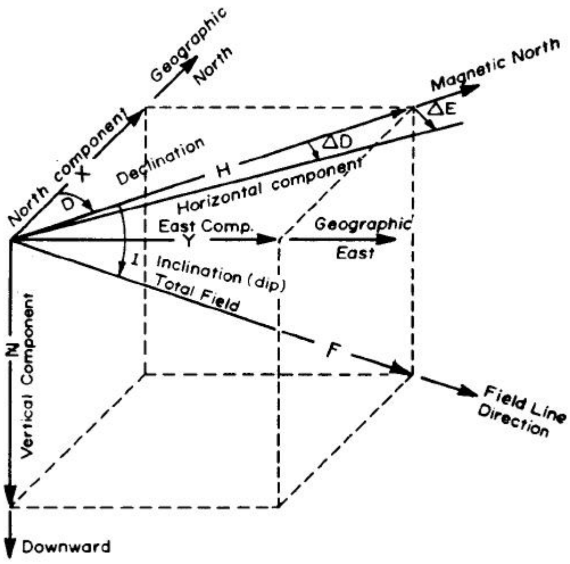

Figure 1: Components of the geomagnetic field measurements. Source: Campbell, 1997.

The asymmetric (ASY) and the symmetric disturbance indice describe the geomagnetic disturbance field in mid-latitudes with high-time resolution. These indice are derived for both H and D components, that is, for the components in the horizontal (dipole pole) direction H (SYM-H, ASY-H) and in the orthogonal (East-West) direction D (SYM-D, ASY-D). These indice are calculated by averaging the disturbance component at each minute (Fredericksburg, Boulder, Tucson, Memambetsu, Martin de Vivies and Chambon-la-Forêt). Some of these stations can be eventually replaced by others depending on the availability and the condition of the data of the month (WDC-Kyoto, 2008).

Principal Component (PC) Analysis

Among the several available methods of analysis, PCs is a particularly useful tool in studying large quantities of multi-variate data. PCs analysis is used to decompose a time-series into its orthogonal component modes, the first of which can be used to describe the dominant patterns of variance in the time series (see e.g. Keiner and Yan, 1997).

PCs are derived as the eigenvectors of the correlation matrix between the variables. Their form depend directly on the interrelationships existing within the data itself. The first PC is that linear combination of the original variables, which when used as a linear predictor of these variables, explain the largest fraction of the total variance. The second, third PC, etc., explain the largest parts of the remaining variance (Murray et al., 1984).

Consider M variables $x_m(t)$, which might represent the geomagnetic observations at M stations as functions of time. Let these be observed at N times, $i = 1, 2, \ldots, n$. We can construct the $m \times n$ matrix as follow:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \quad (1)$$

The center of gravity of the m points is $\bar{x}$ where the ith coordinate is

$$\bar{x_i} = \frac{1}{m}\sum_{j=1}^{m} x_{ij} \quad (2)$$

The points measured from their center of gravity, $v_{ij} = (x_{ij} - \bar{x}_i)$ can be written

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nm} \end{bmatrix} \quad (3)$$

Dividing with element of the v_{ij} by the standard deviation s_i, we rewrite each element of V as:

$$v_{ij} = v_{ij}/s_i \quad (4)$$

After, we compute de correlation matrix of the V matrix. The correlation matrix is a symmetric matrix, since the correlation of column i with column j is the same as the correlation of column j with column i.

$$C = \frac{1}{N}[VV^T] \quad (5)$$

We obtained the PCs as the eigenvectors of the correlation matrix C by resolving:

$$C\vec{e} = \vec{e}\lambda \quad (6)$$

In this case, λ is an eigenvalue and $\vec{e}$ is an eigenvector.

We can summarize several the results of the expressions above (Shlens, 2005):

- C is a square symmetric $m \times m$ matrix.
- The eigenvalues λ are the variance of particular measurements types.
- The eigenvectors $\vec{e}$ are the principal components.

A property of PCs which make them particularly appealing is that, unlike conventional orthogonal representation as: the Fourier decomposition, Tschebycheff, spherical harmonics, they do not require any predetermined form (Murray et al., 1984).

