USGS
science for a changing world

Report of Geomagnetic Pulsation Indices for Space Weather Applications

By Z. Xu, J.L. Gannon, and E.J. Rigler

Open-File Report 2013–1166

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2013

Suggested citation:
Xu, Z., Gannon, J.L., and Rigler, E.J., 2013, Report of geomagnetic pulsation indices for space weather applications: U.S. Geological Survey Open-File Report 2013–1166, 22 p., http://pubs.usgs.gov/of/2013/1166/.

Contents

Figures

Tables

Report of Geomagnetic Pulsation Indices for Space Weather Applications

By Z. Xu, J.L. Gannon, and E.J. Rigler

Introduction

The phenomenon of ultra-low frequency (ULF) geomagnetic pulsations was first observed in the ground-based measurements of the 1859 Carrington Event (Stewart, 1861) and has been studied for over 100 years. Pulsation frequency is considered to be "ultra" low when it is lower than the natural frequencies of the plasma, such as the ion gyrofrequency. ULF pulsations are considered a source of noise in some geophysical analysis techniques, such as aeromagnetic surveys (Wanliss and Antoine, 1995) and transient electromagnetics, so it is critical to develop near real-time space weather products to monitor these geomagnetic pulsations. The proper spectral analysis of magnetometer data, such as using wavelet analysis techniques, can also be important to Geomagnetically Induced Current (GIC) risk assessment.

The classification of ULF pulsations as Pc1, Pc2, Pc3, Pc4, and Pc5 for continuous pulsations and Pi1 and Pi2 for irregular pulsations is shown in table 1 (Jacobs and others, 1964). The sources of pulsations are different for each frequency band. Higher-frequency pulsations are caused by fluctuations and instabilities in the equatorial ionosphere and magnetosphere. Lower frequency pulsations are caused by other complicated phenomena, such as a local wave-particle instability or from coupling of wave energy propagating through the magnetosphere; they are produced either in the solar wind/magnetosheath or at the magnetopause/boundary layer (Anderson, 1993, 1994). Pulsations can be measured using ground-based magnetometers or space-based *in-situ* observations of magnetic and electric fields.

Table 1. The classification of geomagnetic pulsations. (s, seconds; mHz, megahertz)

	Pulsation Classes						
	Continuous pulsations					Irregular pulsations	
	Pc1	Pc2	Pc3	Pc4	Pc5	Pi1	Pi2
T	0.2–5s	5–10s	10–45s	45–150s	150–600s	1–40s	40–150s
f	0.2–5Hz	0.1–0.2Hz	22–100mHz	7–22mHz	2–7mHz	0.025–1Hz	2–25mHz

In order to monitor geomagnetic pulsations, indices with different frequency bands (Pc3, Pc4, and Pc5) are developed using 1-second (sec) geomagnetic data filtered by wavelet analysis. These indices are being developed as a part of the Geomagnetic Hazard Map project at the U.S Geological Survey (USGS) Geomagnetism Program. A correlation study of the wavelet-based indices is also carried out for pairs of different geomagnetic observatories. In addition to the indices, a prototype disturbance map is produced by applying interpolation between pulsation indices at geomagnetic observatories over the continental United States. For space weather applications, the map is an initial

step to present the spatial changes of geomagnetic disturbance patterns within certain frequency bands. Future improvements may include using data with higher spatial resolution and applying more advanced interpolation methods. In addition, the USGS Geomagnetism Program is capable of collecting, transporting, and disseminating the geomagnetic data from 14 observatories in real time with high temporal resolution and accuracy, which enables the development of pulsation indices for real-time monitoring.

Data and Methods

Raw Geomagnetic Data

The data used in this project are USGS 1-sec ground magnetometer data including: (1) one quiet period (12/03–12/04 or day 338–339) in year 2008, (2) one storm period (03/08–03/10 or day 068–070) in year 2008, and (3) one storm period (10/24–10/27 or day 297–300) in year 2011. The data used for this study were collected from 13 USGS-operated observatories (listed in table 2). These data are sampled at 100 hertz (Hz) and averaged to 1-sec values, with a resolution of at least 0.01 nanoTesla (nT). Examples of the data are shown in appendix 1.

The data from the year 2008 are definitive data, from which the artificial spikes are removed and missing data are flagged. The data from year 2011 are preliminary data with some artificial spikes. These raw data need to go through a pre-processing for missing or spiking data before applying further treatment.

Table 2. Observatories list.

Station ID	Station Name	Latitude	Longitude	Geomagnetic Latitude	Geomagnetic Longitude
BOU	Boulder	40.14°N	105.24°W	48.06°N	38.69°W
BSL	Stennis	30.35°N	89.63°W	39.66°N	19.63°W
BRW	Barrow	71.32°N	156.62°W	69.68°N	112.56°W
CMO	College	64.87°N	147.86°W	65.36°N	97.24°W
FRD	Fredericksburg	38.20°N	77.37°W	47.95°N	6.06°W
FRN	Fresno	37.09°N	119.72°W	43.24°N	54.03°W
GUA	Guam	13.59°N	215.13°W	5.48°N	143.91°W
HON	Honolulu	21.32°N	158.00°W	21.59°N	89.70°W
NEW	Newport	48.27°N	117.12°W	54.57°N	54.45°W
SIT	Sitka	57.06°N	135.33°W	60.20°N	78.62°W
SHU	Shumagin	55.35°N	160.46°W	54.24°N	102.36°W
SJG	San Juan	18.11°N	66.15°W	27.92°N	6.53°E
TUC	Tucson	32.17°N	110.73°W	39.56°N	43.23°W

Data Pre-Processing for Missing Data or Spikes

To correct for missing data or spikes, the H and Z components of the magnetic field are interpolated with a “Piecewise cubic Hermite interpolation” method (Fritsch and Carlson, 1980). The results show that the spikes (which are recognized by a threshold algorithm) and missing data (which are flagged as 999999) are effectively removed from the raw data. An example is shown as figure 1. The

pre-processed geomagnetic data are then ready for the next step: frequency separation through wavelet analysis.

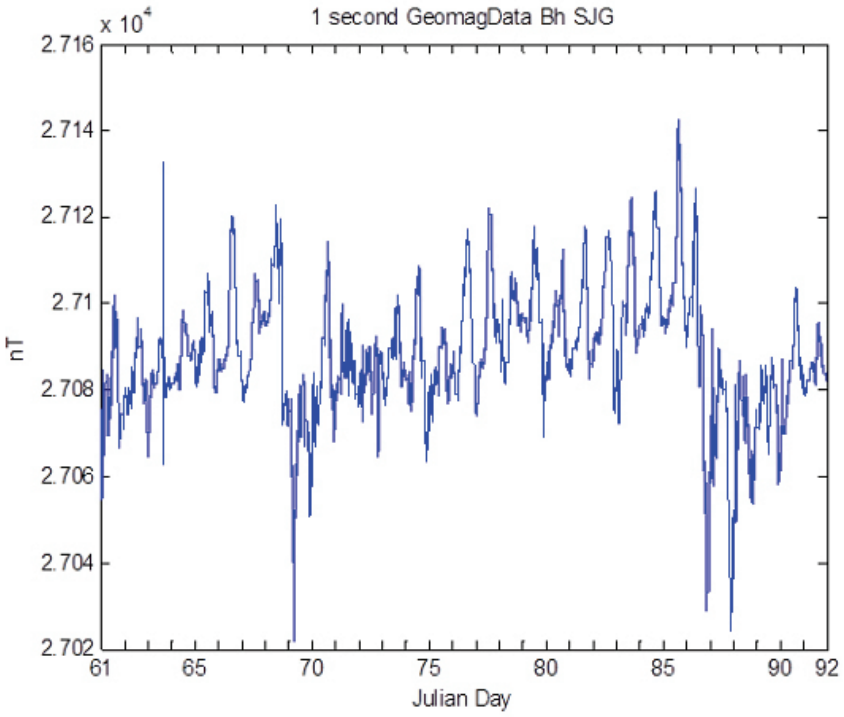


Figure 1. An example of pre-processed geomagnetic data (H component of B field) from the San Juan (SJG) magnetic observatory from Julian day 61–92, 2008. The spikes are removed and filled using the cubic method. (BH, horizontal magnetic field)

