PRACTICE PAPER

Data Tracking Analysis of the Geomagnetic Fixed-Station Network in China

Suqin Zhang[1], Changhua Fu[1], Xiuxia Zhang[2] and Yi Zhang[3]

[1] Institute of Geophysics, China Earthquake Administration, Beijing, 100086, CN

[2] Earthquake Administration of Jiangsu Province, Nanjing, 210014, CN

[3] China Earthquake Networks Center, Beijing, 100045, CN

Corresponding author: Changhua Fu (13521519246@139.com)

Data tracking analysis is an important mechanism for increasing data analysis capacity and eliminating interference from observational data. In this study, the technique was applied to the geomagnetic fixed-station network to improve the efficiency and accuracy of analysis to extract useful information. This paper introduces the scope, workflow, analysis platform, abnormal variation status, and results of the geomagnetic data tracking analysis. We present some typical examples of abnormal variations in addition to our proposals for future work.

Keywords: data tracking analysis; abnormal variation; event records

Introduction

Geomagnetism is an observation-based scientific discipline, where fixed geomagnetic observations are the main source of geomagnetic data (Xu 2003). As investments are made in nationwide construction and development, the scale of the geomagnetic fixed-station network is continuously expanding in China. The variety and time span of observations are also increasing, which has provided abundant and valuable first-hand data for scientific research, while at the same time creating greater demand for data analysis and processing capacity. In addition, rapid urbanization and construction development has negatively affected the geomagnetic observation environment (Xie 2011). Instrument failure and the effects of other interfering factors (i.e., noise) pose severe challenges to the quality of observation data (Yao 2015). These data changes caused by instrument failure, environmental interference, and human interference are defined as abnormal changes. Therefore, identifying and excluding the effects of instrument failure, environmental interference, and human interference from massive data, while increasing the quality of observation data, efficiently extracting useful data, and providing better data sharing services for scientific research have become important work in the Geomagnetic Network of China (GNC) (Rasson, Toh & Yang 2010). To meet these demands, the GNC implemented data tracking analysis in 2014. This paper focuses on the scope and process of data tracking analysis, the data tracking analysis platform, and the abnormal variation patterns of the data obtained from the observatories.

Data Tracking Analysis Scope and Workflow

A tracking analysis event library was developed using distributed technology system architecture composed of four levels of information nodes: stations, local earthquake administrations, the China Earthquake Networks Center, and the GNC (Zhang et al. 2016), which enabled a tracking analysis software system to be deployed. The seismic information network was used to implement the exchange and collection of observation event records from the entire network and initially carry out the observation data tracking analysis from the national geomagnetic network (GNC 2014).

The specific tracking procedure is as follows. Staffs at the observation stations are responsible for data mining, identifying abnormal records and creating observion event records. Data mining, which is the foundation of the tracking analysis platform, includes the extraction of anomalous longitudinal

spectral amplitudes (Yao 2015), wavelet analysis, Fourier correlation analysis (Wan 2007) and extreme value identification (Zheng et al. 2012). The recognition rate of abnormal data is up to 95.2%. For observations with abnormal dynamic changes, the observation diary and working diary are combined to examine the performance and work status of the observation instrument. The observational processes of the staff are investigated and any changes in the observation environment are evaluated in order to identify the reasons for the abnormal changes to the data. The results of this analysis are stored in the station's observation event record database. The local earthquake administrations are responsible for the initial examination for every analysis event record created by the observatory staff. The observatory staff can see the audit results on the tracking analysis platform as soon as they are saved to the database. If the audit results are deemed correct, the observatory staffs reanalyze their data. At the same time, comprehensive analysis of the regional network is carried out, with monthly and annual bulletins being generated. The GNC is responsible for rechecking the quality of the tracking analysis events, providing an analysis of the national network, and generating monthly and annual tracking analysis bulletins. The CENC is responsible for coordinating and promoting this work, developing and improving the analysis software platform, and ultimately providing external services. The entire data tracking analysis procedure is shown in **Figure 1**.