Data set

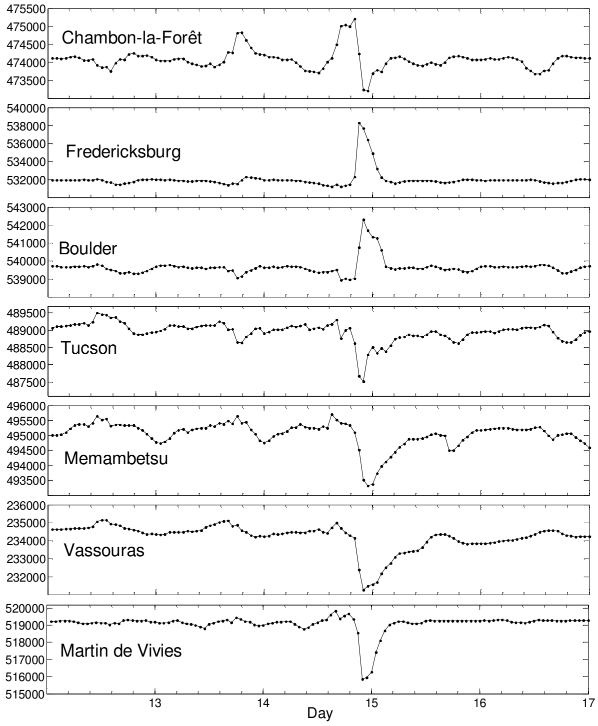

Figure 2: The F component magnetograms obtained CLF, FRD, BOU, TUC, MMB, VSS and AMS.

In this paper, we used ground magnetic measurements to study the variance between different magnetic stations. We choose seven stations that belong to the INTERMAGNET programme (http://www.intermagnet.org). The stations considered in this analysis are: Chambon-la-Forêt (CLF), Fredericksburg (FRD), Boulder (BOU), Tucson (TUC), Memambetsu (MMB), Vassouras (VSS) and Martin de Vivies (AMS). The geographic and geomagnetic coordinates of these magnetic stations are given in Table 1.

Table 1. Magnetic stations considered in the analysis.

ABB CODE	GEO LAT	GEO LONG	GEOMAG LAT
CLF	48.03	2.26	49.84
FRD	38.20	-77.37	48.40
BOU	40.13	-105.23	48.40
TUC	32.17	-110.73	39.94
MMB	43.91	144.19	35.35
VSS	-22.40	-43.65	-13.29
AMS	-37.80	77.57	-46.40

Source: http://swdcwww.kugi.kyoto-u.ac.jp/wdc/obsdata .html (2009)

In order to develop this analysis, one hour time resolution magnetograms obtained at these seven stations were used. The datasets have been obtained from the INTERMAGNET. A five-day interval of the total geomagnetic field (F component) was considered as dataset as shown by Figure 2.

The period of analysis is from 13 to 17 of July, 2000. During these days occurred a very intense storm or superstorm. This storm started on July 15, 2000 and the main phase reached a minimum Dst = -301 nT at 01:00 UT on July 16, 2000. Figure 3 show the Dst index for the whole month of July that belong to solar cycle 23 (period of solar maximum).

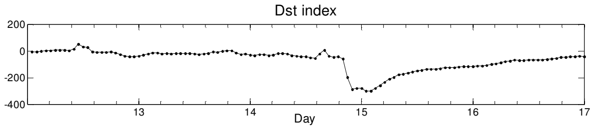

Figure3: Dst index for July 13 - 17, 2000.

Methodology

The methodology to process the magnetograms was based on the following steps:

(1) The data set was organized as an $m \times n$ matrix, where m is the number of stations and n is the time series of the total geomagnetic field observations for five-day interval with 1 hour time resolution.
(2) The mean (x_i) was subtract off for each row of observation.
(3) The correlation matrix between the stations was calculated.
(4) The eigenvectors of the correlation matrix were calculated (Principal Components).
(5) The eigenvalues are calculated and they represent the variance of the particular PC types.

The first eigenvectors of PCs represent the mostly the variability associated with the magnetic station which are characterized by a higher variance. In this analysis, we choose the first three principal components (PCs) because they contained most of the variance.

The analysis was performed using only the data from the stations that compound the asymmetric and symmetric indice (six stations) plus the data from the Vassouras station.