Wavelet Analysis

Geomagnetic data contain different spectral components that are related to complicated current systems in the ionosphere and magnetosphere. To separate these spectral components, we use wavelet analysis on the cleaned data. This decomposes the original data into different frequency bands while keeping the localized time-domain information. Wavelet analysis is a suitable tool for such data, which exhibits impulsive, multi-scale, and other non-stationary spectral features. This technique includes a wide range of applications, such as cross-wavelet analysis, multi-resolution analysis (MRA), and timescale analysis (Torrence and Compo, 1998). We use wavelet analysis to filter and decompose the geomagnetic data into different time series of specific frequency bands of interest for geomagnetic pulsation studies. These time series are called "details" and "smooth" (Percival and Walden, 2000). The detail D_i corresponds to frequencies in the range of 2^{-i-1} to 2^{-i} cycles per second, which corresponds to physical time scales between 2^i and 2^{i+1} sec (where i is 1,2,3…, the level of details). The smooth S_i is composed of the sample mean and frequency components below 2^{-i-1} cycles per second, and is smooth in appearance compared to the original time series. Figure 2 shows an example of separated frequency band output due to wavelet analysis decomposition. The top panel shows that the original geomagnetic data is decomposed into time varying components (details) with different frequency bands; the time-localized features are still in each band.

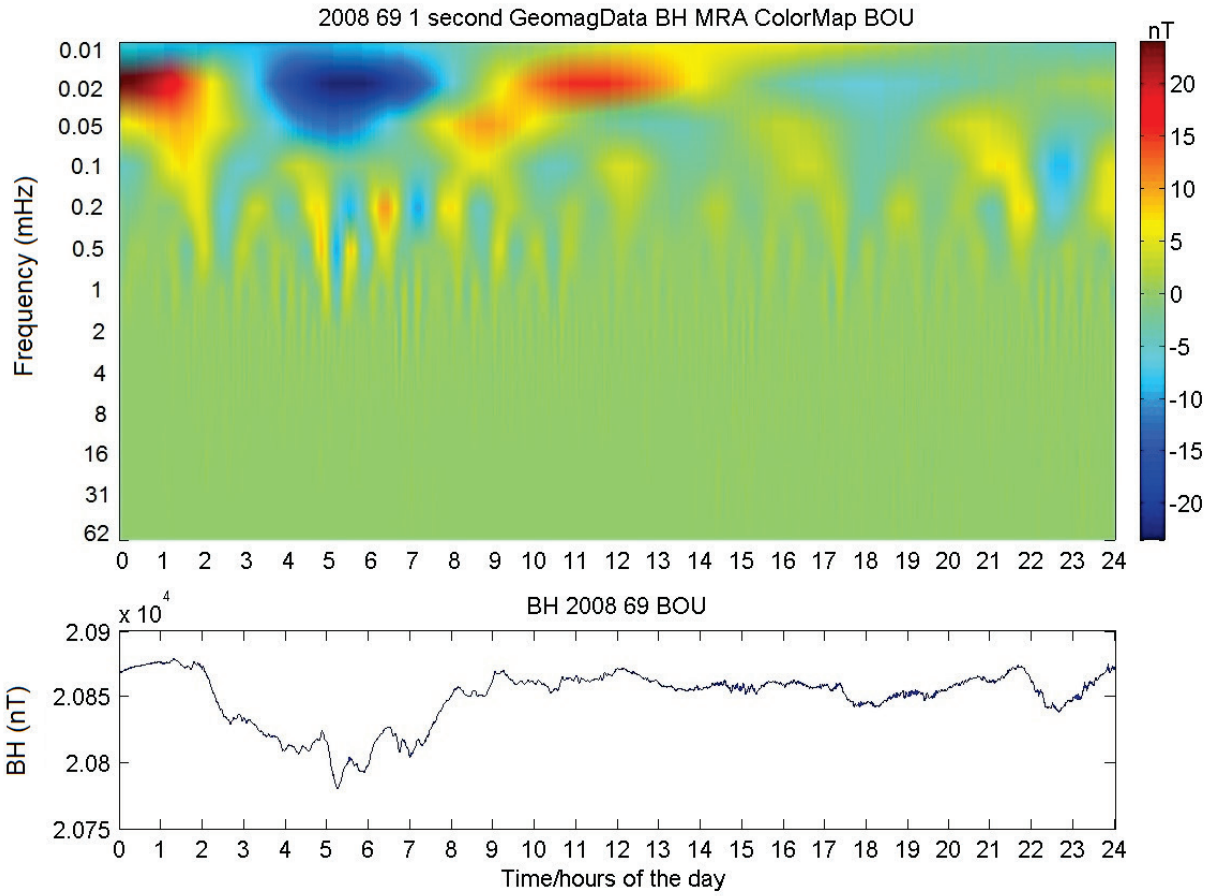


Figure 2. The original geomagnetic data is decomposed into variations which have different periods. The top panel shows the decomposed variations with different periods in BH; the bottom panel shows the original data of BH component on Day 69, 2008, at Boulder (BOU) Observatory. (nT, nanoTesla; BOU, Boulder; BH, horizontal magnetic field)

The wavelet package used in this analysis is the "Maximum Overlap Discrete Wavelet Transform" (MODWT) in "waveslim" package written in the R language (*http://www.r-project.org/*). The MODWT addresses some shortcomings of the Discrete Wavelet Transform (DWT), such as sample size restriction and sensitivity to the starting points of signal series (Xu and others, 2008). The procedures of wavelet analysis are as following:

1) Load the cleaned data (in this case, a three-day span of data) into the wavelet processing program and proceed with wavelet MRA analysis.

2) Perform MODWT wavelet analysis to get details (D_i) and smooth (S_i) of MRAs.

3) Output the MRA results into formatted txt-files (detailed codes in appendixes 2 and 4).

Note that, although the input data samples for this wavelet MRA analysis are two to four days, the input data sample can be as short as 1 hour (hr) to study the geomagnetic pulsations when dealing with the near-real-time circumstances, because the high time-resolution (1 sec) data provide enough data points for the wavelet analysis. A flow chart for the calculation process is shown in appendix 2.

After the cleaned geomagnetic data are split into frequency bands using wavelet analysis, geomagnetic pulsation indices are calculated for the specific frequency bands of interest. Considering the frequency bands for different geomagnetic pulsations, the Pc3 index is developed based on the details D4 and D5 of the MRAs, which contain the range of periods between 10–45 sec; the Pc4 index is developed based on the details D6–D7 of the MRAs, which contain the range of periods between 45–150 sec; and the Pc5 index is developed based on the details D8–D9 of the MRAs, which contain the range of periods between 150–600 seconds (table 3).

Table 3. Geomagnetic pulsations and their related wavelet details. (s, seconds; mHz, megahertz)

Pulsation	Period	Frequency	Wavelet Details	Wavelet Periods
Pc3	10–45s	22–100mHz	D4–D5	16–64s
Pc4	45–150s	7–22mHz	D6–D7	64s–256s
Pc5	150–600s	2–7mHz	D8–D9	256s–1,024s

Results and Validation

Pc3, Pc4, and Pc5 Indices for Different Observatories

Examples of these indices are shown in figure 3. Figure 3 *A*, *B*, and *C* show the geomagnetic disturbances within the frequency bands in Pc3, Pc4, and Pc5 ranges, respectively, for three mid-latitude observatories of Boulder (BOU), Bay Saint Louis (BSL), and Fredericksburg (FRD) from Day 68 to 70 in 2008. The disturbances are compared with the USGS 1-minute Dst geomagnetic disturbance index, which indicates there was a medium storm on Day 69 with a magnitude of about –80 nT. The magnitude of geomagnetic pulsation is about 0.5–1 nT for Pc3, 1–1.5 nT for Pc4, and over 2 nT for Pc5. During this storm, the pulsations happened simultaneously at these mid-latitude observatories. There are Pc4 and Pc5 pulsations but no Pc3 at the sudden commencement around 03:00UT on 2008–69. This is because the source of Pc3, which is related to the currents in the equatorial magnetosphere, is different from the source of Pc4 and Pc5, which is in the solar wind/magnetosheath or magnetopause/boundary layer.

