https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

Open Access Earth System Science Data Discussions

Rescue and quality control of historical geomagnetic measurement at SheShan Observatory, China

Suqin Zhang[1], Changhua Fu[1], Jianjun Wang[2], Guohao Zhu[3], Chuanhua Chen[4]

[1]Institute of Geophysics, China Earthquake Administration, Beijing, 100081, China
[2]Earthquake Administration of Gansu Province, Lanzhou, 730000, China
[3]Shanghai Earthquake Agency, Shanghai, 200062, China
[4]Earthquake Administration of Shangdong Province, Jinan, 250014, China

Correspondence to: *Changhua Fu (*angie@email.com)

Abstract. The Sheshan geomagnetic observatory (IAGA code SSH) , China was built in Xujiahui, Shanghai in 1874 and moved to Sheshan, Shanghai at the end of 1932. So far, SSH has a history of nearly 150 years. It is one of the earliest geomagnetic observatory in China and one of the geomagnetic stations with the longest history in the world.In this paper, we present the rescue and quality control of the historical data at Sheshan observatory (SSH) from 1933 to 2019. The rescued data are the absolute hourly mean values (AHMVs) of D, H and Z components. Some of these data are paper-based records, and some are stored in a floppy disk in BAS, DBF, ACCESS and other file formats. After digitization and format transformation, we imported the data into the Toad database to achieve the unified data management. We performed statistics of continuity rate, visual analysis, outliers detects and data correction on the stored data. Then we conducted the consistency test of daily variation and secular variation（SV）by comparing the corrected data with these of the reference observatory data, and the computational data of the COV-OBS model, respectively. The consistency test revealed fairly good agreement. However, the individual data should be used with caution, because these data are suspicious values, but there is not any explanation or change registered in the available metadata and logbooks. Finally we present an example of the datasets in discriminating geomagnetic jerks. The digitized and quality-controlled AHMVs data are available at: https://doi.org/10.5281/zenodo.6584285 (zhang et al, 2022).

1 Introduction

Geomagnetic observation data contains abundant solar-terrestrial spatial information, which is widely used in geoscience and space science research. The observation data with time resolution of one second to one hour are usually used to study various short-period magnetic fields such as pulsation, geomagnetic crochet, geomagnetic bay and magnetic storm (Zhao et al., 2019), and to monitor and predict the electromagnetic environment in solar-terrestrial space. At the same time, it also has important applications in detecting underground electrical structures and evaluating the impact of geomagnetic induced current (GIC) on underground metal pipe network, transmission network, communication cables, high-speed railway lines and other major projects (Kappenman, 1996; Bolduc et al., 1998, 2002; Boteler et al., 1998; Liu et al., 2008, 2016; Liu et al.,

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

2009; Guo et al., 2015). Observation data with time resolution of 1h to hundreds of years are usually used for the study of geomagnetic field and its secular variation, such as geomagnetic jerk (Courtillot and Mouël, 1984; Xu, 2009), magnetic pole movement, dipole magnetic moment change, westward drift, etc., which are of great significance for understanding the material flow inside the core and at the core mantle boundary.

The development and application of geomagnetism depends on long-term data accumulation. The long-term operation of geomagnetic observation stations is very important for the study of the geomagnetic field (Linthe et al., 2013). It is especially valuable to study the variation characteristics of the geomagnetic field from decades to hundreds of years (Clarke, 2009; Zhang et al., 2008). Using the latest scientific and technological means to analyze the geomagnetic continuous observation data as long as possible, so as to obtain the change information of geomagnetic field, has always been a method often used by scientific researchers. However, not all data can be directly provided to researchers, because some data still exist only in the form of hard copy, and even some data face the risk of serious damage and loss due to improper storage conditions. Therefore, it is very important to rescue and digitize these data as soon as possible. High quality data are the basis of scientific research and the prerequisite for obtaining valuable results (Linthe et al., 2013). Scientists around the world pay more and more attention to the accumulation of observation data, the rescue of historical data and the sharing of scientific data resources (Curto and Marsal, 2007; Peng et al., 2007; Korte et al., 2009; Dawson et al., 2009; Morozova et al., 2014, 2020; Sergeyeva et al., 2020; Dong et al., 2009; Zhao et al., 2017; Thomson, 2020).

The rescue, recovery, digitization and the quality control of historical geomagnetic data are of extraordinary importance for the geomagnetic community (Rasson, et al., 2011). This paper presents the collection, collation, digitization, the quality control and the correction of the historical data of Sheshan Geomagnetic Observatory (International Association of Geomagnetism and Aeronomy code SSH) from 1933 to 2019. SSH Geomagnetic Observatory is the geomagnetic station with the longest history in China. Although many efforts have been made (Gao et al.,1993;), the existing data are still insufficient. Our work aims at filling the lack of observation data at SSH station since 1933, presenting the absolute hourly mean values (AHMVs) data collected from 1933 to 2019.

This paper is organized as follows. Section 2 describes the data acquisition method, providing information about SSH observatory history, data sources and digitizing methodology. Section 3 describes the quality control of the digitized data. Section 4 describes the correction of the selected homogeneity. Section 5 describes the examination of the corrected series by comparing with reference series and section 6 presents application examples of the datasets. Concluding remarks are made in section 7.

2 Data Production Methods

2.1 The Sheshan Geomagnetic Observatory

The first step of the data rescue process is to collect resources scattered in different locations, which exist in various forms, including data and metadata that may have an impact on data rescue (observatory location, instrument changes, etc.). We

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

conducted a careful examination of the bibliographic documents stored in the SSH, Geomagnetic Network of China (GNC) and reference room of Institute of Geophysics, China Earthquake Administration (IGP, CEA). It took us nearly two months to collect resources. The documentation consulted includes *Observatory Communication Journal*, *Geomagnetic Observation Report*, *Chronicles of China Geomagnetic Observatory* and postal letters. The metadata is mainly stored in the *Chronicles of China Geomagnetic Observatory and Geomagnetic Observation Report*. An example of the cover of the bibliographic documents is shown in Fig. 1. The data of 1933-1954 were recorded in the *Geomagnetic Observation Report*. The observation was interrupted from April 1945 to December 1946 due to war. The data of 1955-1994 were stored in the DBF format. The data of 1995-2001 were stored in the BAS format. The data from 2002 to 2006 were stored in the Access database in the MDB format. The data of 2007, 2008 and 2010 were lost for unknown reasons. The data from September to December 2011, and July to October 2019 were missing due to the failure of absolute observation instruments. Data for other years from 2009 to 2019 are stored in Oracle database.

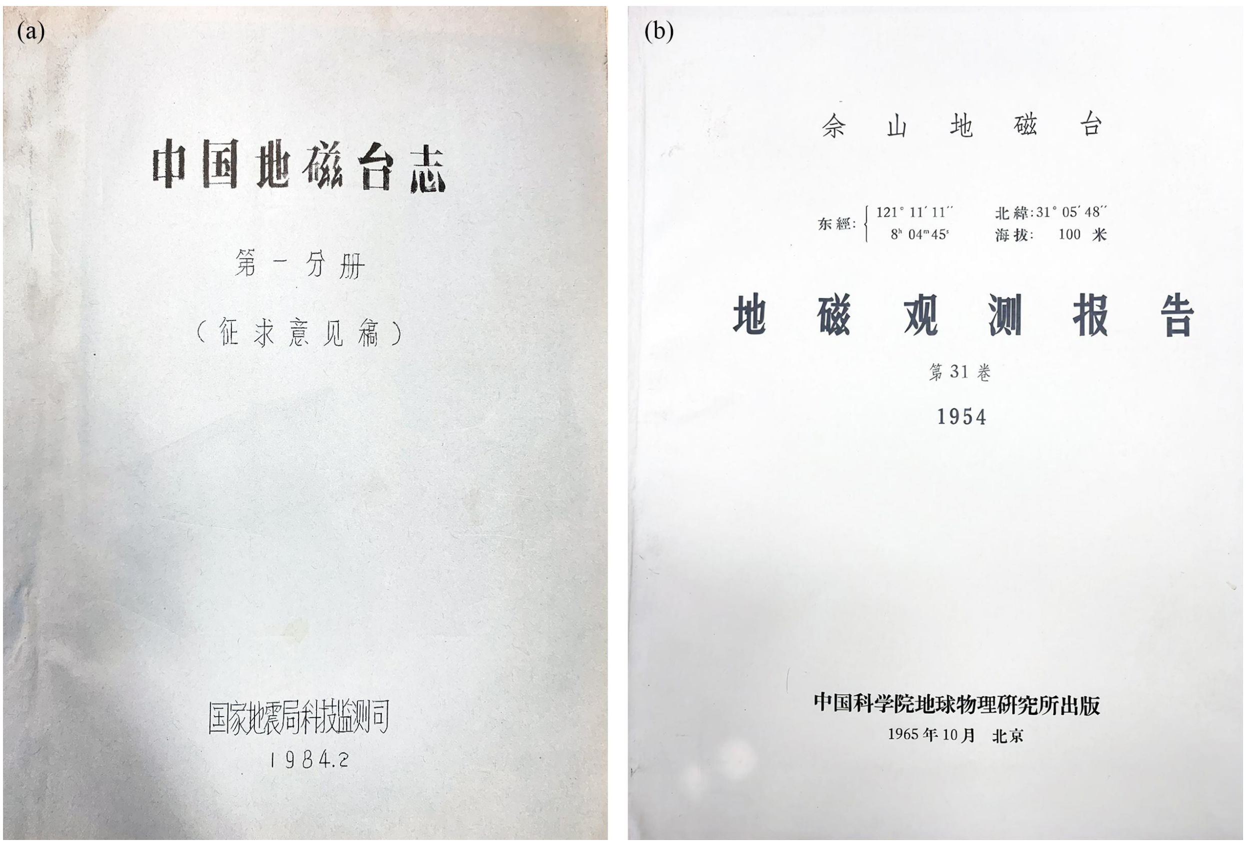

Figure 1 (a)Cover of the *Chronicles of China Geomagnetic Observatory* (Department of science, technology and monitoring, CEA, 1984) and (b)*Geomagnetic Observation Report* (Institute of Geophysics, Chinese Academy of Sciences, 1965)