Results and discussion

In this section, we will discuss the PCs analysis for the set of data before mentioned. In Figure 4, we show the correlation matrix between the stations: CLF, FRD, BOU, TUC, MMB, VSS and AMS. The correlation matrix lists each of stations down in the column and across in the row. So the first column represents the station of Fredericksburg, the second, the station of Boulder and so on. The same happens for the rows. The diagonal of the correlation matrix always consists of ones. That's because these are the correlations between each variable and itself. To locate the correlation for any pair of station, find the value in the table for the row and column intersection for those two stations. The lower part of the symmetric correlation matrix of the stations is

$$C = \begin{bmatrix} 1 & & & & & & \\ -0.199 & 1 & & & & & \\ -0.282 & 0.809 & 1 & & & & \\ -0.196 & -0.705 & -0.285 & 1 & & & \\ 0.011 & -0.824 & -0.833 & 0.485 & 1 & & \\ -0.094 & -0.845 & -0.838 & 0.527 & 0.801 & 1 & \\ 0.355 & -0.851 & -0.830 & 0.383 & 0.676 & 0.635 & 1 \end{bmatrix}$$

Figure 4, 5 and 6 show the pattern for the first three most important PCs calculated for the six stations that compound the asymmetric and symmetric indice (CLF, FRD, BOU, TUC, MMB, AMS) plus the Vassouras station. The map used to illustrate the station was taken from the WCD-Kyoto site (1999). The values for the five first principal components are shown in Table 2. The PCs code was written by Shlens J. (2005) using the GNU/Octave program. For computing the eigenvalues and eigenvectors, GNU/Octave uses a several step process which began with an Hessenberger decomposition, followed by a Schur decomposition.

Table 2. The three first EOFs for the six stations plus Vassouras station.

ABB CODE	PC 1	PC 2	PC 3	PC 4	PC 5
CLF	0.075	0.794	0.386	0.415	-0.081
FRD	-0.457	0.017	-0.242	0.017	0.070
BOU	-0.429	-0.208	0.346	-0.038	0.119
TUC	0.287	-0.462	0.716	0.120	0.005
MMB	0.421	-0.087	-0.263	0.372	0.736
VSS	0.421	-0.188	-0.288	0.216	-0.646
AMS	0.406	0.264	0.087	-0.791	0.123

We observe that the spatial pattern for the first PCs present the same phase oscillation for almost all geomagnetic stations except for the FRD and BOU. The PCs lower amplitudes belong to the CLF and TUC. For the second PCs, CLF has the larger amplitude of oscillation. In the third PCs, FRD, MMB and VSS are oscillating in phase with similar amplitudes. Also, BOU and CLF are oscillating in phase with similar amplitudes. The fourth and fifth PCs present quite dissimilar pattern,

Final remarks

The EOF technique is a very useful method for compacting large data sets and for diagnosing the dominant patterns of variability in geophysical data sets.

The main conclusions in this analysis can be summarized as follow:

- In Figure 4, we could observe that the stations of the north hemisphere are oscillation in phase opposition with the stations of the south hemisphere. BOU and FRD were oscillating in opposition of phase with the rest of the stations.

This may suggest a possible longitudinal behavior.

- The first PC corresponds 65% of the total variance, approximated.
- In Figure 5 the behavior corresponds to 20% of the total variance, the second PC. We can observe that CLF presented the larger amplitude of oscillation. Also, it is possible to suggest a latitudinal oscillation behavior.
- Comparing the variance between the two first PCs, we can notice that the longitudinal effect is 3 times more than the latitudinal effect.
- For the third PCs, FRD, MMB and VSS; and BOU and CLF were oscillating in phase with similar amplitudes
- The fourth and fifth PCs showed quite dissimilar pattern.

The results obtained are encouraging. But the present study dealt with just a few stations and not enough data (just 5 days). In the next step, we will present a further study using more data and more stations to do a complete statistical analysis. So, the physical process involved will be analyzed.

The first interpretation of the results suggest that PCs can be used to characterize the statistical relationships between magnetic stations, but need further study.