The specific wavelet transform, MODWT, that is applied in this project has the advantages of sample-size restriction and insensitivity to the starting points of signal series, making it more flexible on data-span requirements and not as sensitive to the starting points of the signal series as the regular DWT. These advantages of wavelet analysis plus the real-time USGS magnetometer data facilitate using these indices as near real-time monitoring tools. These indices can be produced as a daily or hourly online data display chart for space weather applications.

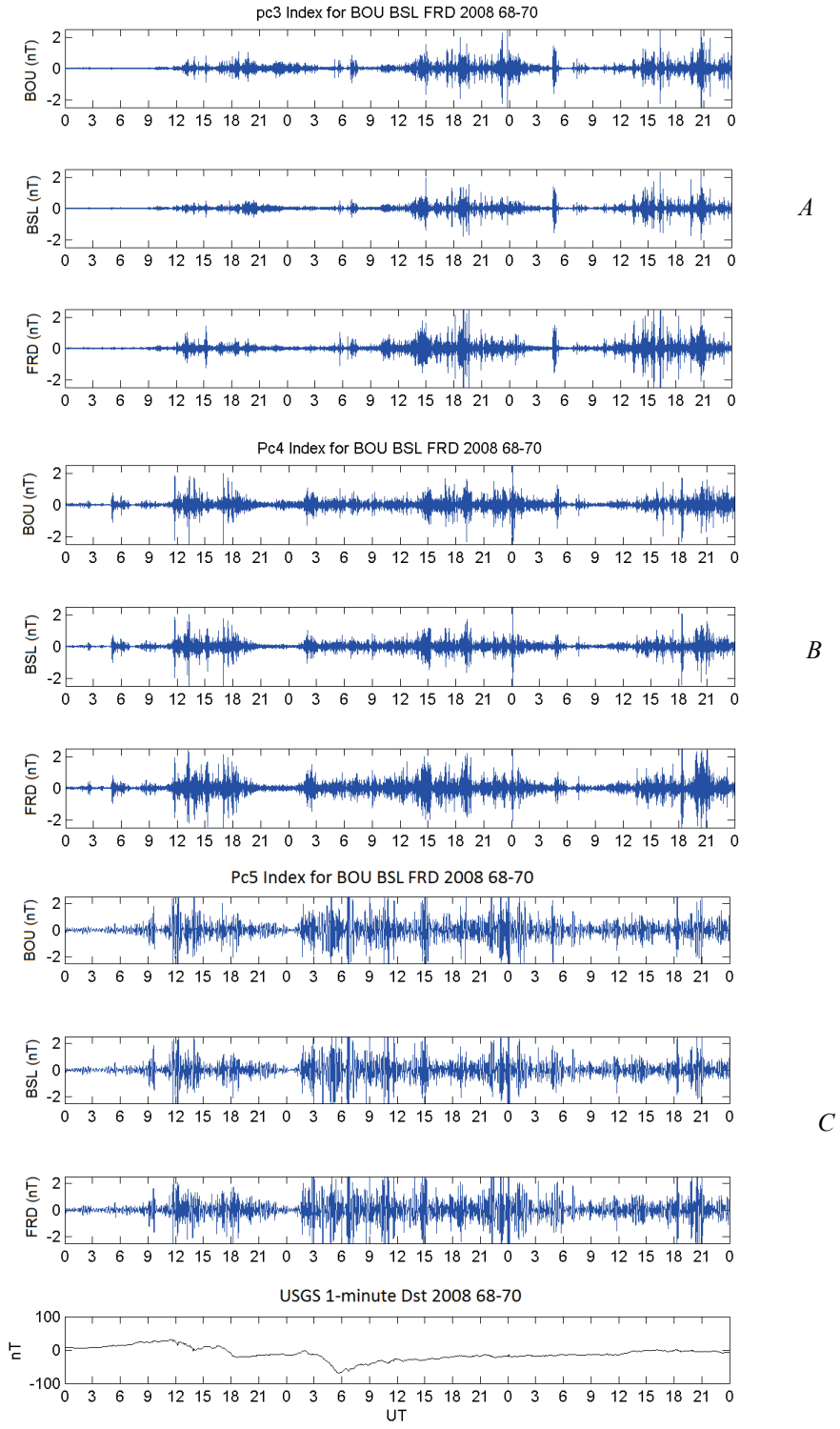


Figure 3. The geomagnetic disturbances within the frequency bands in *A*, Pc3; *B*, Pc4; and *C*, Pc5 range for three mid-latitude observatories of BOU, BSL, and FRD from Day 68–70 in 2008. (BOU, Boulder; BSL, Bay Saint-Louis; FRD, Fredericksburg; nT, nanoTesla; UT, Universal Time)

Correlation Study for Variations of Different Frequency

In order to study the relationship of geomagnetic variations at different frequency bands at different geographic locations, a correlation study was performed with the decomposed data extracted using wavelet analysis. The correlation coefficients between geomagnetic variations of the same frequency bands at different observatories are calculated for a quiet period, day 338 to 339 in 2008, and a storm period, day 68 to 70 in 2008. Figure 4 A and B show the correlation coefficients for the horizontal components of the measured magnetic field (BH) with different frequency bands between Boulder and Bay Saint Louis (BOU–BSL), Boulder and Fresno (BOU–FRN), and Boulder and Tucson (BOU–TUC), which hold the highest correlation among all observatories.