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

Open Access Earth System Science Data Discussions

The Sheshan Geomagnetic Observatory (SSH) is presently run by IGP, CEA, and has been in operation for almost 150 yr. Its predecessor is Shanghai Xujiahui Observatory and Jiangsu Kunshan Lujiabang observatory. It was established in Xujiahui in 1874 and began continuous geomagnetic observation since then. It was moved from Xujiahui to Lujiabang in 1908, and was moved from Lujiabang to Sheshan in December 1932. The Sheshan Geomagnetic Observatory is located in Sheshan (Latitude: 31.1° N, Longitude:121.2°E), 20km to the southwest of Shanghai city. The geology of the vicinity of the observatory is Upper Jurassic to Down Cretaceous Andesite. The gradient of the field is about 2-3nT/m. The earliest absolute houses and recording room were built in 1932, they are made of non-magnetic material. The regular observation began in 1933.

Table 1 shows the absolute and relative instruments in SSH observatory from 1933 to 2019 and the measured geomagnetic elements at different periods. The information was retrieved from the bibliographic documents mentioned above. The first instrument set included as absolute instruments a Elliott (D measurements), a Smith (H measurement) and a Schulze (I measurement) since 1933. The continuous recordings of magnetic variations of D, H and Z were obtained respectively with a horizontal variometer (Toepfer) and a vertical intensity variometer (Godhavn) since 1933. Later, a few replacements of instruments took place in SSH observatory (Table 1). During this period, many jumps were seen in the relative recorded data due to the adjustment of the variometer, lightning stroke, earthquake and other reasons. These jumps have been corrected by the baseline, so that the absolute value is not affected. By the 2000, the SSH observatory was equipped with digital instruments. On Jan.1, 2003, the Schmidt Standard Theodolite was replaced by DIM-100/353766 Fluxgate Theodolite.

Table 1 Summary of instruments in the period from 1933-2019 at SSH

Component	Absolute measurements		Relative measurements	
	Date	**Instrument name and type**	**Date**	**Instrument name and type**
D	1933-1969.6	Magnetometer (Elliott/49)	1933-2000	horizontal variometer (Toepfer)
	1969.6-2002	Standard Theodolite(Schmidt /572144)	2000-2019	Fluxgate magnetometer (FGE)
	2003-2009	Fluxgate Theodolite (DIM-100/353766)		
	2009-2019	Fluxgate Theodolite (MINGEO DIM)		
H	1933-1992	Magnetometer (Smith/35416)	1933-2000	horizontal variometer (Toepfer)
			2000-2019	Fluxgate magnetometer

				(FGE)
I	1933-1992	Geomagnetic induction instrument (Schulze/42)		
	1993-2009	Fluxgate Theodolite (DIM-100/353766)		
	2009-2019	Fluxgate Theodolite (MINGEO DIM)		
Z			1933-2000	vertical intensity variometer (Godhavn)
			2000-2019	Fluxgate magnetometer (FGE)
F	1981-1985	Proton Precession Magnetometer (CHD-5/10)		
	1985-2002	Proton Precession Magnetometer (CZM-2)		
	2002-2019	Overhauser effect Magnetometer (GSM-19F)		

2.2 Data digitization

Because some records are handwritten or manual mimeographed, it is impossible to automate the digitization process. In order to facilitate the digitization and further application of these records, all the documents were photographed. It is also useful for re-verifying the data when the documents were not available. Using the character recognition program to recognize the photos and compare the consistency with the paper data, it is found that the recognition effect of character is not ideal. It may be due to the light color of the handwriting, or some of the handwriting is fuzzy and unclear. Therefore, the digitization was mainly performed by key input. We digitized the AHMVs of the three components of declination components (D), horizontal (H), and vertical (Z). We designed a set of Excel templates to unify the data entry format. The digital templates are very similar to the original data source in order to keep track of our work. The input templates include three workbooks, which are used to store the AHMVs of one year, including the AHMVs of D component, H component and Z component. Every AHMVs workbook consists of 14 worksheets, including text description, data worksheets from January to December of every year and automatic summary worksheet. The monthly data worksheet header includes the station code, measuring point ID, date, large value, 24 hourly mean values. The 24 AHMVs can be calculated by adding the large values to the 24 hourly mean values respectively. Missing values were marked as '99999'. An example of the Excel tables with digitized data is presented in Fig. 2. The "key input" approach is slower but has the lower error rate (Capozzi, 2020). After each month of data entry, the digitized data have been cross-checked with the original source values in order to identify and

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

remove transcription errors. After half a year's efforts, we input the data from 1933 to 1954 into the Excel template and completed the digitization of paper data.

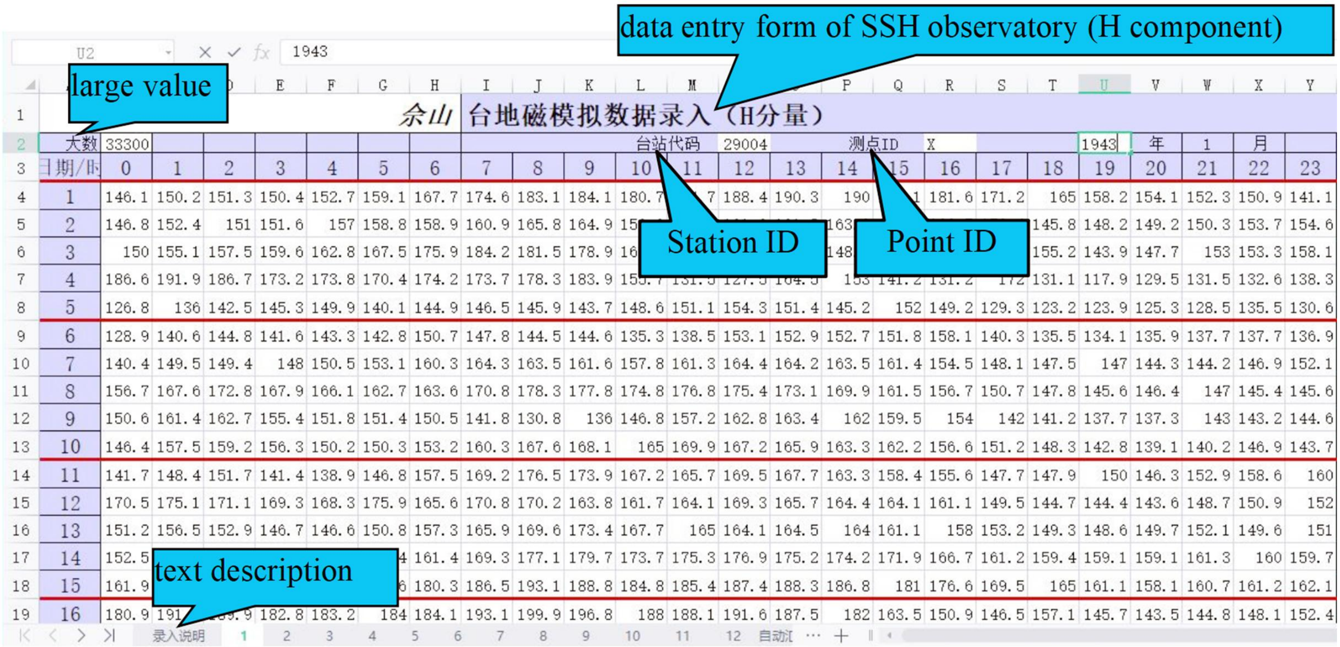

Figure 2. An example of the Excel tables with digitized data

2.3 Import data stored in various formats into Oracle Database

We developed a data import software (Fig. 3) to import data stored in various formats (XLS, DBF, BAS and MDB) into a unified Oracle database. In this way, all AHMVs are stored in the same database in a unified format. We call these stored data without any correction as the original AHMVs data. It is convenient for the subsequent analysis and application to store the data in the same database. This allows us to examine the data using Geomagnetic Data Processing Software (GDPST) (zhang, 2016), which was developed by Geomagnetic Network of China. It is developed based on Oracle database, and provides a convenient way in data processing, comparative analysis. The software also has the functions of the data query, data backup, and data download, etc.