With the goal of comprehensively monitoring data with different resolutions (e.g., minute data, hourly average, and daily averages), tracking analysis is divided into weekly, monthly, and annual analyses. Weekly analysis is conducted based on original and preprocessed minute data with a focus on analyzing event types that are unresolved using preprocessing, such as subway interference and inter-day steps. Monthly analysis is conducted based on hourly average data, and annual analysis is based on daily average data. The focus of the monthly and annual analysis is to analyze long-term trends in abnormal variation events that are imperceptible in weekly analysis, such as data drift due to unstable instrument piers and long-term trend abnormalities due to gradual changes in the observation environment.
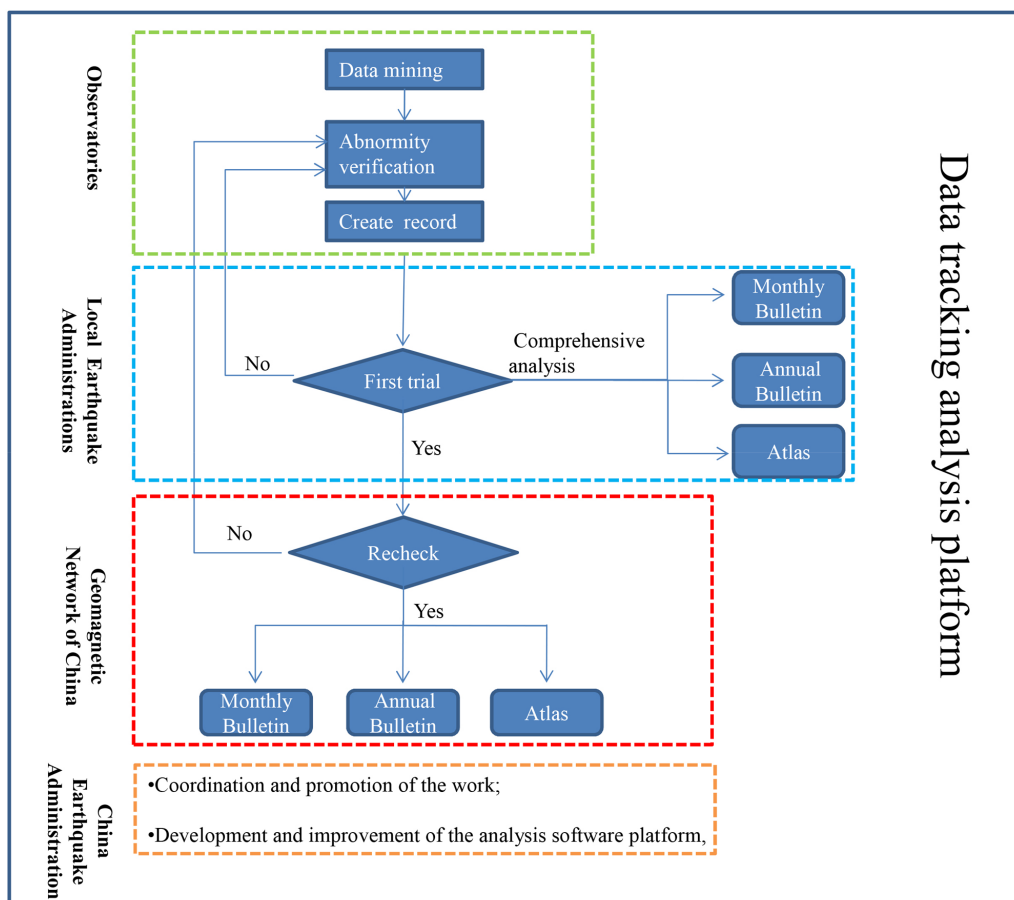
Figure 1: Flow diagram of data tracking analysis at GNC.

Data Tracking Analysis Platform

Geomagnetic data tracking analysis is implemented and completed based on the tracking analysis platform. The data tracking analysis platform was jointly developed by the CENC, GNC, local earthquake administrations, and station staff (Li et al. 2016). The platform has a series of functions, such as data mining, reconstructing observation events, plotting, reviewing, correcting, and archiving results. The station staffs scan the observation data weekly, monthly, or annually, according to the requirements, and search for observation events with abnormal variation patterns. Using the platform's plotting function, different data types can be output as illustrations, and text, icons, seismic events, and other necessary information can be marked on the figure, thus forming an event entry. The platform also supports the embedding of images created from other software. The analysis results are automatically stored in the tracking analysis event library, and the data management system automatically transfers them to an upper node in the network. Based on the above process, the Local Earthquake Administration (LEA) and GNC complete a series of work including event record audit, feedback, comprehensive analysis, and report output.