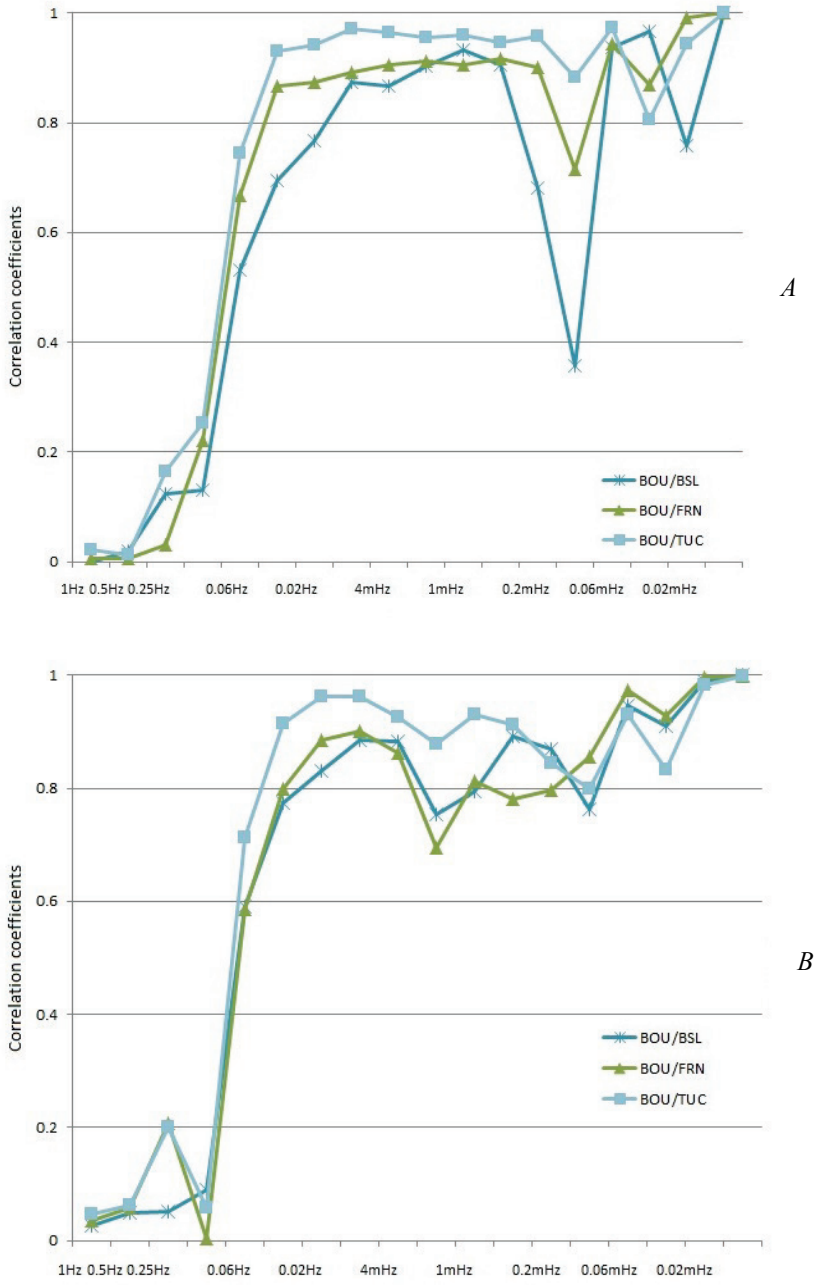


Figure 4. *A*, the correlation coefficients for BH with different frequency bands between BOU–BSL, BOU–FRN, and BOU–TUC for the quiet period, 2008-338–339; *B*, the correlation coefficients for BH with different frequency bands between BOU-BSL, BOU-FRN, and BOU-TUC for the storm period, 2008-68–70. (mHz, megahertz; BOU, Boulder; TUC, Tucson ; FRN, Fresno; BSL, Bay Saint-Louis)

The results show that for both storm and quiet periods, for the horizontal component BH, these three pairs of observatories are highly correlated. The pattern of correlation changes with frequency bands for these two periods but exhibits similarities in general, such as: (1) low-correlation coefficients for high-frequency bands over 0.06 Hz, which is in the mixing range of Pc1 and Pc2; (2) high-

correlation coefficients for low-frequency bands below 22 megahertz (mHz), which is in the diurnal variation range; and (3) high correlation coefficients for 2 mHz–22 mHz, which is in the range of Pc4 and Pc5. There is an obvious increase of correlation coefficients around 60 mHz, which is related to disturbances with periods of 4 hr during the storm. This could be explained by the global enhancement of ring current during storm periods.

The same correlation study is implemented for all 13 observatories, which results in 78 pairs of observatories. For observatories at low- and mid-latitudes, the correlation coefficients of BH are higher when the pairs of observatories are relatively close to each other in distance, such as BOU–TUC and FRD–BSL. These characteristics of correlation varying with frequency are useful for map interpolations, which will be discussed in next section.

The correlation coefficients of the vertical component of the magnetic field (BZ) are more complicated, differ from observatory to observatory, and need to be further investigated with the knowledge of local conductivity environments, magnetotellurics and other geophysical techniques.

Disturbance Map Development

Using the Pc pulsation indices for each observatory, another way to study the relationship between pulsations at different observatories is by looking at the variation in geographic distribution (that is, by building a map). In this way, we present the spatial patterns of geomagnetic pulsations at the same time frame over different locations on the Earth, mainly focusing on the continental United States.

There are several maps of different frequency bands developed including: Pc3, Pc4, and Pc5. Several interpolation methods were applied including: Linear (Triangle-based linear interpolation), Cubic (Triangle-based cubic interpolation), Nearest neighbor interpolation, BiHarmonic Spline (BHS) interpolation (Sandwell, 1987), and Kriging interpolation (Matheron, 1963).

During the testing stage, over 620,000 frames of testing maps were produced for the two storm periods, (2008–68 to 2008–70, and 2011–297 to 2011–300, the format YYYY–DOY), for Pc3, Pc4, and Pc5 frequency bands. We test the accuracy of the interpolation methods by comparing the results of different interpolations for the same frequency and period. The sample results are shown below in Figure 5*A*, *B*, and *C*. Figure 5*A* is the result from linear interpolation. Due to the mathematical properties of linear interpolation, the artificial patterns on the map are easy to recognize. Figure 5*B* is the result of BHS interpolation and 5*C* is of Kriging interpolation. These two methods produced similar results. Due to the mathematical nature of BHS and Kriging, the difference in these two methods is most pronounced near boundary regions (especially the south boundary) on the maps. The BHS method shows continuous variations near boundaries, whereas the results of the Kriging have more impacts coming from the nearest data source. The results show that the Kriging interpolation produces fewer artificial results than other methods.

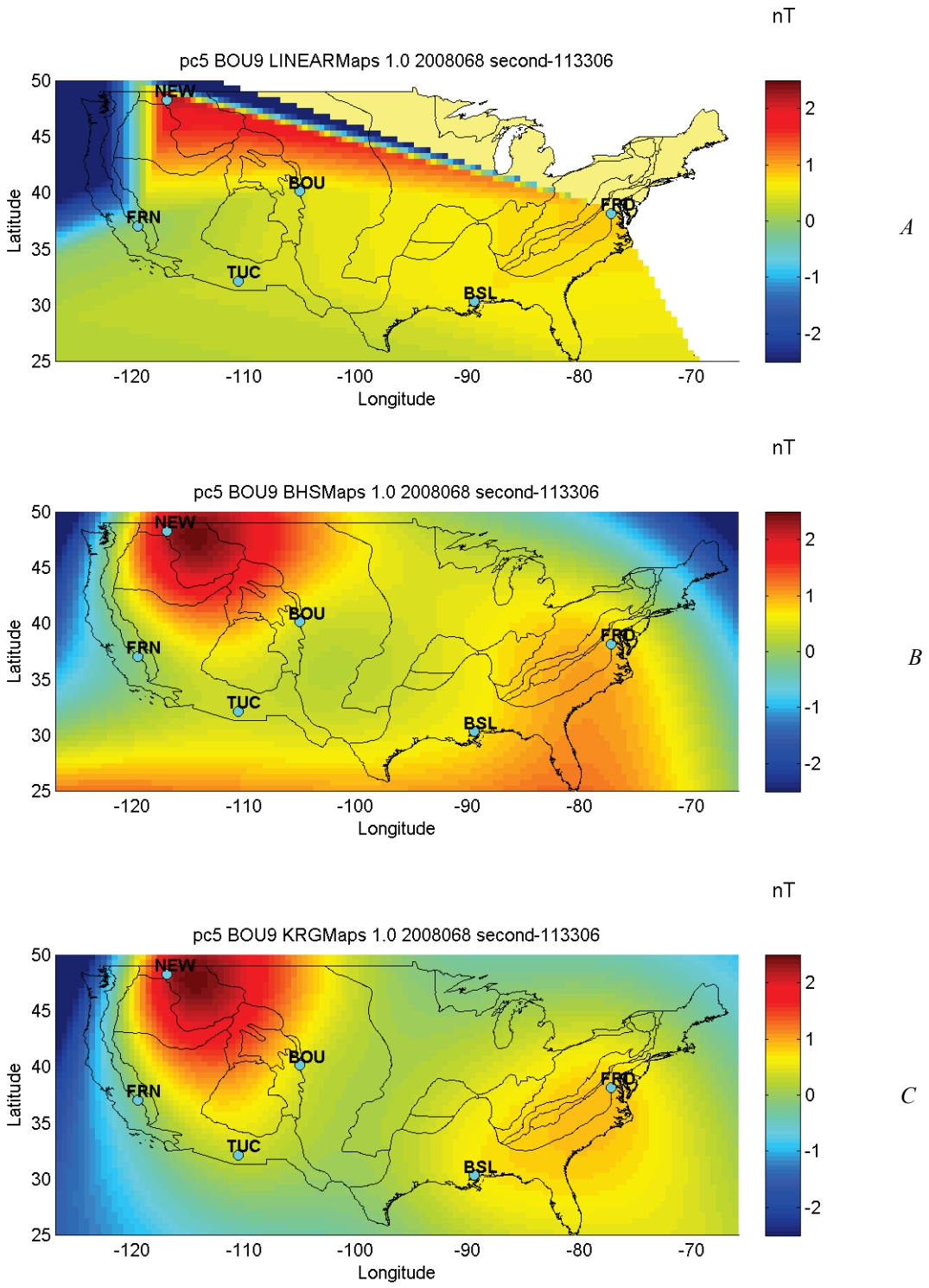


Figure 5. The map of Pc3 pulsation with the interpolation method of A, linear; B, BHS; and C, Kriging for the same time frame on 2008–68. The background lines are the geologic province boundaries of the United States. (BOU, Boulder; FRN, Fresno; TUC, Tucson; BSL, Bay Saint-Louis; FRD, Fredericksburg)

Since the BHS and Kriging results were close by visual comparison, a further detailed validation was performed based on comparison between the interpolated data in the grid near Boulder, Colorado, and the actual magnetic field measurement at Boulder. Examples of the comparison results on 2008–068 are shown in figure 6*A* and *B* for Pc5, figure 7*A* and *B* for Pc4, and figure 8*A* and *B* for Pc3. These scatter plots show the correlations between the interpolations of Pc indices with the original Pc indices from the Boulder observatory (X-axis). The black straight lines are the linear fitting lines.