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

Figure 3. Data import software

3 Quality control of digitized data

The aim of quality control (QC) is to check the continuity and reliability of geomagnetic observation data. The quality of geomagnetic data is often affected by the changes in the instrument or environmental conditions of the measurements, for example repair or re-calibration of the instrument, instrument replacement, station relocation, gradual changes of the observation environment, changes in observing process, etc., (Morozova et al., 2014, Zhang et al., 2016). The majority of such changes can lead to sudden breaks and jumps in the series of geomagnetic data, or gradual biases from the real geomagnetic characteristics. These phenomena are called inhomogeneities (liu et al., 2012; Morozova et al., 2014). Correction of inhomogeneities before any subsequent analyses is highly desirable (Mestre, 2013).

In this study, the QC was performed in order to check the quality of the rescued data. The inspection contents include evaluating the continuity of data, the accuracy of daily variation, and analyzing the influence factors of data quality.

3.1 the continuity of data

Based on the original AHMVs, the annual continuity is calculated, using the following formula (1):

$$C = (W_o - W_m)/W_o \qquad (1)$$

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

Here, C is the continuity of the AHMVs, W_o is the number of expected data in the chosen period, W_m is the number of missing data, see Fig. 4 for the continuity of data. The series in this study consists of data measured from 1933 to 2019. It has 5 gaps having a total of 67 monthly data missing, the number of missing data accounts for 6.5% of the total. "Every trial to correct data can produce unwished secondary effects in the result" (Linthe, 2013), so all gaps were not replaced by interpolation in this study. War and instrument failure are the main reasons for the lack of observation.

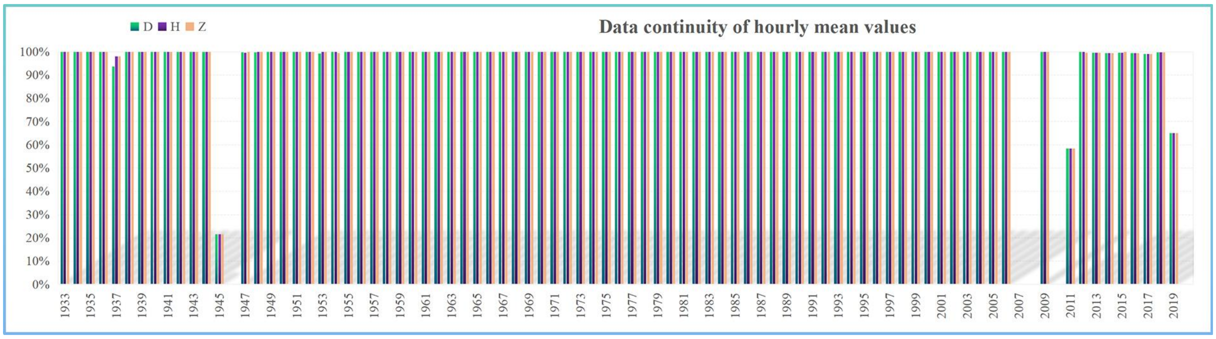

Figure 4. The continuity of the AHMVs from 1933 to 2019

3.2 the accuracy of the rescued data

We have designed a strict quality control (QC) procedure to ensure the accuracy of the rescued data. It consists of the following three steps:

1. Preliminary analysis of the series, detection of outliers.

In order to avoid the adverse impact of extreme data on the overall trend, we filter out clearly obvious outliers by the appropriate filtering function of Excel, such as the missing values which were marked as'99999', obvious input error, and so on.

2. Visual analysis of the series and their first time derivative at different timescales.

After removing the obvious outliers, we plot AHMVs of geomagnetic field components D, H and Z for all time from 1933 to 2019 (see Fig. 5). It can be seen from the figure that D, H and Z components have obvious trend. It is the secular variation （SV） in geomagnetic field with time. The noise in the plots mainly comes from the activity of external field. The most significant influence is on both the horizontal components D and H; its influence on the vertical component is minor. In the plots, we do not see obvious step and peak interference.

We checked the AHMVs plots of SSH month by month and found that sometimes the geomagnetic changes were quiet and regular, sometimes violent and irregular, and most of the days the geomagnetic changes were superimposed on the regular quiet day changes with some disturbances of different shapes and amplitude. As shown in Fig. 6, taking the AHMVs record

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

of January 1955 as an example, it can be seen from the figure that the record includes both regular periodic quiet day changes and complex perturbation, and the geomagnetic storms of January 18-19 are violent perturbation changes.

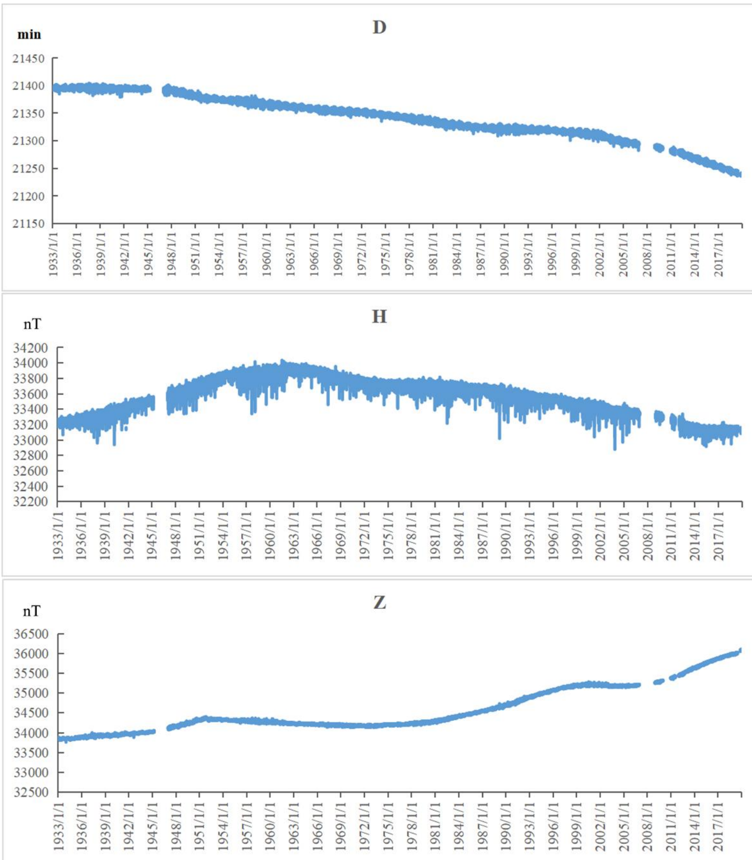

Figure 5. The AHMVs plots of D, H and Z components for all time from 1933 to 2019

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

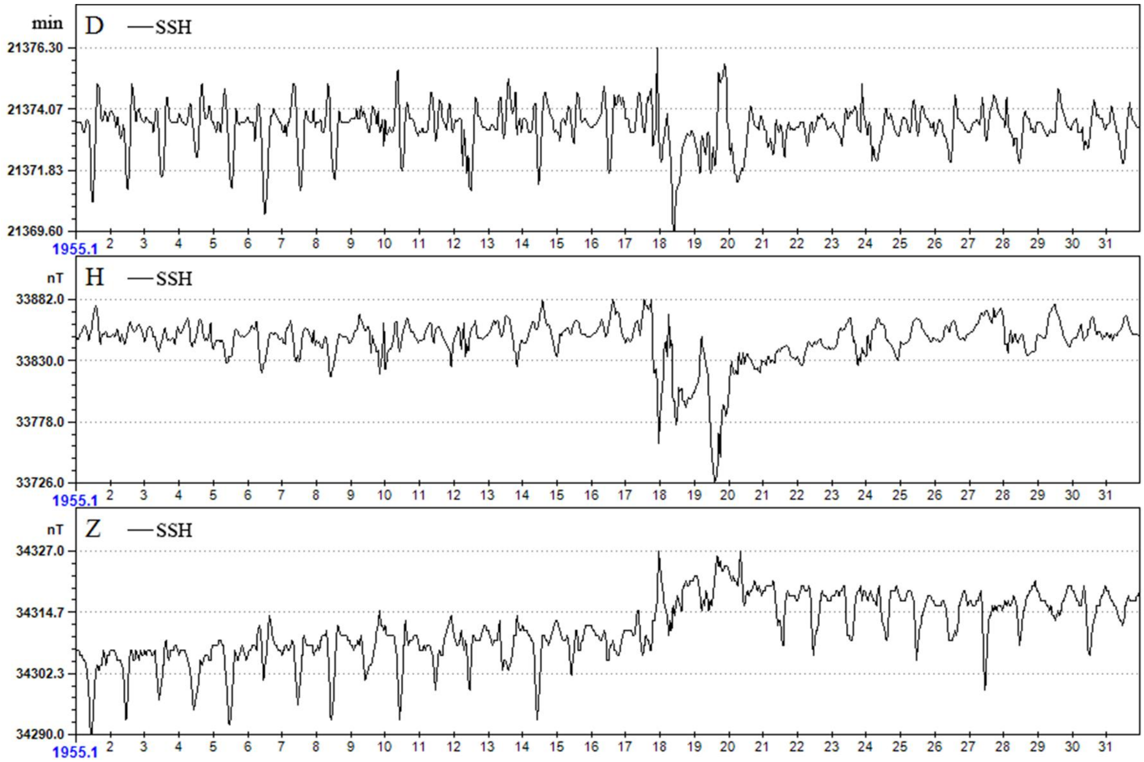

Figure 6. The AHMVs record of January 1955 at SSH

The absolute daily mean values (ADMVs) are computed from the AHMVs. It can be seen from the ADMVs curves (Fig. 7) that D, H and Z components have the same long-term change with AHMVs curve. The red dotted line is a 6th-order polynomial fitting curve. However, the trend will dominate all possibly existing inhomogeneities, making them undetectable. In order to eliminate the impact of trends and detect the inhomogeneities more effectively, we plot the first time derivative (FTD) of D, H and Z components (Fig. 8). We calculated the FTD using the consecutive values of daily series (Morozova et al., 2014). FTD plots are also particularly useful in evaluating artificial noise, especially interference in the shape of steps or spikes (Linthe et al., 2013; Pang et al., 2013; Chen et al., 2014). It can be seen from the figure that the data after the FTD eliminates the trend change, and the data are steady, going up and down within a certain range. ΔD varies within ±5 min, ΔH varies within ±200 nT, and ΔZ varies within ±30 nT.