Each tracking analysis record includes the type of abnormal variation pattern (as described in the next section), influencing factors, the start and end time of the event, a description of the variation pattern, the extent of the variation pattern, the name of the event inspection staff, the diary completion status, the type of data used, abnormal images, and any other pertinent information. For abnormal events with unknown causes, the station is required to conduct an on-site verification and provide the verification report for the abnormality.

Abnormal Variation Patterns

The causes of abnormal dynamic variations of observation data are classified using a comprehensive analysis and summary of evolution trends and characteristics of the geomagnetic network observation data combined with the physical attributes of the events. Categories include the observation system, natural environment, site environment, human interferences, and unknown causes. There are many factors for the various types of abnormal dynamic variations, and **Table 1** gives a variety of common factors that cause abnormal changes. Notably, changes due to unknown reasons may have explanations that are currently unknown because of limitations in our current knowledge. For example, some variations may be suspected earthquake precursor anomalies, which require tracking and analysis to provide a reference for subsequent understanding.

The main purpose of data preprocessing is to remove artificial disturbances contained in the daily variations. The observatory staffs monitor the daily variations every morning and manually remove artificial disturbances from the records. It can be seen from the following examples that, unlike data preprocessing, the abnormal information is extracted automatically from the mass observation data by visualization data mining methods; the characteristics of the typical abnormal signals are subsequently summarized and analyzed systematically. The automated system improves the capacity to recognize abnormal data and the efficiency with which they are analyzed, which provides great convenience for data users to remove noise from bulk data. Due to limited space, this study only provides some typical and representative abnormal variation event patterns.

**Table 1:** Examples of factors influencing various types of abnormal variations.

| Influencing factors	Examples
Observation system malfunction	Sensor malfunction, host computer malfunction, data requisition malfunction, communication unit malfunction, lightning avoidance system malfunction, battery malfunction, insufficient power source, AC/DC switching interference, simulation devices malfunction.
Natural environment	Temperature, humidity.
Site environment	Infrastructure construction, vehicles, subway and light rail, HVDC, earth resistivity, power supply interference, factory operations.
Human interference	Calibration, maintenance equipment, agriculture.
Unknown reasons	The cause of changes cannot be confirmed after exclusion of effects due to observation system malfunctions, the natural environment, the site environment, or human interference.

Subway and light rail

The fluxgate magnetometers at Sheshan station in Shanghai are affected by the operation of subway routes 1, 2, and 9 every day at 04:25–23:59 Beijing time, with interference from the subway causing vertical component (Z), horizontal component (H), and declination (D) variations in the range of 5.6 nT, 0.72 nT, and 0.26 minute, respectively. An example interference pattern is shown in **Figure 2**.