Acknowledgments

V. Klausner wishes to thanks CAPES for the financial support of her PhD. The authors are very thankful to Eng. V. E. Menconi for the computing support. Also,the authors would like to thank CAPES/PAEP (proc. number: 0880/08-6), CNPq (proc numbers: 486165/2006-0, 308680/2007-3 and 309017/2007-6) and FAPESP (proc number: 2007/07723-7), and INTERMAGNET programme for the datasets used in this work.

References

Campbell W. H., 1997, Introduction to geomagnetic fields, Cambridge University Press, Cambridge, UK.

Domingues M. O., 1993, Métodos de análise de séries temporais geofísicas: aplicações no estudo das variações interanual e interdecadal do nível médio do mar no Atlântico Norte Extra-tropical: tese (Mestrado) - INPE, 1993. São José dos Campos, SP: INPE, 1993 9898318 (INPE-5549-TDI/535).

Gonzalez, W. D., J. A. Joselyn, Y. Kamide, H. W. Kroehl, G. Rostoker, B. T. Tsurutani and V. M. Vasyliunas, 1994, What is a geomagnetic storm?, Journal of Geophysical Research, vol 99, NO. A4, pages 5771 – 5792.

Keiner, L. E., and X. H. Yan, 1997, Empirical Orthogonal Function Analysis of Sea Surface Temperature Patterns in Delaware Bay, IEEE Transactions on Geoscience and Remote Sensing, vol 35, NO. 5.

Mendes Jr., O., M. O. Domingues, A. Mendes da Costa and A. L. C. Gonzalez, 2005, Wavelet analysis applied to magnetograms: Singularity detections related to geomagnetic storms, Journal of Atmospheric and Solar-Terrestrial Physics, vol 67, pages 1827 – 1836.

Murray, G. W., J. C. Muller and H. J. Zwally, 1984, Matrix portioning and EOF/Principal Components Analysis of Antartic Sea Ice Brightness Temperatures, Technical Notes NOAA.

Parkinson W. D., 1983, Introduction to geomagnetism, Scottish Academic Press Ltd., Edinburgh and London, UK.

Preisendorfer R. W., 1988, Principle Component Analysis in Metereology and Oceanography, Elsevier, New York, USA.

Shlens J., 2005, A tutorial on Principal Component Analysis, http://www.snl.salk.edu/~shlens/pub/notes/pca.pdf.

Tascione T. F., 1988, Introduction to the space environment, Orbit Book Company, INC., Florida, USA.