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

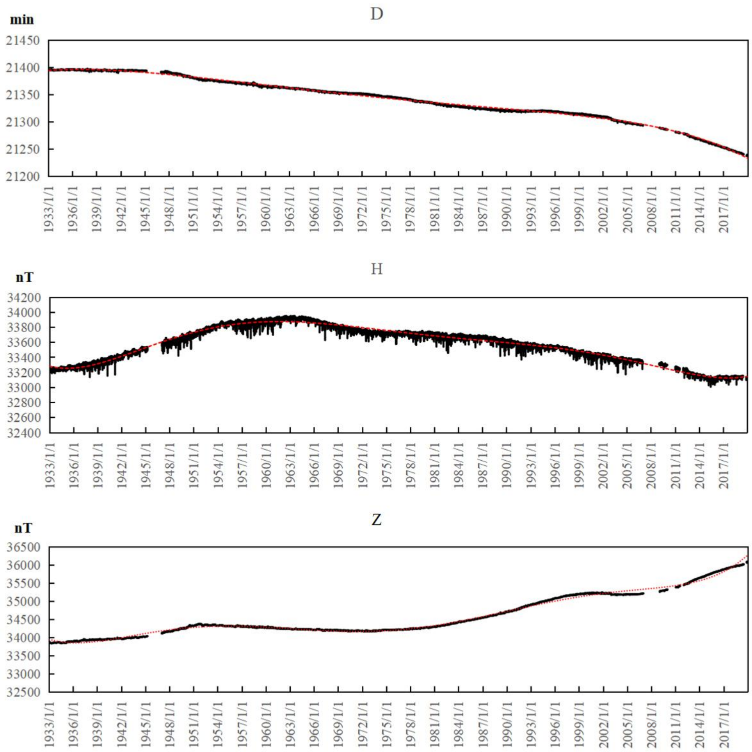

Figure 7. The ADMVs plots of D, H and Z components for all time from 1933 to 2019

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

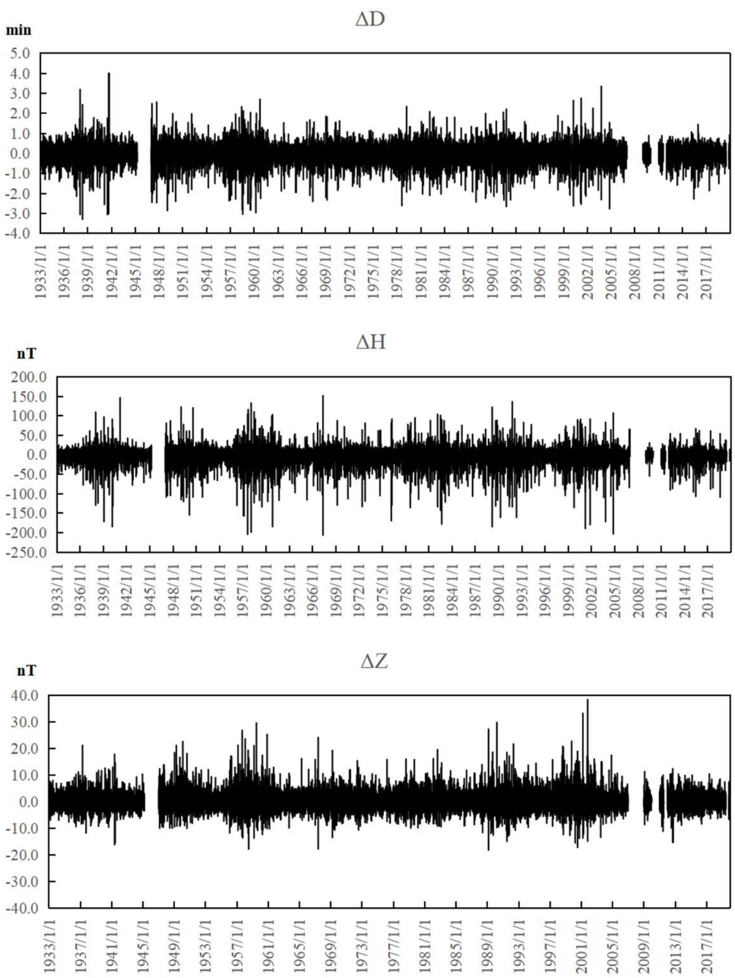

Figure 8. The FTD plots of D, H and Z components

3. The tolerance test detects the outliers and compared with geomagnetic indices.

We present a histogram of the FTD between 1933 and 2019 in Fig. 9, which aims at detecting the outliers further. This distribution has been easily modeled through the Gaussian probability density function (red curve). Red vertical lines indicate the lower and upper limits obtained by applying the criteria (3σ-μ) and (3σ+μ), where σ and μ are the standard deviation and mean for all time. We found that more than 98% of the FTD data points fall within the range of three times the standard deviations.

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

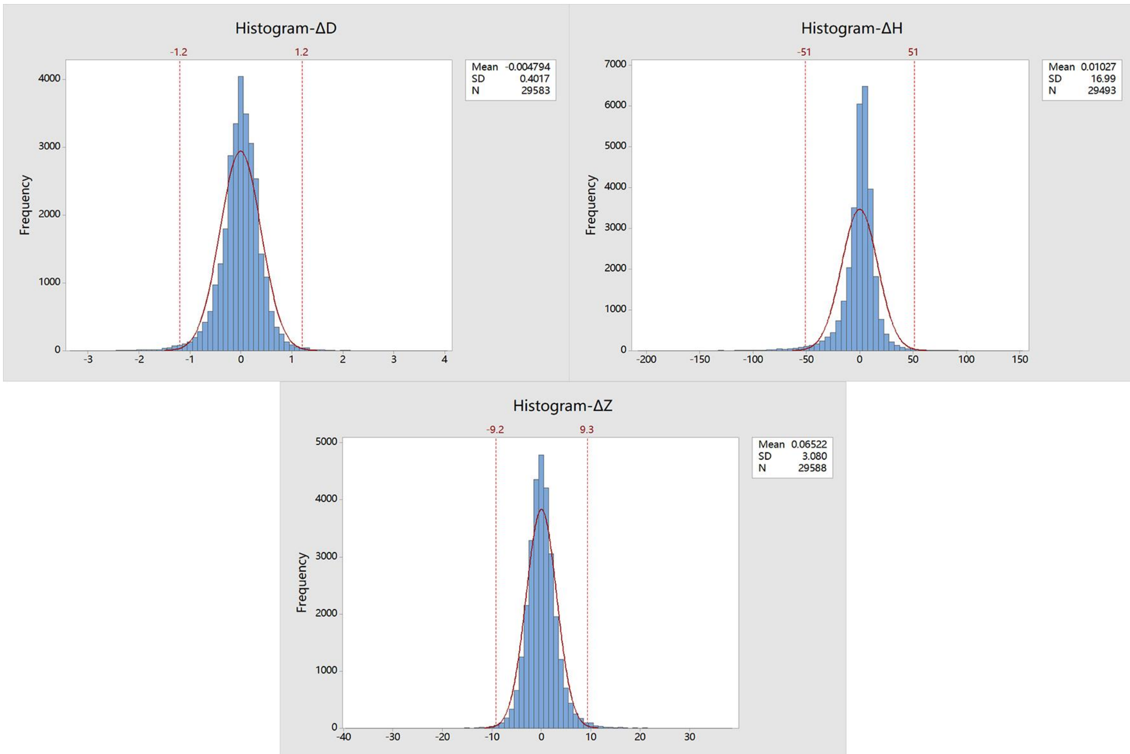

Figure 9. The histogram of the FTD of between 1933 and 2019

For the FTD data exceeding three times the standard deviation, we defined them as FTD outliers. We compared them with the AP indices of the day and the previous two days, and tried to find out the cause of the FTD out of tolerance, and took corresponding measures: a) When the Ap indices are greater than or equal to 12, it was considered to be caused by geomagnetic activity. The AHMVs at the corresponding time were not corrected. b) When the AP indices are less than 12, we carefully looked for the cause of each FTD outlier by comparing the daily variation curves of multiple stations (Sheshan, Chongming, Wuhan, Guangzhou, or Nanjing), and further consult the available documentation (*Observatory Communication Journal*, *Geomagnetic Observation Report*, *Chronicles of China Geomagnetic Observatory* and postal letters). A total of 168 FTD outliers were found, of which the D component appeared 63 times, the H component appeared 44 times and the Z component appeared 61 times. Preliminary analysis found that when the absolute value of D component FTD was less than 2', no abnormalities were found in the daily variation curves, accounting for about 46.9% of the total; and 48.4% of the outliers were related to magnetic perturbations; and only 2 values were questionable. All FTD outliers were related to magnetic perturbations at the H component. When the absolute value of Z component FTD was less than 14nT, no abnormalities were found in the daily variation curves, accounting for about 65.6% of the total; 21.3% of the outliers were related to magnetic disturbances; and the remaining 13.1% were questionable. A total of 10 FTD outliers were questionable