HVDC transmission interference

From 03:12 UTC on July 13 to 10:03 UTC on July 14, 2016, the GM4 fluxgate magnetometers at Wuhan station were affected by interference from HVDC circuit debugging. The interference levels for H, Z, and D were 0.0 nT, 8.5 nT, and 0.0 nT, respectively. **Figure 3** compares the curves of raw data and pre-processed data before and after processing and highlights the described interference levels.
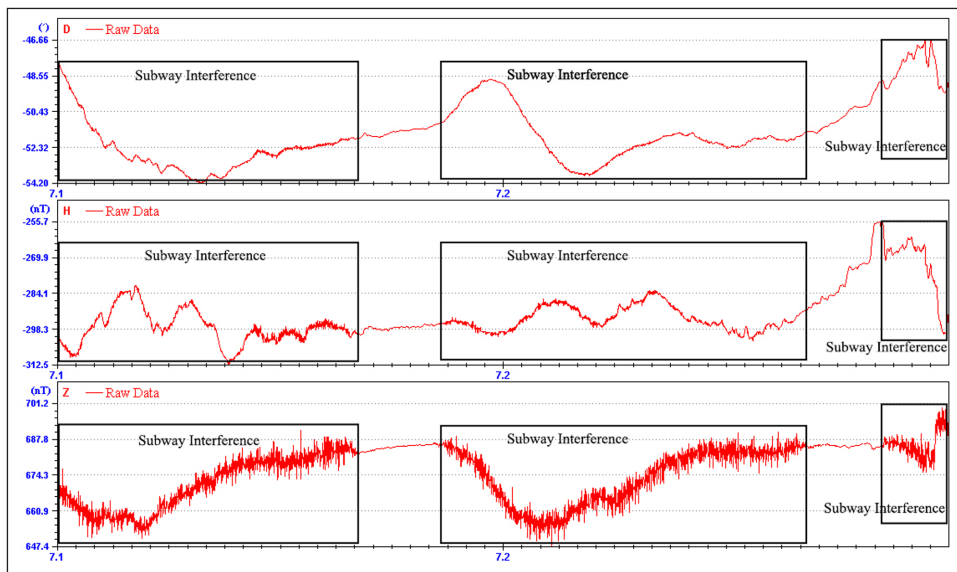
Figure 2: An example of the pattern of subway interference at Sheshan station.
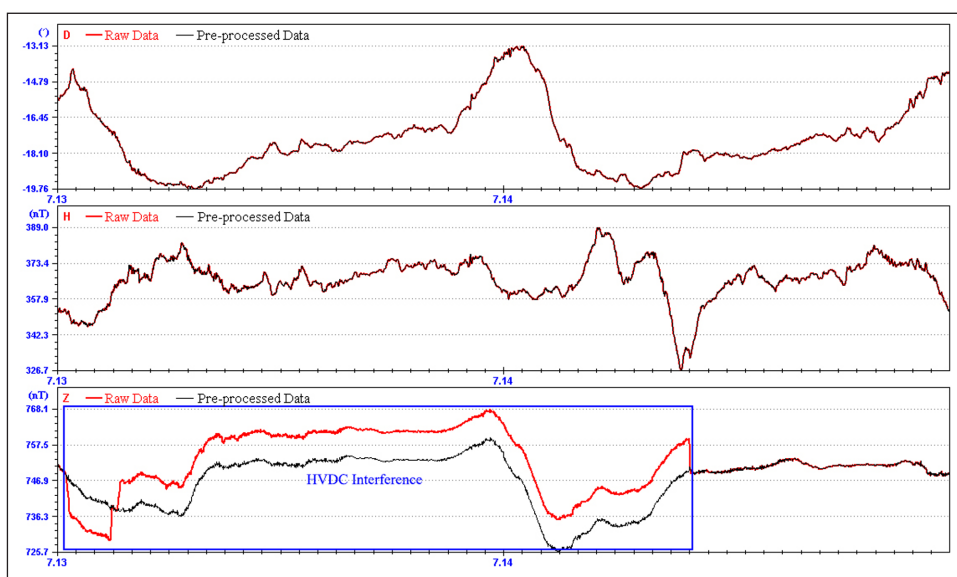
Figure 3: An example of an interference pattern caused by HVDC circuit debugging at Wuhan station.

Natural environment effects

The hourly average data curves from Wanzhou station and other surrounding stations were compared, see **Figure 4**, which indicates that the D and Z component data from Wanzhou station had significant drift. After comparing the data curves with the temperature data from the recording station, the temperature variations at the Wanzhou station were found to be greater than at other stations. The greater temperature variability was due to poor insulation of the recording room where the probe was installed, resulting in data drift.