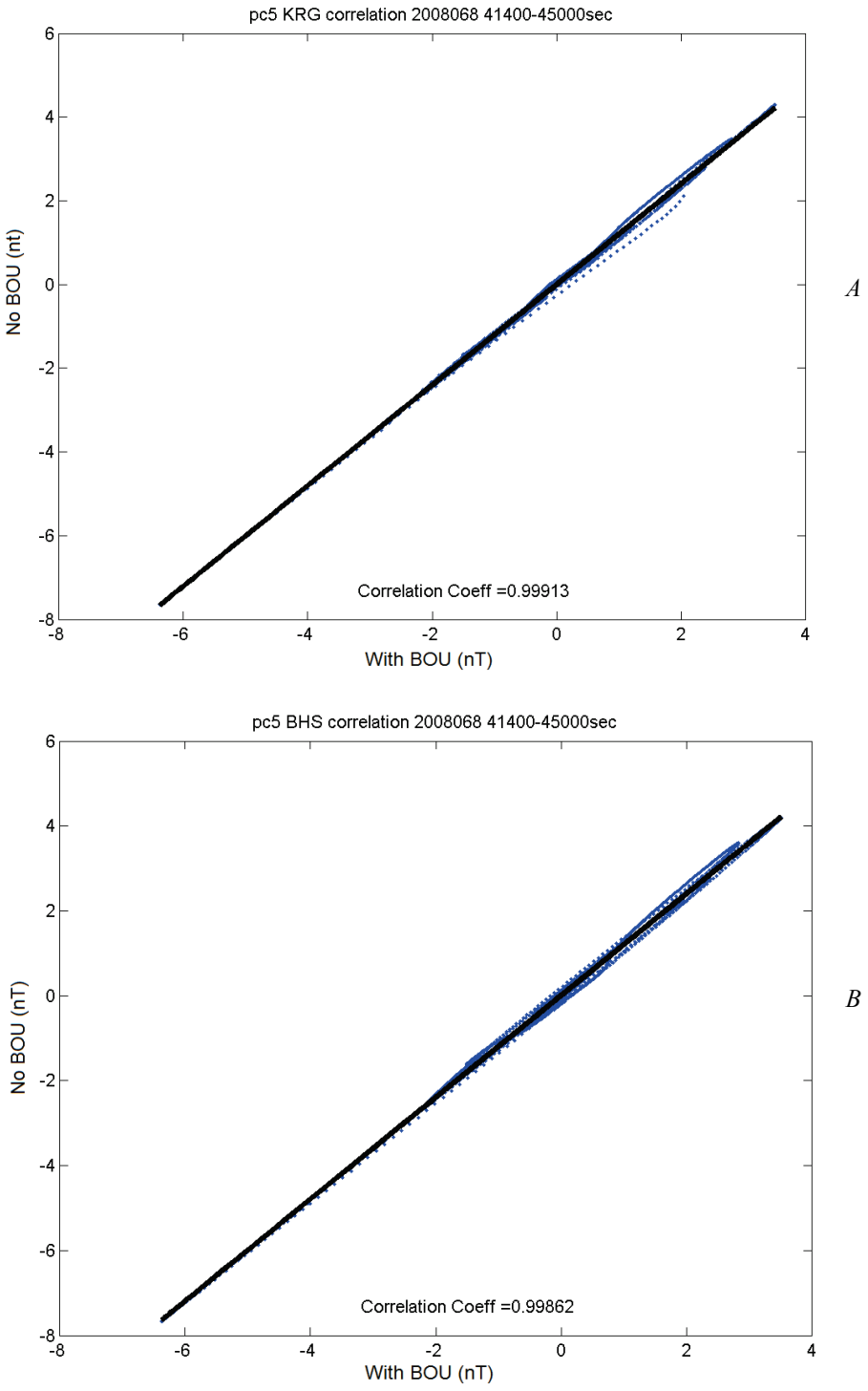


Figure 6. The comparison between original and interpolated Pc5 indices. *A*, Kriging interpolation; *B*, BHS interpolation on 2008-068. (BOU, Boulder; nT, nanoTesla)