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

(see table 2) and no relevant and useful information was also recorded in the available documentation. The AHMVs at the corresponding time are not corrected. The AHVMs for the day the outlier appear and the day before were marked as questionable data, the quality flag is QC=Q. As shown in the Fig. 10, taking the FTD outliers of September 21 and 22, 2012 as an example, which obvious deviated from the original data series. Due to the lack of complete documentation, the questionable data were not corrected, just made the marks in the datasets, QC=Q. c) When the AP indices is less than 12, and a change is registered in the available documentation. These data can be accepted to be corrected. In Table 3, we list the date of the data to be corrected and the possible reasons recorded in the observatory logbooks and annual books. It is a inhomogeneity. It took place only on 1 January 2003, a modern absolute instrument named Fluxgate Theodolite DIM-100 replaced Geomagnetic Induction Instrument Schulze. This change led to a sudden step like change in size of about 3' and 40 nT in D and Z records respectively.

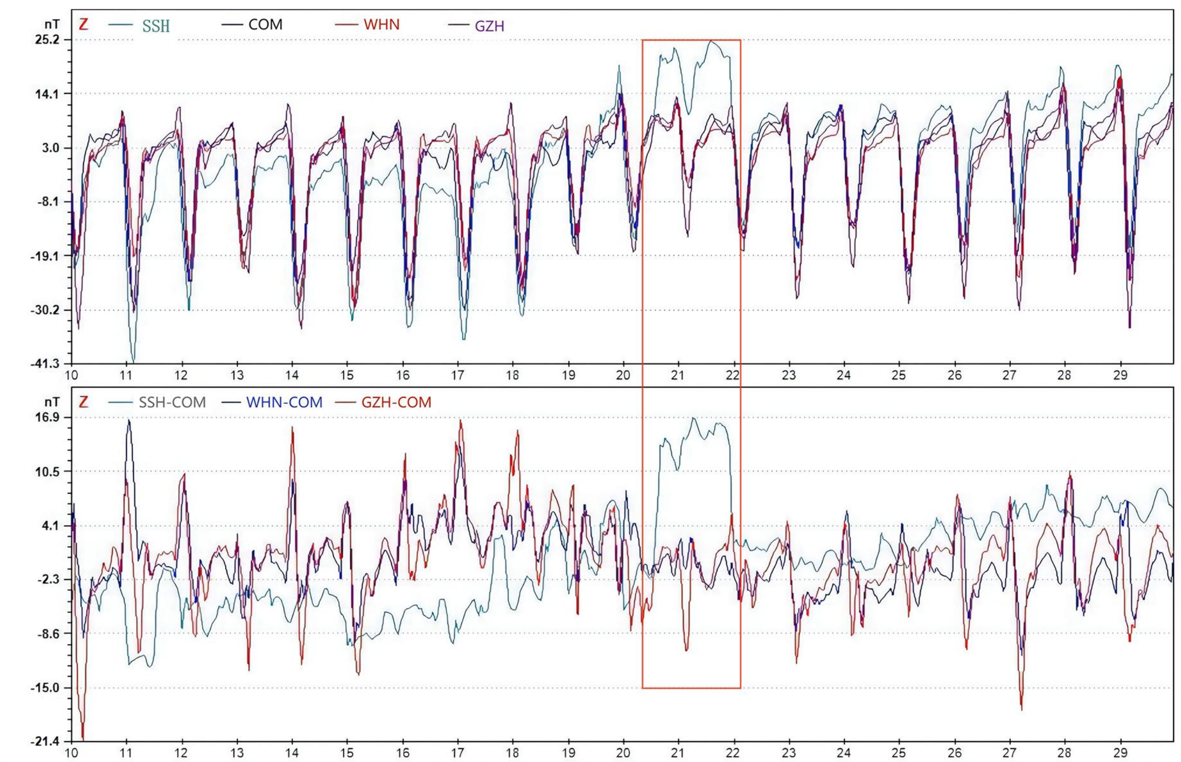

Figure 10. The contrast curve of SSH,COM, WHN and GZH from July 10 to 29, 2012

Table 2 List of dates where AHMVs marked as questionable data

Elements	**D**	**Z**
Date of the questionable AHMVs data	1958/6/19; 1958/7/29	1941/6/4; 1941/6/5; 1947/7/5; 1958/7/29; 1991/9/15; 2009/2/11; 2012/9/21; 2012/9/22

Table 3 Date and the possible reasons recorded of the data to be corrected

Date	**Metadata**	**Time interval**	**Correction values**	
2003.1.1	Instrument replacements	2003.1.1–2019.12.31	D: -2.7'	Z: -39.5nT

4 Correction of the selected homogeneity

As was mentioned above, we corrected only the data of D and Z components that occurred after 1 January 2003. The step-like change was corrected with substitution of the values by the mean values computed from preceding and following periods of 3 months. The break arises due to installation of a new instrument in 2003.1.1. The corrected value is 2.7' for D component and 39.5nT for Z component (see table 3). As shown in Fig. 11 and Fig. 12, we give the hourly and daily mean values curve of D and Z components before and after correction from January 1, 2001 to December 31, 2004. It can be seen that the homogeneity of corrected data have been greatly improved.

The AHMVs curve shows obvious annual and seasonal variations. The seasonal variation shows the variation range is large in summer and small in winter. The ADMVs curve of D component shows an obvious long-term trend of slow decline from 2001 to 2004. No obvious change characteristics can be seen in Z curve.