WCD-Kyoto, 2008, http://swdcwww.kugi.kyoto-u.htmlu.ac.jp/index.html

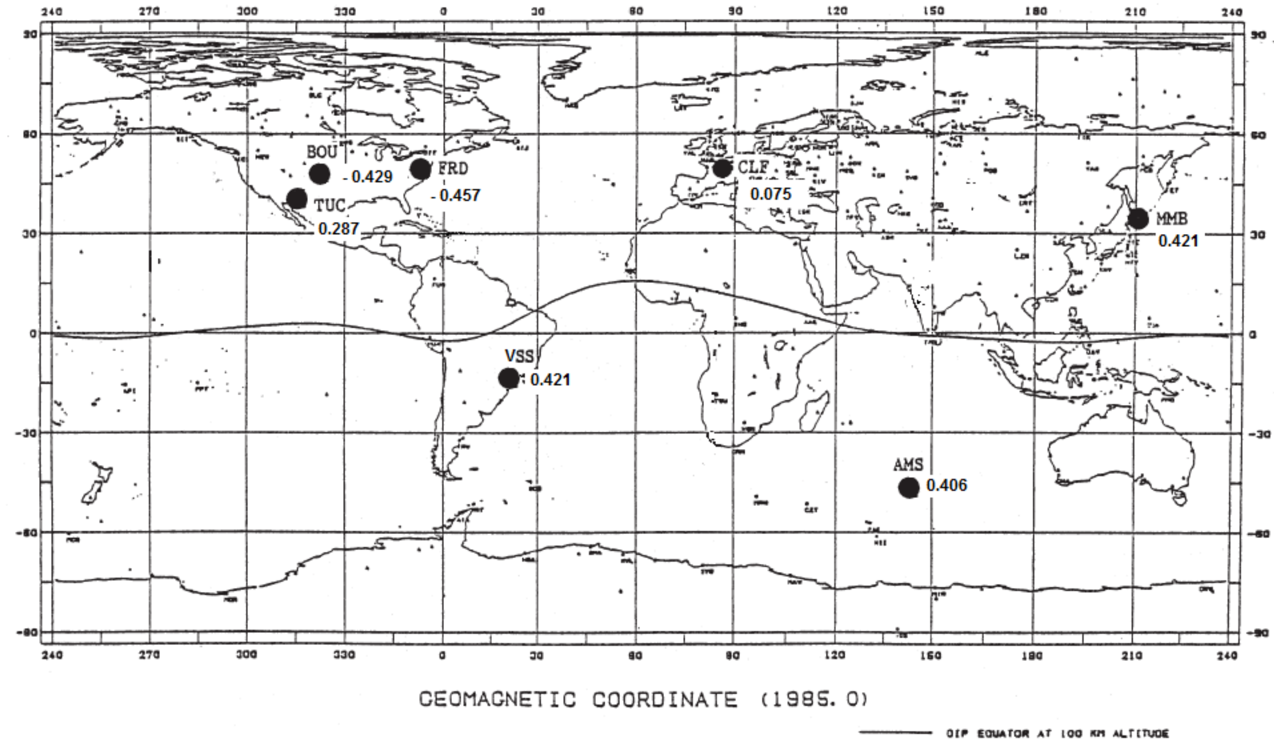

Figure 4: The spatial pattern of the first principal component at the seven geomagnetic stations. The first PC explain of the 64.68% variance.