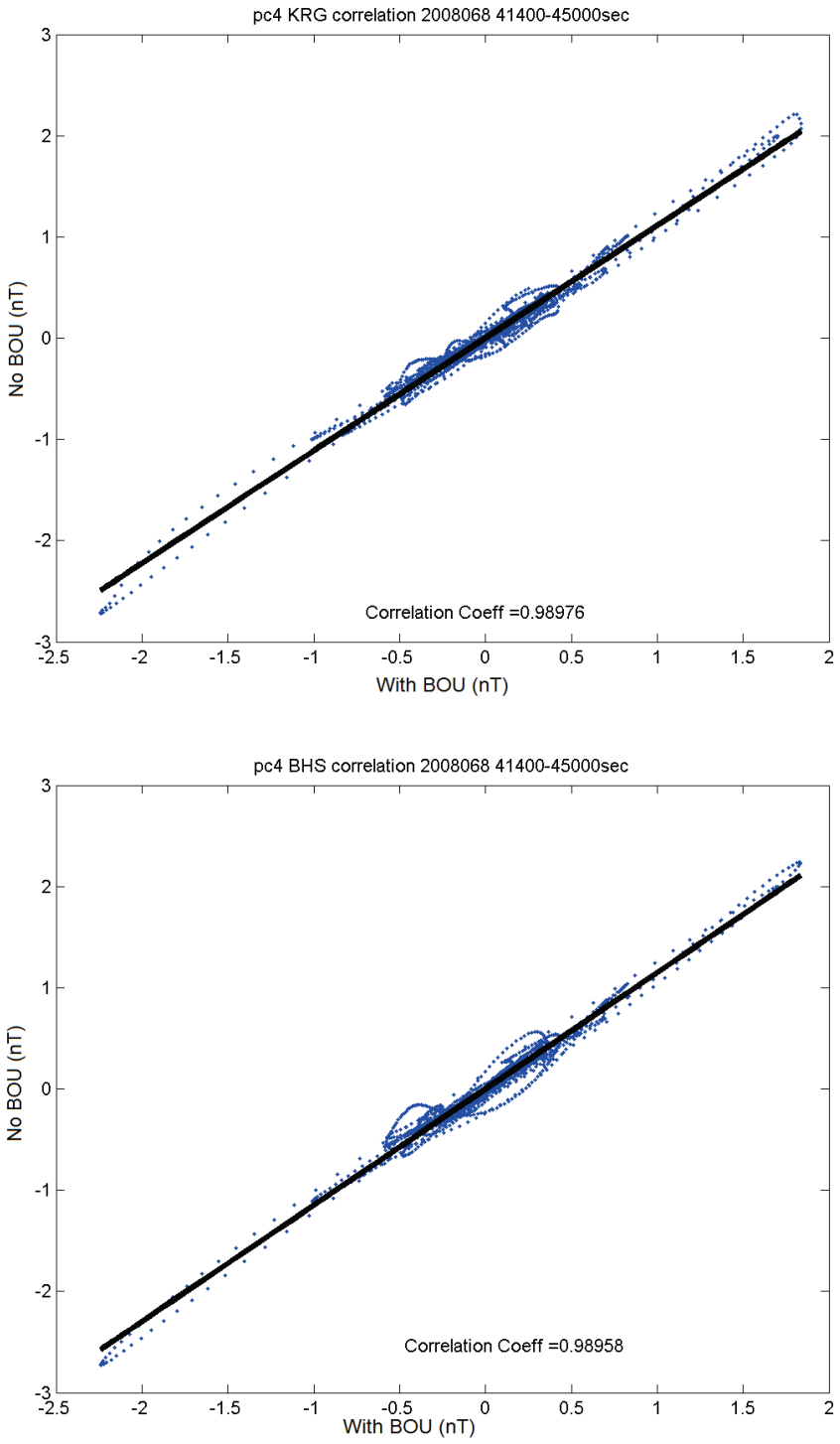
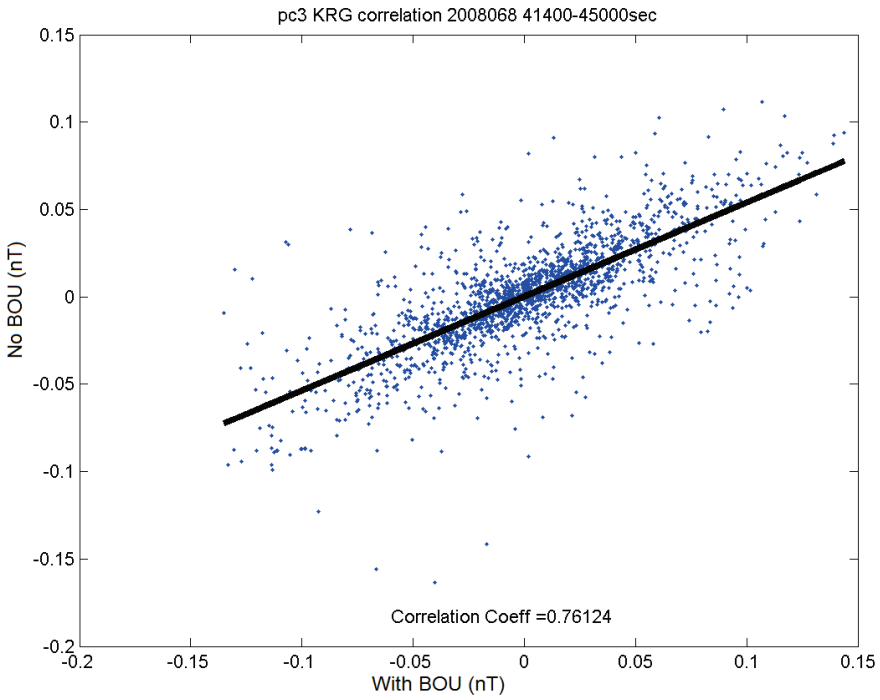
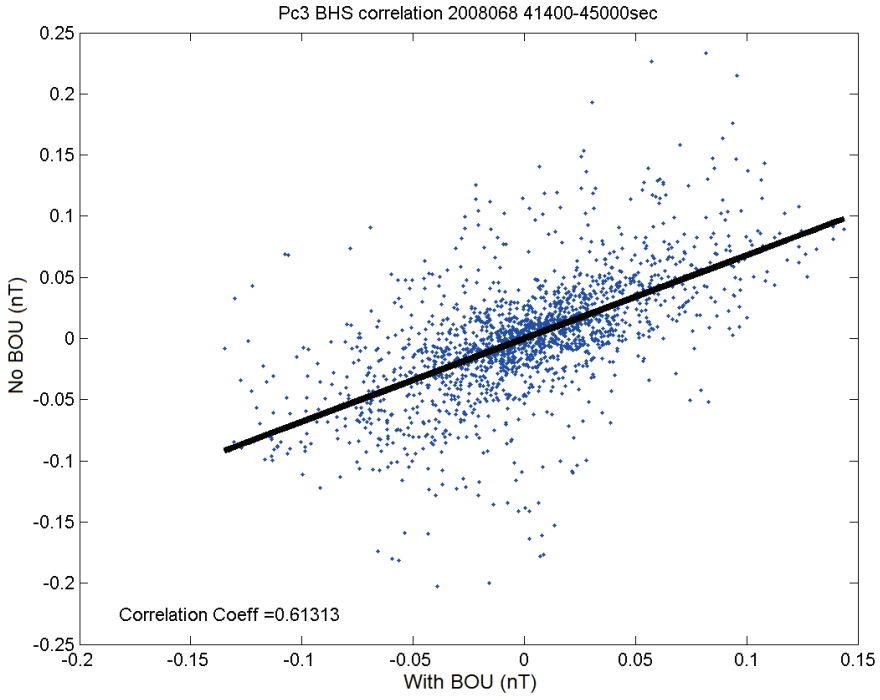


Figure 7. The comparison between original and interpolated Pc4 indices. *A*, Kriging interpolation; *B*, BHS interpolation on 2008–068. (BOU, Boulder; nT, nanoTesla)



A



B

Figure 8. The comparison between original and interpolated Pc3 indices. *A*, Kriging interpolation; *B*, BHS interpolation on 2008-068. (BOU, Boulder; nT, nanoTesla)

The results for the Pc5 and Pc4 frequency bands show good agreement between interpolated values and original data at BOU. This is true for both interpolation methods, although slightly better for KRG than for BHS. For the Pc3 band, for both interpolation methods, there are obvious differences between the interpolated results and the original data. One possible explanation is that the Pc3 band is higher in frequency than Pc4/5, which means shorter scale length for the interpolation method to respond to the rapid changes that happened at the source points. As shown in the previous statistical correlation study between BH variations with different frequencies (fig. 4), the correlation coefficients are higher for the Pc4/Pc5 bands (2–22 mHz) than the Pc3 band (22–100mHz). The correlation study also supports the idea that the Kriging method is more reasonable. In the Kriging method, the data source with the closer distance has higher weight in the interpolation, and the correlation study shows that the data sources of BH from locations with closer distance have higher correlation coefficients to each other.

A detailed test procedure was run for the Kriging method for the relative error between the data from BOU and the interpolated time series at that location for Pc3, Pc4, and Pc5 during the storm period from 2011–297 to 2011–300. The method is described in appendix 4 and results are shown as table 4. The difference between the interpolated and original data in this storm-time example is small; this level of difference is consistent with the previous tests shown in figures 6–8.

Table 4. Relative errors between interpolation with/without data from BOU for Pc3, Pc4, and Pc5 on 2011-298. (BOU, Boulder; KRG, Kriging; UT, universal time)

Pc3		Pc4		Pc5	
KRG					
Period(UT)	Relative Error	Period(UT)	Relative Error	Period(UT)	Relative Error
01–03	19%	01–03	26%	01–03	20%
03–06	21%	03–06	13%	03–06	22%
06–09	13%	06–09	15%	06–09	27%
09–12	24%	09–12	8%	09–12	9%
12–15	31%	12–15	6%	12–15	9%
15–18	33%	15–18	15%	15–18	5%
18–21	37%	18–21	19%	18–21	6%
21–24	30%	21–24	17%	21–24	9%
BHS					
Period(UT)	Relative Error	Period(UT)	Relative Error	Period(UT)	Relative Error
01–03	34%	01–03	52%	01–03	40%
03–06	30%	03–06	22%	03–06	35%
06–09	20%	06–09	26%	06–09	49%
09–12	26%	09–12	13%	09–12	20%
12–15	38%	12–15	10%	12–15	13%
15–18	40%	15–18	26%	15–18	7%
18–21	37%	18–21	25%	18–21	11%
21–24	25%	21–24	17%	21–24	21%

The pulsation maps can be produced as a near real-time map because of the strengths of the wavelet procedure. But we need to note that, due to the limitation of the spatial resolution of the source data, the interpretation of these maps is more useful as a space-weather-monitoring tool than a scientific modeling study. The confidence coefficients can be produced for the map, which have 100 percent confidence at the observatory source locations and decrease when the location moves away from observatories. The future improvements on the map will be done more effectively by increasing the spatial resolution of data points rather than by manipulating the interpolation methods.