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

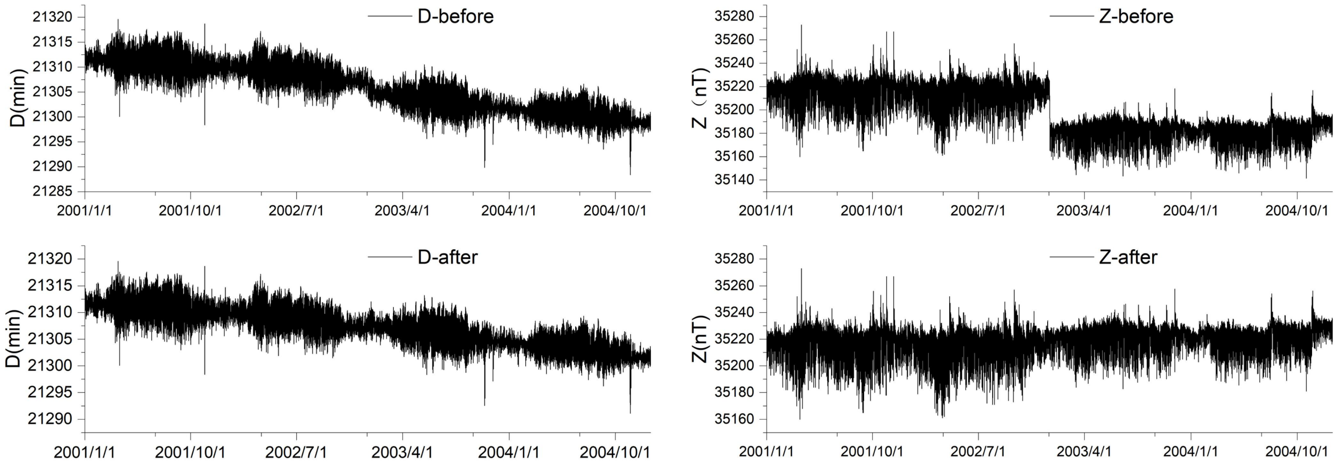

Figure 11. Plot of hourly mean values before and after correction

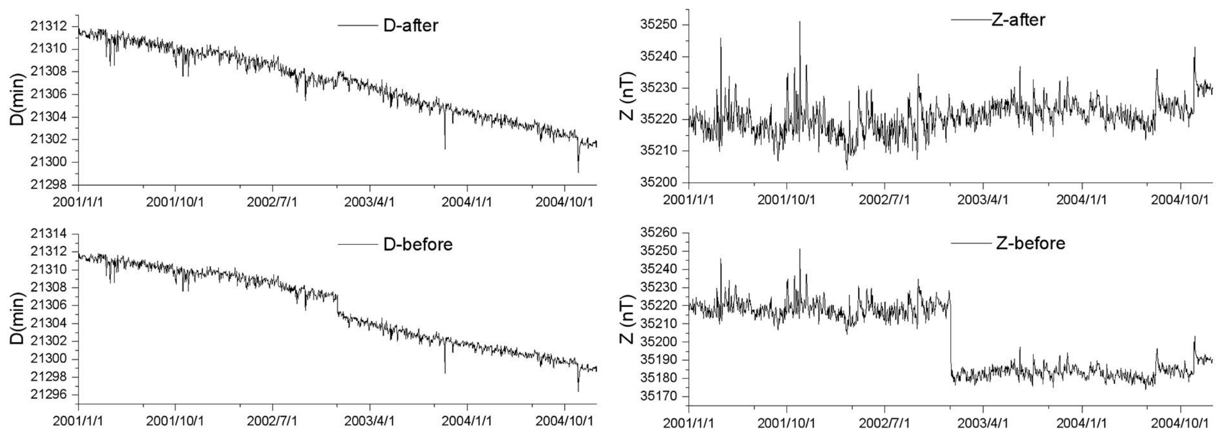

Figure 12. Plot of daily mean values before and after correction

5 Examination of the corrected series by compare with reference series

Inter-comparison of geomagnetic elements time series from adjacent observatories or geomagnetic models is also an important method to test accuracy and stability of data (Curto and Marsal, 2007). It can be used to detect spikes outliers, fluctuations or jumps at different time scales (Dawson, 2009; Chulliat et al., 2009; Zhang et al., 2016). Firstly, we compared the rescued data and those data from the reference station. In China, the regular observation of most geomagnetic observatories began in the 1980s. Only eight geomagnetic observatories were established during the international geophysical year (Rasson, 2011). Among the eight observatories, Guangzhou observatory (GZH) is closer to Sheshan observatory. It started observation early, and the quality of observation data are good. So we selected GZH as the reference

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

station. The GZH located in Guangzhou City, Guangdong Province, about 1240 kilometers northeast of SSH. It began geomagnetic observation in 1957. Due to the interference of Guangzhou metro operation, the new site Gaoyao Liantang Town was selected in 1996. The construction of the new station was completed at the end of 2001. The geomagnetic observation records officially began on January 1, 2002 in new place.

We used GDPST which offers very useful diagnostic procedures of the data quality to plot inter-comparison of values curve and their difference curve from SSH and GZH observatories on hourly and daily timescales to detect data with potential quality issues year by year. As an example, we present AHMVs, ADMVs and their difference curves of SSH and COM observatories in Fig. 13. At Fig. 13 in the upper panel (a)the AHMV and their difference curves are depicted, while in the lower panel (b) the ADMVs and their difference curves are plotted. On hourly scales the single components D, H and Z of SSH and GZH behave roughly identical, and their difference series slowly fluctuates (due to geomagnetic activity) around a certain range. Spikes are caused in most cases by external perturbations. On daily scales the components D, H and Z show roughly identical, but their differences coincide clearly with the variation of the geomagnetic field. It is because the distance between two observatories too far to offset completely the influence of internal and external source fields in different regions.

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

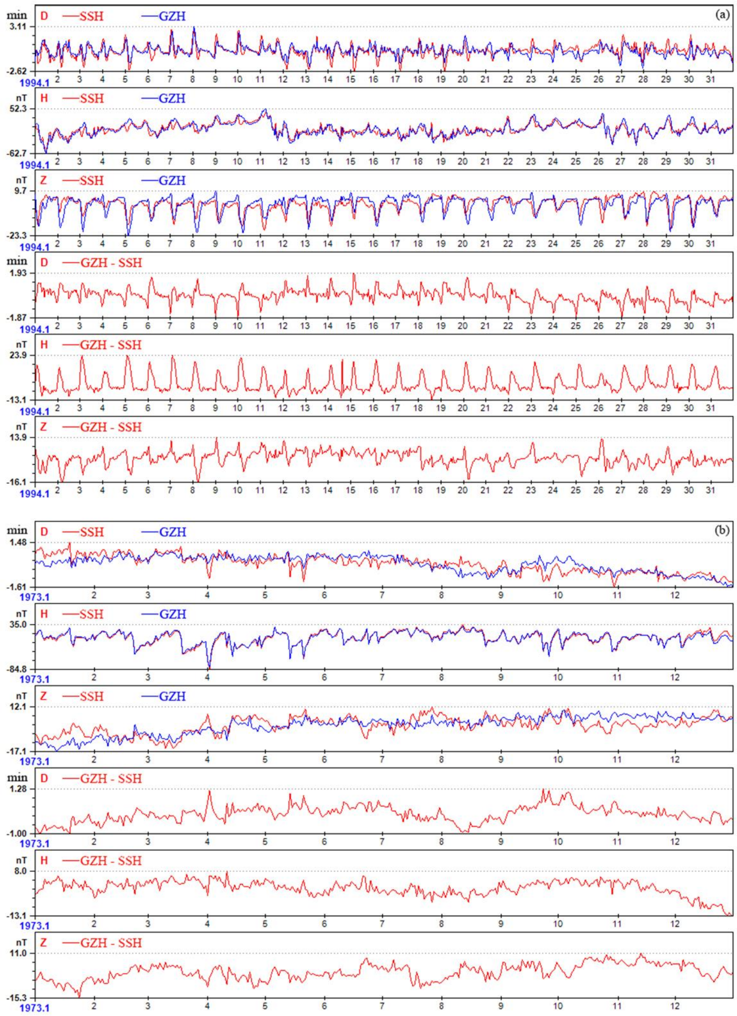

Figure 13. AHMVs, ADMVs and their difference curves of SSH and COM observatories. (a) the AHMVs and their difference curves. (b) the ADMVs and their difference curves

Comparing the measured values with the calculated values of the model for a long time scale is not only an important means to check the quality of the observational data, but also an important means to evaluate the accuracy of the model (Zhang, 2008; Thébault, 2010; Chen, 2012). One of the aims of geomagnetic observatories is the monitoring of SV (Reda et al., 2011). Secondly, to monitor the SV of SSH, we compared annual means values (AMV) series curve of X, Y, Z components calculated from the rescued records and these data calculated from the COV-OBS model (Gillet et al., 2013). As can be seen from the Fig. 14, the change trend of X, Y and Z components from SSH and COV-OBS model are very consistent. X component increased year by year before 1962 and generally decreased after 1962; From 1933 to 2019, the Y component shows a general downward trend and the Z component shows a general upward trend. There are differences between the AMV of SSH and these

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

data calculated from the COV-OBS model, the X component varies from -210 nT to -276 nT, the Y component varies from 17 nT to 94 nT, and the Z component varies from 198 nT to 289 nT. According to the preliminary analysis, the main reasons for the large difference between SSH and COV-OBS model may be the local magnetic anomaly in Sheshan area, the uneven distribution of global stations, the lack of modeling data and so on. This fully illustrates the importance of continuous and high-quality data in magnetic field modeling.

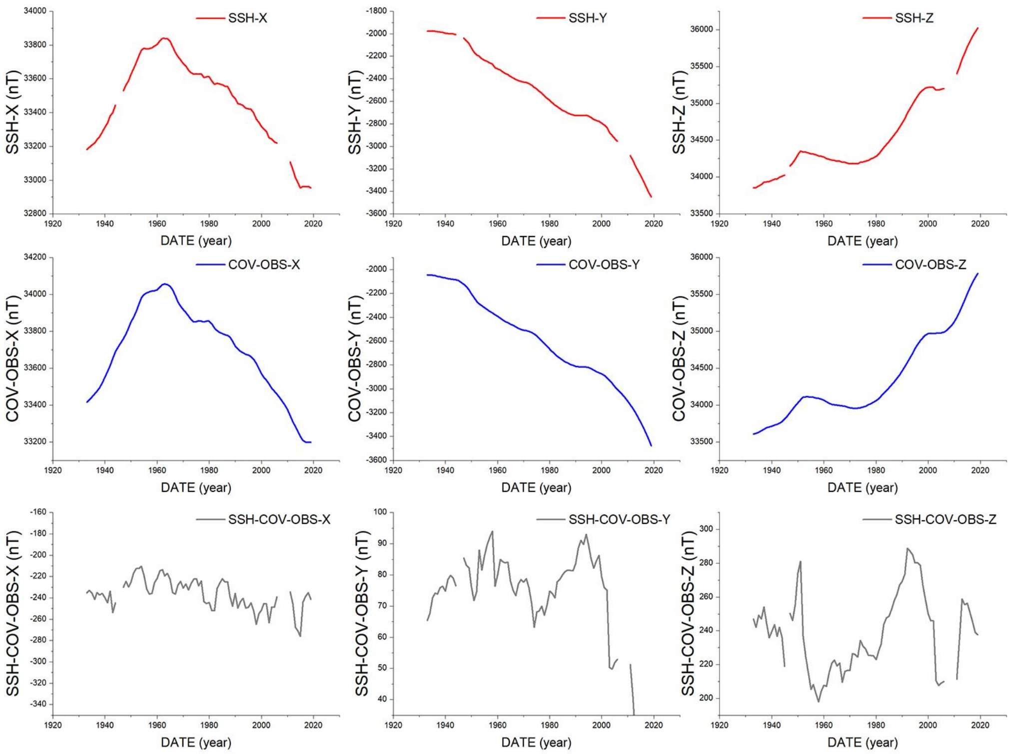

Figure 14. The AMV and their difference curves of X, Y, Z components calculated from the rescued records and the COV-OBS model