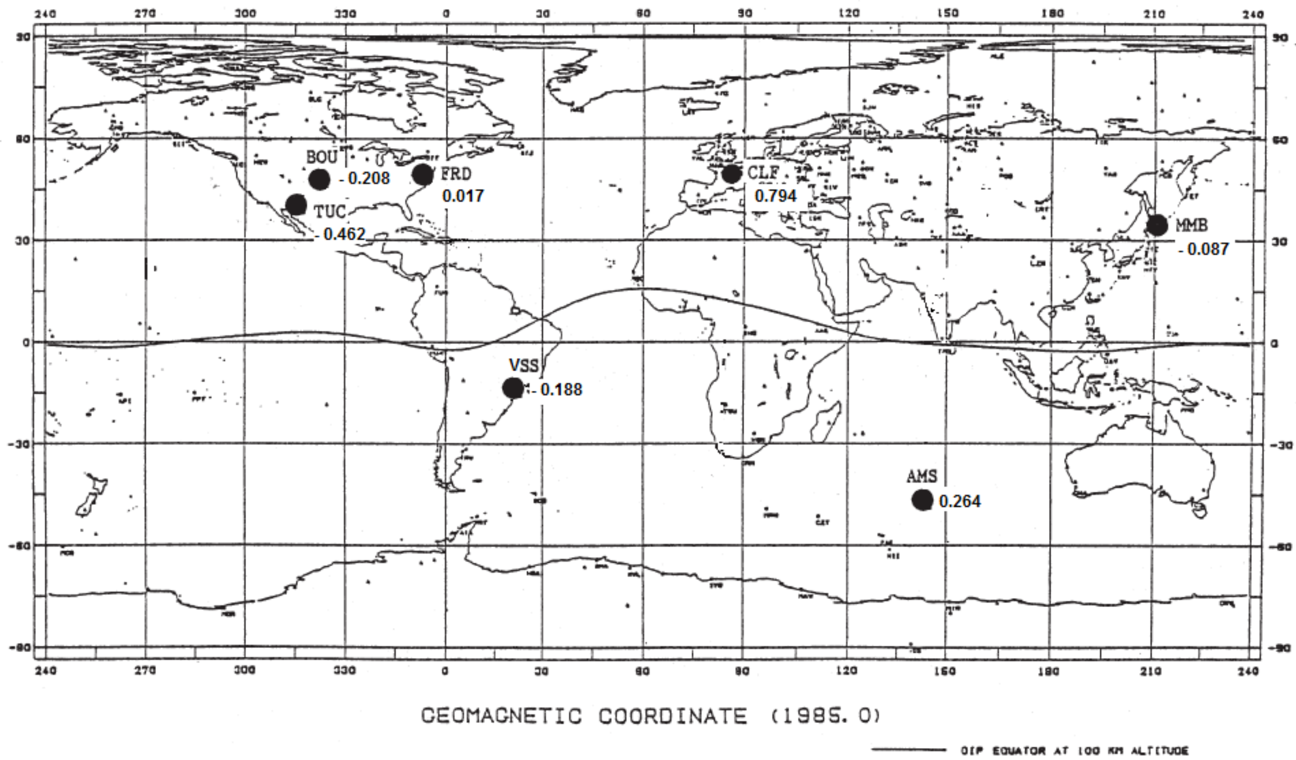

Figure 5: The spatial pattern of the second principal component at the seven geomagnetic stations. The second PC explain of the 18.90% variance.

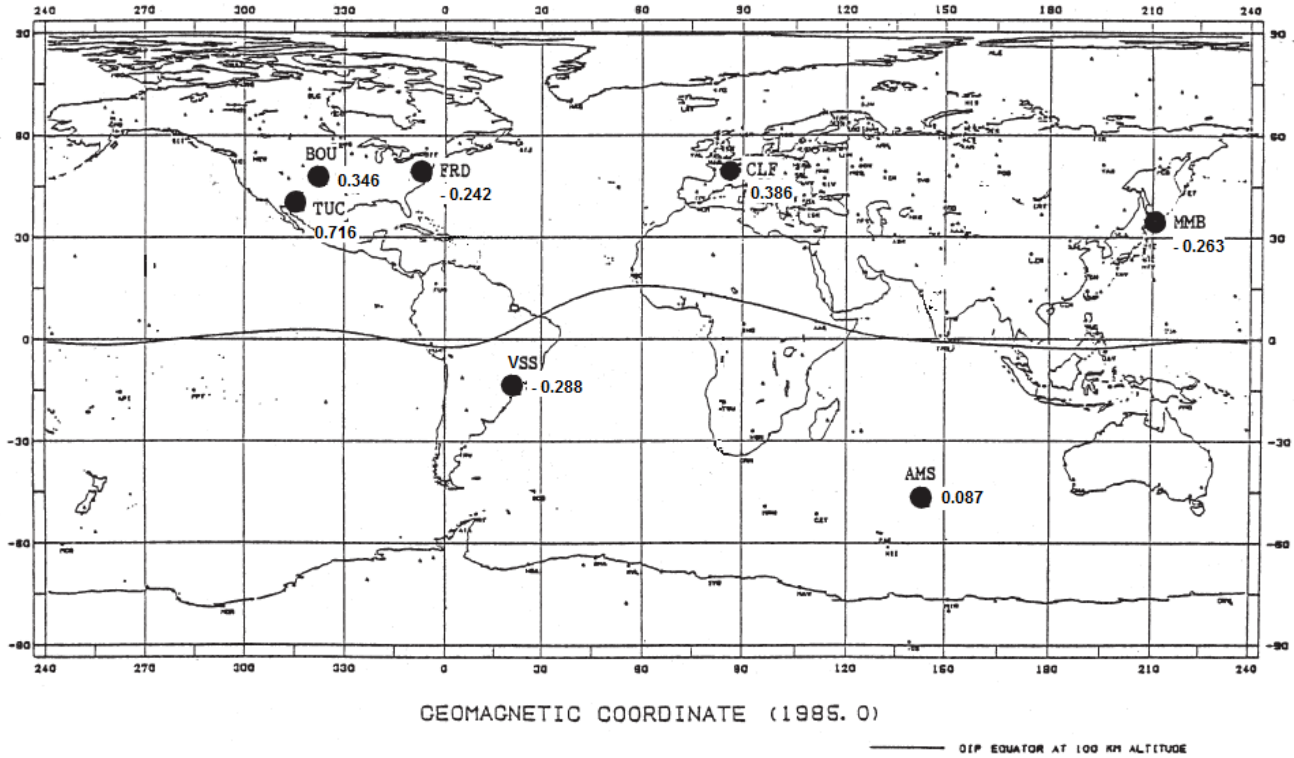

Figure 6: The spatial pattern of the thrid principal component at the seven geomagnetic stations. The third PC explain of the 9.31% variance.