Summary

Geomagnetic pulsation indices are created by applying wavelet analysis to 1-sec geomagnetic data provided by the USGS Geomagnetism Program. The indices cover the frequency band of Pc3, Pc4, and Pc5 pulsations and can be applied as a near real-time online display for space weather monitoring tools. The statistical-correlation study for geomagnetic variations with different frequency at different locations is performed for horizontal (BH) magnetic field components. The map of geomagnetic pulsations separated by frequency band is created as an initial step in monitoring the spatial pattern changes of geomagnetic variations over the continental United States. This map may prove useful as a space weather monitoring tool. Further improvements can be made through increasing spatial resolution of data.

References Cited

Anderson, B.J., 1993, Statistical studies of Pc 3-5 pulsations and their relevance for possible source mechanisms of ULF waves: Annales Geophysicae, v. 11, p. 128–143.

Anderson, B.J., 1994, An overview of spacecraft observations of 10 s to 600 s period magnetic pulsations in the Earth's magnetosphere, *in* Solar wind sources of magnetospheric ultra-low-frequency waves, eds. Engebretson, M.J., Takahashi, K., and Scholer, M.: American Geophysical Union Geophysical Monograph 81, p. 25–43.

Fritsch, F.N., and Carlson, R.E., 1980, Monotone piecewise cubic interpolation: SIAM Journal on Numerical Analysis, v. 17, p. 238–246.

Jacobs, J.A., Kato, Y., Matsushita, S., and Troitskaya, V.A., 1964, Classification of geomagnetic micropulsations: Journal of Geophysical Research, v. 69, p. 180–181.

Matheron, Georges, 1963, Principles of geostatistics: Economic Geology, v. 58, p. 1246–1266.

Percival, D.B., and Walden A.T., 2000, Wavelet methods for time series analysis: Cambridge University Press, Cambridge, 594 p.

Sandwell, D.T., 1987, Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data: Geophysical Research Letters, v. 2, p. 139–142.

Stewart, B., 1861, On the great magnetic disturbance which extended from August 28 to September 7, 1859, as recorded by photography at the Kew Observatory: Philosophical Transactions of the Royal Society London, p. 423–430.

Torrence, Christopher, and Compo, G.P., 1998, A practical guide to wavelet analysis: Bulletin of the American Meteorological Society, v. 79, p. 61–78.

Wanliss, J.A., and Antoine, L.A.G., 1995, Geomagnetic micropulsations—Implications for high resolution aeromagnetic surveys: Exploration Geophysics, v. 26, p. 535–538.

Xu, Z., Zhu, L., Sojka, J.J., Kokoszka, P., and Jach, A., 2008, An assessment study of the wavelet-based index of magnetic storm activity (WISA) and its comparison to the Dst index: Journal of Atmospheric and Solar-Terrestrial Physics, v. 70, no. 11-12, p. 1579–1588.

Appendixes

I. Data file example

1) BOU2008068.sec (quiet period in year 2008)

```
BOU 2008 068 08-MAR-08 HEZF 0.01nT File Version 2.00
00000 2089226  -4943 4831140 5318027
00001 2089224  -4940 4831139 5318027
00002 2089224  -4940 4831138 5318029
00003 2089224  -4940 4831139 5318027
00004 2089225  -4941 4831138 5318026
00005 2089227  -4941 4831138 5318029
... ... ... ... ... ... ... ... ... ... ... ... ...
86396 2086898  -4228 4831404 5317336
86397 2086897  -4229 4831405 5317336
86398 2086895  -4228 4831404 5317335
86399 2086895  -4228 4831405 5317335
```

2) bou20110914vsec.sec (storm period in year 2011)

```
Format         IAGA-2002                 |
Source of Data     United States Geological Survey (USGS)    |
Station Name      Boulder                  |
IAGA CODE       BOU                     |
Geodetic Latitude   40.137                    |
Geodetic Longitude   254.764                   |
Elevation       1682                    |
Reported        HDZF                    |
Sensor Orientation   HDZF                     |
Digital Sampling    100 Hz                    |
Data Interval Type   Average 1-Second               |
Data Type       variation                |
# DECBAS        7406  (Baseline declination value in    |
#           tenths of minutes East (0-216,000)).     |
# This data file was constructed by Golden GIN.           |
# Final data will be available on the INTERMAGNET DVD.       |
# Go to www.intermagnet.org for details on obtaining this product. |
# CONDITIONS OF USE: The Conditions of Use for data provided    |
# through INTERMAGNET and acknowledgement templates can be found  |
# at www.intermagnet.org                    |
DATE    TIME     DOY   BOUH   BOUD   BOUZ   BOUF  |
2011-09-14 00:00:00.000 257   20852.02  -33.36 47910.49 52786.01
2011-09-14 00:00:01.000 257   20851.94  -33.36 47910.52 52786.00
2011-09-14 00:00:02.000 257   20851.86  -33.36 47910.56 52786.02
... ... ... ... ... ... ... ... ... ... ... ... ...
2011-09-14 23:59:57.000 257   20853.81  -33.56 47909.18 52785.92
2011-09-14 23:59:58.000 257   20853.83  -33.56 47909.18 52785.92
2011-09-14 23:59:59.000 257   20853.86  -33.56 47909.16 52785.92
```

II. R codes for wavelet analysis

1) Main R code

```
##############################################                    Comments
#load data
file.name<-"OBSYYYYDDDDDDBH.txt"                                  Cleaned geomagnetic data
data<-read.table(file.name,comment.char=comment.char)             Load data into R console
Bh<-as.vector(t(data))                                            Bh is geomagnetic data
##############################################
#wavelet initial
wf<-"la8"                                                         Wavelet function
boundary<-"reflection"                                            Boundary condition
N<-length(Bh)                                                     Number of data points
n.levels<-floor(log(N,2))                                         Highest level of MRAs
# apply wavelet transform
Bh.wt <- modwt(Bh, wf=wf,n.levels=n.levels,boundary=boundary)     MODWT function
# obtain MRA
Bh.wt.mra<-mra.wt(Bh.wt)                                          Data vector for MRA
##############################################
#output
write.table(Bh.wt.mra, output.file, sep="\t")                     Output MRA into txt file
##############################################
```