6 Application examples of SSH datasets

We calculated the first time derivative using a difference between two consecutive years as SV, and plot the first time derivative from the rescued data and COV-OBS model to detect possible geomagnetic jerks over the past 90 years. Geomagnetic jerks are defined as as V-like or Λ-like changes in the SV and occur in a time period of a few months (Courtillot and LeMouël, 1984; Morozova et al., 2014; Kang et al., 2020). "The geomagnetic jerks are due to interactions of the core field and the rapid time-varying core flow" (Kuang et al., 2011). Since Malin and Hodder (1982), Courtillot and Le MouëL (1984) discovered the geomagnetic jerk in 1969, ten jerks have been detected in observatories from 1933 to 2020, of which 1969, 1978 (Alexandrescu et al.,1996), 1991 (De Michelis et al., 1998), 1999 (Mandea et a.l, 2000; Zhang et al., 2008), 2003 (Olsen et al., 2007; Feng et al.,

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

2018; He et al, 2019), 2007 (Kotzé et al. 2010; Chulliat et al., 2010) and 2014 (Brown et al., 2016; Kloss and Finlay, 2019; Finlay et al., 2016; Kang et al., 2020) were global events. In addition, there were two local events, which occurred in 1949 (Mandea et al., 2000) and 2011 (Chulliat and Maus, 2014; Kotzé and Korte, 2016). In 2017, there are similar characteristics of geomagnetic jerks, which may be a new geomagnetic jerk (He et al, 2019; Pavón- Carrasco et al., 2021). The jerks are more easily seen in the eastward component（Y） of geomagnetic secular variation. Not all jerks can be detected around the world, some seem to be seen only in limited regions (Morozova et al., 2014), and its occurrence time is not exactly same at each observatory.

Figure 15 gives the SV of the annual Y series from SSH and COV-OBS model, and the blue dotted line is the third-order moving average curve of Y component SV at SSH. Nine jerks are clearly seen in the plot. They occurred in 1950, 1971, 1978, 1993, 1999, 2004, 2008, 2013 and 2018. Except for the jerk occurred in 1978 and 1999, other events were a little later is seen at SSH. Between 2008 and 2017, only one jerk (in 2013) is seen at SSH, which is inconsistent with the study of other scholars mentioned above. They observed jerks event in 2011 and 2014 respectively.

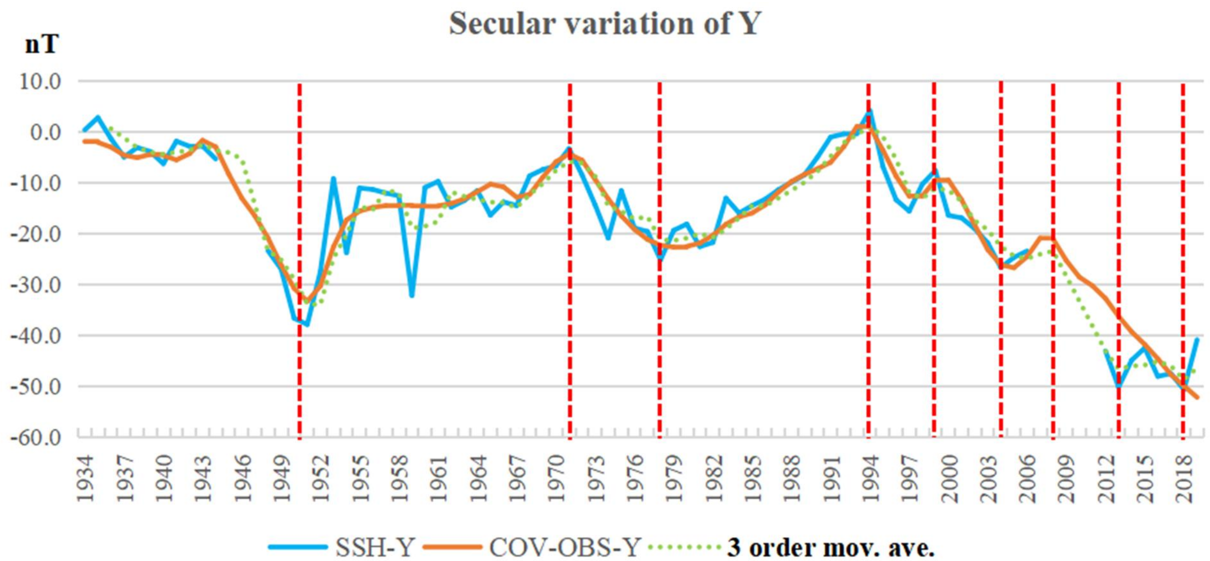

Figure 15. The SV of the annual Y series from SSH and COV-OBS model

7 Conclusion

This paper presents the acquisition process, the quality control, data correction, the quality examination and application examples of the datasets of SSH from 1933 to 2019. The quality examination results show that the corrected data have a good agreement with the reference observatory data and the model data. This fully indicates that the rescued data are of good quality. The datasets are valuable for studying the geomagnetic daily variation, geomagnetic field model construction and secular variation. It should be noted that these data (see table 2) are used with caution for the reasons mentioned above. A few problems were found in the acquisition of geomagnetic historical datasets: ① The rescue of paper data is a time-consuming and laborious work. For example, the font color of some reports is light, which makes automatic recognition difficult. We can only recognize and input by key. ② Some metadata of the SSH has been missing, which brings difficulties to the identification and correction of data. Therefore, we believe that we should do our best to do the rescue work of historical data, to avoid the irreparable losses over time. Our goal for the next phase is to rescue the historical data from 1874-1932 at SSH, as well as from other observatories, to provide more high-quality data for the geomagnetic science community.

Data availability

The digitized and quality-controlled AHMVs data are available at: https://doi.org/10.5281/zenodo.6584285 (zhang et al, 2022). The data are provided in Microsoft Excel format, including the observed absolute hourly mean values files of the three components (D, H and Z) at SSH for 1933-2019, and also the metadata files about the datasets.

Author contribution

SZ performed data correction, quality assessment and the analysis of application example . SZ prepared the manuscript. CH calculated X, Y, Z components from the COV-OBS model and performed revision of the manuscript. GZ was responsible for the digitization of paper data from 1933 to 1945 of SSH, and provided digital series of the geomagnetic components 2009 and 2011. CC designed a set of Excel templates. JW developed the data import software.

Funding information

This work was funded by the National Natural Science of Foundation of China (grant no. 41974073) and the Macau Foundation and the pre-research project of Civil Aerospace Technologies of China (grant no. D020308).

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgement

We thank SSH, GNC and reference room of Institute of Geophysics, China Earthquake Administration for providing the valuable data resources.We are also grateful to anonymous reviewers. We sincerely thank all the staffs who have ever worked or are currently working at SSH.

References

Alexandrescu, M., Gibert, D., Hulot, G., Le Mouël, J. L., and Saracco, G.: Worldwide wavelet analysis of geomagnetic jerks, J Geophys Res, 101（B10）, 21975–21994, 1996.

Bolduc, L., Langlois, P., Boteler, D., and Pierre.: A study of geomagnetic disturbance in Quebec. I. General results, IEEE Transactions on Power Delivery, 13, 1251-1256, 1998.

Bolduc, L., Langlois, P., Boteler, D., and Pirjola, R.: A study of geoelectromagnetic disturbances in quebec. II. Detailed analysis of a large event, IEEE Transactions on Power Delivery, 15(1), 272-278, 2002.

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

Boteler, D. H., Pirjola, R. J., and Nevanlinna, H.: The effects of geomagnetic disturbances on electrical systems at the Earth's surface, Advances in Space Research, 22, 17-27,1998.

Brown, W., Beggan, C., and Macmillan, S.: Geomagnetic jerks in the Swarm Era[C]//Proceedings of the ESA Living Planet Symposium, Czech, Prague, 9−13, 2016.

Capozzi, V., Cotroneo, Y., Castagno, P., De Vivo, C., and Budillon, G.: Rescue and quality control of sub-daily meteorological data collected at Montevergine Observatory (Southern Apennines), 1884–1963, Earth Syst. Sci. Data, 12, 1467–1487, https://doi.org/10.5194/essd-12-1467-2020, 2020.

Chen, B., Gu, Z. W., Gao, J. T., Yuan, J. H. and Di, C. Z.: Geomagnetic secular variation in China during 2005-2010 described by IGRF-11 and its error analysis, Progress in Geophysics, 27(2), 512-521, 2012. (in Chinese).

Chen, J., Jiang, Y. L., Zhang, X. X., Chen, C. H., Yang D. M., and Liu H. F.: The design of HVDC discrimination and processing system for geomagnetic network,Seismological and Geomagnetic Observation and Research, 35(3/4), 271-274, 2014. (in Chinese).