Data requisition malfunction

The proton vector magnetometer FHD-2B at Wushi station had a data requisition malfunction on March 9, 2016. After reactivating the data requisition, errors appeared in the instrument parameters, resulting in steps in the H component data from 05:33 UTC on March 9, 2016 to 06:40 UTC on March 10, 2016. After
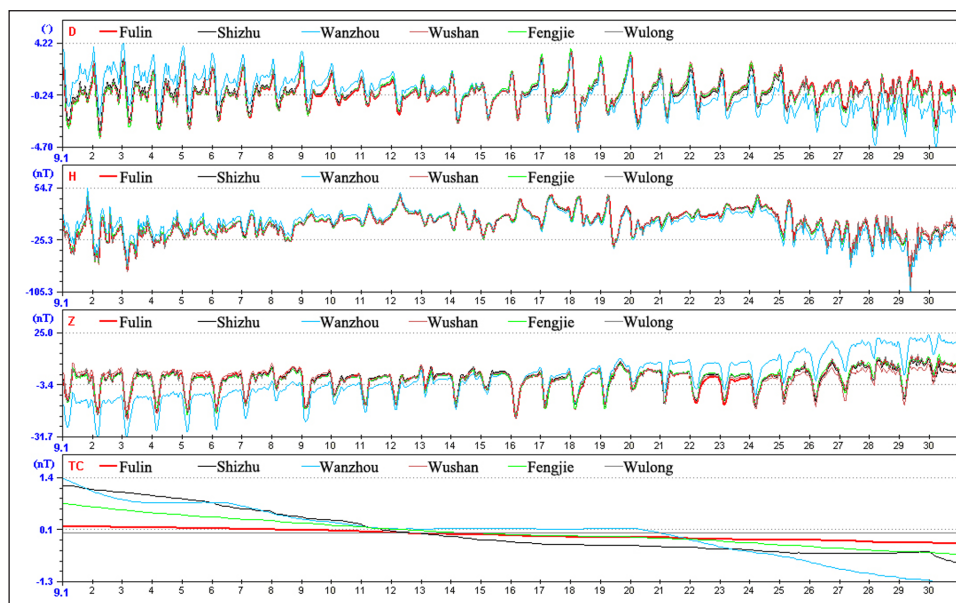
Figure 4: An example of data drift caused by poor insulation of the recording room at Wanzhou station.
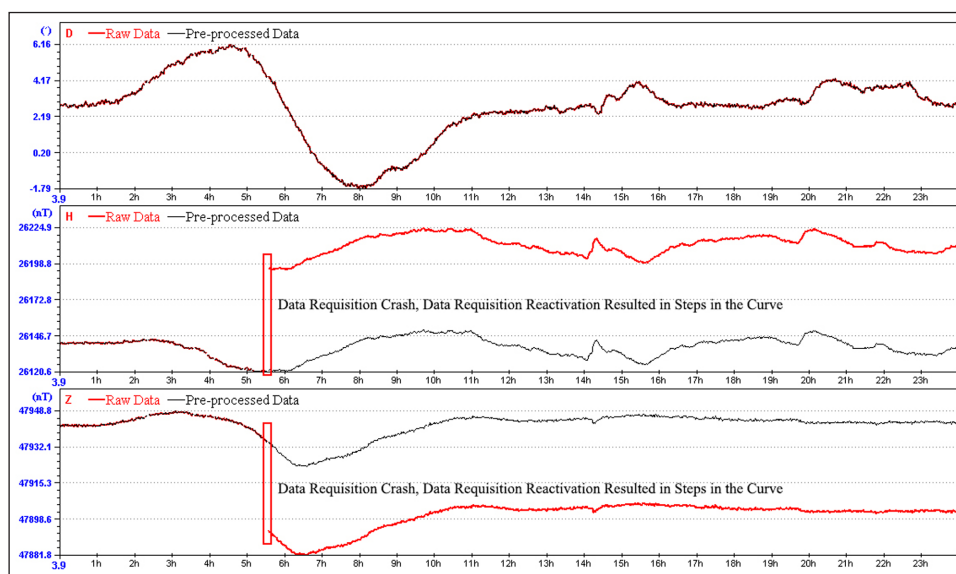
Figure 5: An example of interference patterns due to a data requisition crash at Wushi station.
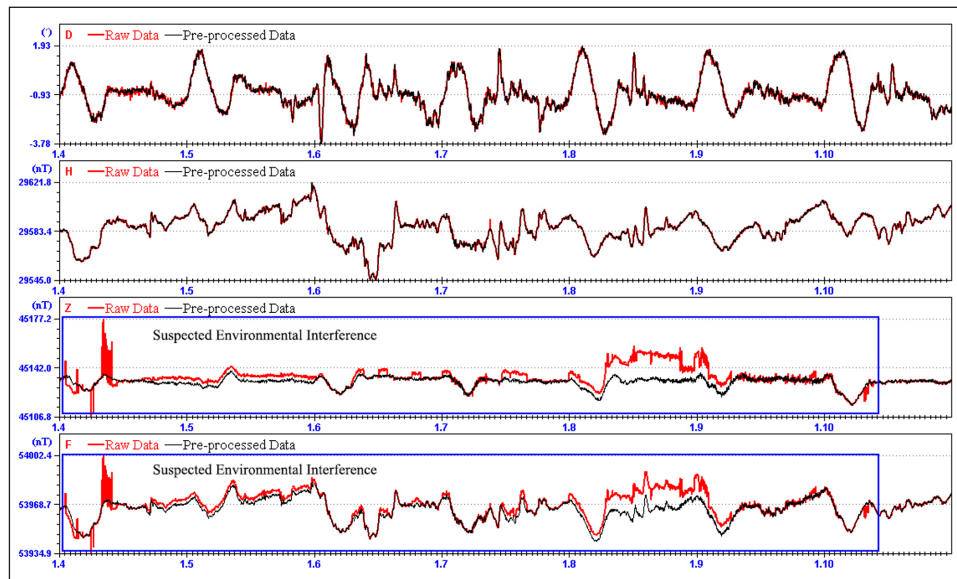
Figure 6: An example of suspected environmental interference at Taiyuan station.