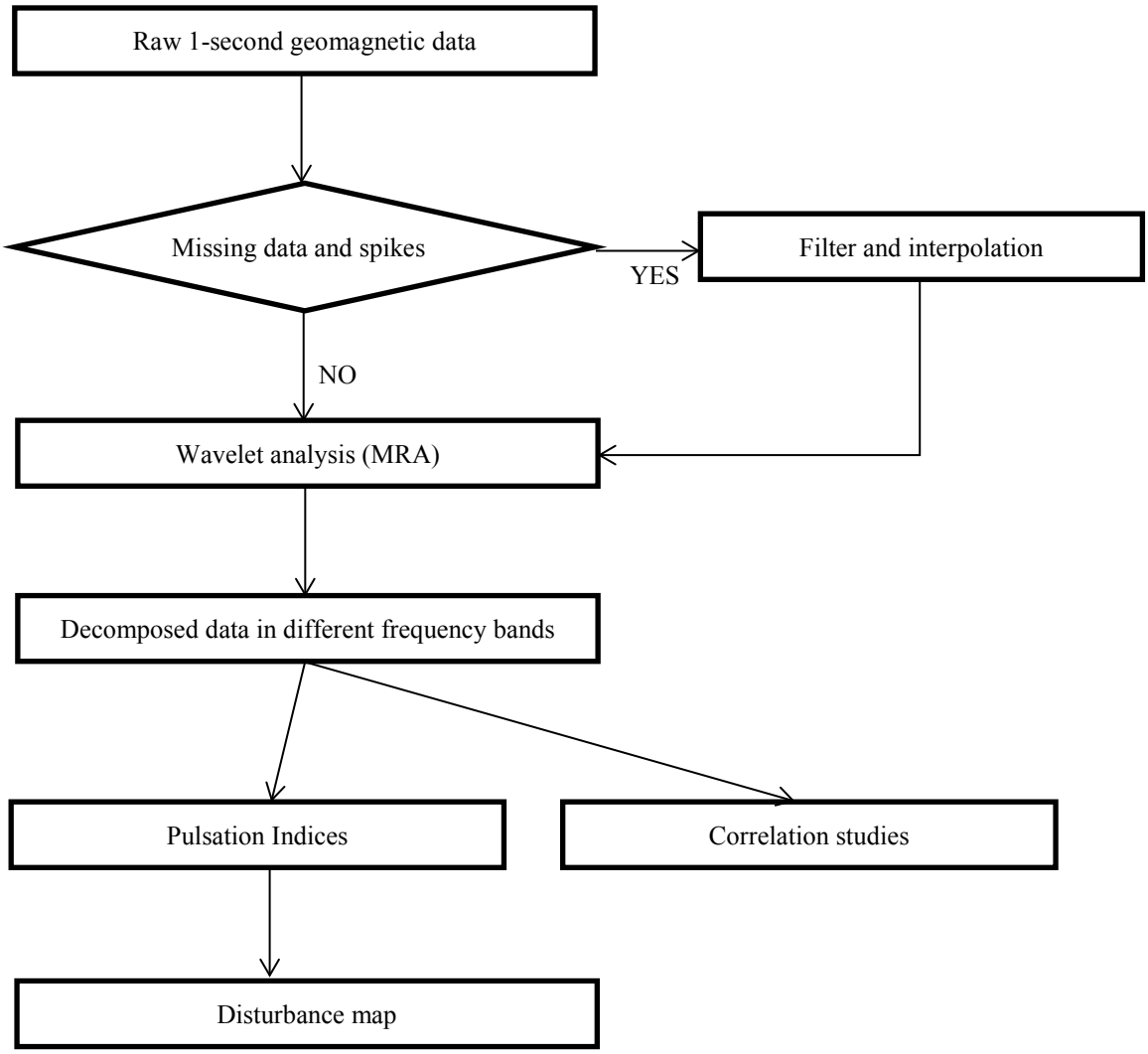
2) Subroutine for MODWT

```
##############################################                    Comments
function (x, wf = "la8", n.levels = 4, boundary = "periodic")     Function input
{
  switch(boundary, reflection = x <- c(x, rev(x)), periodic = invisible(),
    stop("Invalid boundary rule in modwt"))                       Check the validation of
                                                                  boundary conditions
  N <- length(x)                                                  Number of data points
  storage.mode(N) <- "integer"
  J <- n.levels                                                   Levels of wavelet transform
  if (2^J > N)
    stop("wavelet transform exceeds sample size in modwt")        Check data length enough or not
  dict <- wave.filter(wf)                                         Select wavelet filter
  L <- dict$length                                                Length of data
  storage.mode(L) <- "integer"                                    Data type
  ht <- dict$hpf/sqrt(2)                                          High-pass filter coef.
  storage.mode(ht) <- "double"                                    Data type
  gt <- dict$lpf/sqrt(2)                                          Low-pass filter coef.
  storage.mode(gt) <- "double"                                    Data type
  y <- vector("list", J + 1)                                      Initial output data vector
  names(y) <- c(paste("d", 1:J, sep = ""), paste("s", J, sep = ""))
                                                                  Initial output data name
  W <- V <- numeric(N)                                            Initial coefficient vector
  storage.mode(W) <- "double"                                     Data type
  storage.mode(V) <- "double"                                     Data type
  for (j in 1:J) {                                                Loop of each level
    out <- .C("modwt", as.double(x), N, as.integer(j), L,
      ht, gt, W = W, V = V, PACKAGE = "waveslim")[7:8]            Wavelet transform
    y[[j]] <- out$W                                               Wavelet coefficients
    x <- out$V                                                    Scaling coefficients
  }
  y[[J + 1]] <- x                                                 Scaling coefficients
  class(y) <- "modwt"                                             Set output wavelet class
  attr(y, "wavelet") <- wf                                        Set output wavelet function
  attr(y, "boundary") <- boundary                                 Set output boundary condition
  return(y)                                                       Output result vector
}
##############################################
```

3) Subroutine for Multi Resolution Analysis (MRA)

```
##########################################                    Comments
mra.wt<-function(x.wt)                                       Function input
{
  wf<-attr(x.wt,"wavelet")                                   Wavelet function
  J<-length(x.wt)-1                                          Wavelet levels
  method<-attr(x.wt,"class")                                       Wavelet method
  boundary<-attr(x.wt,"boundary")                            Boundary condition
  x.mra <- vector("list", J + 1)                             Initial MRA vector
  zero <- vector("list", J + 1)                                    Initial zero vector
  names(zero) <- c(paste("d", 1:J, sep = ""), paste("s", J,  Initial zero vector names
    sep = ""))
  class(zero) <- method                                            Initial zero vector class
  attr(zero, "wavelet") <- wf                                      Initial wavelet function
  attr(zero, "boundary") <- boundary                               Initial boundary condition
  zero[[J + 1]] <- x.wt[[J + 1]]                             Initial value
  if (method == "modwt") {                                   For MODWT
    for (k in 1:J) zero[[k]] <- numeric(n)                         no changes to data length
    x.mra[[J + 1]] <- imodwt(zero)                                 Inverse MODWT for smooth
  }
  else {                                                           For other DWTs
    for (k in 1:J) zero[[k]] <- numeric(n/2^k)               Data length changed to 2k
    x.mra[[J + 1]] <- idwt(zero)                             Inverse DWT for smooth
  }
  for (j in J:1) {                                           Loop for J to 1 levels MRA
    zero <- vector("list", j + 1)                            Initial zero vector
    names(zero) <- c(paste("d", 1:j, sep = ""), paste("s",
      j, sep = ""))                                          Initial zero vector names
    class(zero) <- method                                    Initial zero vector class
    attr(zero, "wavelet") <- wf                              Initial wavelet function
    attr(zero, "boundary") <- boundary                       Initial boundary condition
    zero[[j]] <- x.wt[[j]]                                   Initial wavelet coefficients
    if (method == "modwt") {                                       For MODWT
      if (j != 1) {                                          No changes to coefficient length
        for (k in c(j + 1, (j - 1):1)) zero[[k]] <- numeric(n)
      }
      else {
        zero[[j + 1]] <- numeric(n)                          Detail 0 is “0”
      }
      x.mra[[j]] <- imodwt(zero)                             Inverse MODWT for details
    }
    else {                                                   For other DWT
      zero[[j + 1]] <- numeric(n/2^j)                        Detail 0 is “0”
      if (j != 1) {
        for (k in (j - 1):1) zero[[k]] <- numeric(n/2^k)     Change coefficients length
      }
      x.mra[[j]] <- idwt(zero)                               Inverse DWT for details
    }
  }
  names(x.mra) <- c(paste("D", 1:J, sep = ""), paste("S", J, Set output names
    sep = ""))
  if (boundary == "reflection") {                            Set data length
    for (j in (J + 1):1) x.mra[[j]] <- x.mra[[j]][1:(n/2)]   Boundary conditions
    return(x.mra)                                            Output MRA
  }
  else {
    return(x.mra)                                            Output MRA
  }
}
##########################################
```

III. Schematic flow chart



IV. Average Relative Error Calculation

```
error_threshold = mean(original_value)                    Choose error threshold
number_above_threshold = 0;                               Initialize counter
rel_error_sum = 0                                         Initialize sum of relative
                                                          errors


for i=1:length(original_value)                            Loop over original data
{
  if abs(original_value(i)) > error_threshold             If value exceeds threshold
      {

    rel_error_sum = rel_error_sum + abs( (original_value(i)-interpolated_value(i)) /
original_value(i) );
```

```
        number_above_threshold = number_above_threshold +1;
    }
}

rel_error_mean = rel_error_sum / number_above_threshold;
```

Relative error is calculated RE = abs (original - interpolated) / original

Calculate mean