resetting the parameters, the data returned to normal. **Figure 5** shows a comparison of images before and after data pre-processing. The step correction was –73.5 nT.

Suspected environmental interference

The Z-component of the data curve from the GM4 fluxgate magnetometers at Taiyuan station showed abnormal variation patterns from January 4–10, 2016, **Figure 6**. After examination, three instruments in the station were found to have synchronization changes. This indicated that these changes were not due to instrument malfunctions, and instead were suspected to be due to environmental interference.

Results

Since the launch of the data tracking analysis in 2014, all nodes output event records, monthly bulletins, annual bulletins, and an observation event atlas. In 2014, 2015 and 2016, 1192, 1390 and 1339 event records were archived, respectively. A total of 1116 monthly bulletins and 93 annual bulletins were compiled by LEA and GNC. The typical event information was extracted, and an atlas of typical observation events in 2015 and 2016 was compiled.

Conclusions and Future Work

The data tracking analysis provides a highly efficient and convenient service for data users to extract useful data from massive datasets. However, current tracking analysis work partly relies on human intervention and is not sufficiently automated. Future work should focus on strengthening data interpretation and identification capabilities, utilizing intelligent identification technology to establish data mining and review models, and decreasing the time required to evaluate and share data.

Acknowledgements

Funding Information

Competing Interests

The authors have no competing interests to declare.

References

Geomagnetic Network of China. 2014. Technical guide for geomagnetic observation data tracking analysis (Trial Edition). Unpublished.

Li, ZY, Xiong, DH, Liu, GC, Liang, YQ and **Ye, Q.** 2016. The application of large data mining techniques in Earthquake Precursory Network Observation Data Tracking Analysis. *Seismological and Geomagnetic Observation and Research*, 37(3): 1–6.

Rasson, JL, Toh, H and **Yang, DM.** 2010. The global geomagnetic observatory. In: Mandea, M and Korte, M (eds.), *Geomagnetic Observations and Models*, 1–27. Volume 5 of the series IAGA Special Sopron Book Series: Springer. DOI: https://doi.org/10.1007/978-90-481-9858-0

Wan, YG. 2007. Implementation of digital signal processing based on Matlab. Beijing: Science Press.

Xie, F. 2011. A study on reducing man-made electromagnetic noise in geomagnetic field observation. Unpublished thesis (PhD), Institute of Geophysics, China Earthquake Administration.

Xu, WY. 2003. Geomagnetism. Beijing: Earthquake Press.

Yao, XY. 2015. Method of the artificial electromagnetic disturbances identification and data reconstruction. Unpublished thesis (PhD), Institute of Geophysics, China Earthquake Administration.

Zhang, SQ, Fu, CH, He, YF, Yang, DM, Li, Q, Zhao, XD and **Wang, JJ.** 2016. Quality control of observation data by the Geomagnetic Network of China. *Data Science Journal*, 15(15): 1–12. DOI: https://doi.org/10.5334/dsj-2016-015

Zheng, JR, Fan, X, Zhang, XX and **Bao, HY.** 2012. Application of interference identification technology in geomagnetic observation based on extreme characteristic scale. *Progress in geophysics*, 27(2): 528–536