Chulliat, A. and Maus, S.: Geomagnetic secular acceleration, jerks, and a localized standing wave at the core surface from 2000 to 2010, J Geophys Res, 119(3), 1531–1543, 2014.

Chulliat, A., Peltier, A., Truong, F., and Fouassier, D.: Proposal for a new observatory data product: quasi-defifinitive data. 11th IAGA Scientifific Assembly, Abstract Book 94 pp, 2009.

Chulliat, A., Thébault, E., Hulot, G.: Core field acceleration pulse as a common cause of the 2003 and 2007 geomagnetic jerks, Geophys Res Lett, 37: L07301. 2010.

Clarke, E., Flower, S., Humphries, T., McIntosh, R., McTaggart, F., McIntyre, B., Owenson, N., Henderson, K., Mann, E., MacKenzie, K., Piper, S., Wilson, L., and Gillanders, R.: The digitization of observatory magnetograms, Poster presented at: 11th IAGA Scientific Assembly, 23-30 Aug 2009, Sopron, Hungary, 2009.

Courtillot, V. and Le Mouël, J. L.: Geomagnetic secular variation impulses, Nature, 311(5988), 709-716, 1984.

Curto, J. J. and Marsal, S.: Quality control of Ebro magnetic observatory using momentary values, Earth, Planets and Space, 59(11), 1187-1196, 2007.

Dawson, E., Reay, S., Macmillan, S., Flower, S., and Shanahan, T.: Quality control procedures at the World Data Centre for Geomagnetism (Edinburgh) [C]// IAGA 11th Scientific Assembly, Sopron, Hungary, 23-30 Aug 2009, 2009.

De Michelis, P., Cafarella, L., and Meloni, A.: Worldwide character of the 1991 geomagnetic jerk, Geophys. Res. Lett., 25(3), 377–380, 1998.

Department of science, technology and monitoring, CEA: The Chronicles of China Geomagnetic Observatory, 1984. (in Chinese).

Dong, X. H., Li, X. J., Zhang, G. Q., Shi, J., and Liu, C.: The study of digital identification of magnetogram, Seismological and Geomagnetic Observation and Research, 30(6), 49-55, 2009. (in Chinese)

Feng, Y., Holme, R., Cox, G. A., and Jiang, Y.: The geomagnetic jerk of 2003.5: Characterisation with regional observatory secular variation data, Physics of the Earth and Planetary Interiors, 278: 47–58, 2018.

Finlay, C. C., Olsen, N., Kotsiaros, S., Gillet, N., and Tøffner-Clausen, L.: Recent geomagnetic secular variation from swarm and ground observatories as estimated in the chaos-6 geomagnetic field model, Earth, Planets and Space, 68(1), 112, doi:10.1186/s40623-016-0486-1, 2016.

Gao, M. Q. and Hu, Z. Y.: Establishment of historical data and database of rescue Sheshan observatory, Proceedings of the 9th Annual Academic Meeting of China Geophysical Society in 1993, 262, 1993. (in Chinese).

Gillet, N., Jault, D., Finlay, C. C., and Olsen, N: Stochastic modeling of the Earth's magnetic field: Inversion for covariances over the observatory era, Geochemistry Geophysics Geosystems, 14(4):766–786, 2013.

Guo, S. X., Liu, L. G., Pirjola, R. J., and Wang, K. R., and Dong, B.: Impact of EHV power system on geomagnetically induced currents in UHV power system, IEEET Power Deliver, 30, 2163-2170, 2015.

He, Y. F., Zhao, X. D., Zhang, S. Q., Yang, D. M., and Li, Q.: Geomagnetic jerks based on the midnight mean of the geomagnetic field from geomagnetic networks of China, Acta Seismologica Sinica, 41(4), 512–523, doi: 10.11939/jass. 20190009, 2019. (in Chinese).

Institute of Geophysics, Chinese Academy of Sciences: Geomagnetic Observation Report, 1965. (in Chinese).

Kang, G. F., Gao, G. M., Wen, L. M., and Bai, C. H.: The 2014 geomagnetic jerk observed by geomagnetic observatories in China, Chinese Journal of Geophysics, 63(11), 4144-4153, doi:10.6038/cjg2020N0337, 2020. (in Chinese)

Kappenman, J. G.: Geomagnetic storms and their impact on power systems, IEEE Power Engineering Review, 16: 5-8, 1996.

Kloss, C., and C. C. Finlay.: Time-dependent low-latitude core flow and geomagnetic field acceleration pulses, Geophysical Journal International, 217.1, 140-168,doi:10.1093/gji/ggy545, 2019.

Korte, M., Mandea, M., Linthe, H. J., Hemshorn, A., P Kotzé, and Ricaldi, E.: New geomagnetic field observations in the South Atlantic Anomaly region, Annals of Geophysics, 52(1), 65-81, 2009.

Kotzé, PB. and Korte M.: Morphology of the southern African geomagnetic field derived from observatory and repeat station survey observations: 2005−2014, Earth, Planets and Space, 68, 23, 2016.

Kotzé, PB.: The 2007 geomagnetic jerk as observed at the Hermanus magnetic observatory, Phys Comment, 2, 5-6, 2010.

Kuang, W.J. and Tangborn, A.: Interpretation of Core Field Models, in: Geomagnetic Observations and Models (Vol. 5), edited by: Mandea, M. and Korte, M., Springer Sciencee+Business Media, 295-309, 2011.

Linthe, H. J., Reda, J., Isac, A., Matzka, J., and Turbitt, C.: Observatory data quality control-the instrument to ensure valuable research. In: Hejda, P. (Ed.), Proceedings of the XVth IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition and Processing: extended abstract volume, (Boletin Roa / Real Instituto y Observatorio de la Armada en San Fernando; 3, 2013), 173-177, 2013.

Liu, C. M., Liu, L. G. and Pirjola, R.: Geomagnetically induced currents in the high voltage power grid in China, IEEE T Power Deliver, 24, 2368-2374, 2009.

Liu, J., Ma, Z. F., Fan, G. Z., and You, Y.: Research on the comparison of different homogeneity test methods, Meteorological Monthly, 38(9), 1121-1128, 2012. (in Chinese).

Liu, L. G., Ge, X. N., Wang, K. R., Zong, W., and Liu C. M.: Observation studies of encroachment by geomagnetic storms on high-speed railways and oil-and-gas pipelines in China, Sci Sin Tech, 46, 268–275, doi: 10.1360/N092015-00279, 2016. (in Chinese).

Liu, L. G., Liu, C. M., Zhang, B., Wang, Z. Z., Xiao, X. N., and Han, L. Z.: Strong magnetic storm's influence on China's Guangdong power grid, Chinese Journal of Geophysics, 51(4): 976-981, doi: 10.3321/j.issn:0001-5733.2008.04.004, 2008. (in Chinese).

Malin, S. R. C. and Hodder, B. M.: Was the 1970 geomagnetic jerk of internal or external origin? Nature, 296, 726–728, 1982.

Mandea, M., Bellanger, E., and Le Mouël, J. L. A geomagnetic jerk for the end of the 20th century?, Earth Planet Sci Lett，183（3/4）, 369–373, 2000.

Mestre, O., Domonkos,P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Böhm, R., Aguilar, E., Guijarro, J., Vertachnik, G., Klancar, M., Dubuisson, B., and Stepanek, P.: HOMER : a homogenization software – methods and applications, Vol. 117, No. 1, January–March 2013, pp. 47-67

Morozova, A. L., Ribeiro,P., and Pais, M. A.: Homogenization of the historical series from the Coimbra Magnetic Observatory, Portugal, Earth System Science Data, https://doi.org/10.5194/essd-2020-317, 2020.

Morozova, A. L., Ribeiro,P., and Pais, M. A.: Correction of artificial jumps in the historical geomagnetic measurements of Coimbra Observatory, Portugal, Ann. Geophys., 32, 19–40, 2014.

Olsen, N. and Mandea, M. Investigation of a secular variation impulse using satellite data: The 2003 geomagnetic jerk, Earth Planet Sci Lett, 255（1/2）, 94–105, 2007.

Pang, J.Y., Chen, J., Wang, C., Teng, Y.T., Zhao, Y.G., and Li, Z.G.: The principle of power interference and its automatic processing during geoelectrical resistivity observation, Seismological and Geomagnetic Observation and Research, 34(5/6),117-122,2013. (in Chinese).

Pavón-Carrasco, F.J., Marsal, S., Campuzano, S. A., and Torta, M.: Signs of a new geomagnetic jerk between 2019 and 2020 from Swarm and observatory data, Earth, Planets and Space, 73(1):1-11, 2021.

Peng, F,.Shen, X., Tang, K., Zhang, J., Huang, Q., Xu, Y., Yue ,B., and Yang, D.: Data-Sharing Work of the World Data Center for Geophysics, Beijing Data Sci J, 6, 404–407, 2007. (in Chinese).

Rasson, J.L., Toh, H., and Yang D. M.: The Global Geomagnetic Observatory Network, in: Geomagnetic Observations and Models (Vol. 5), edited by: Mandea, M. and Korte, M., Springer Sciencee+Business Media, 1-25, 2011.

Reay, S. J., Clarke, E., Dawson, E., and Macmillan, S., T.: Operations of the World Data Centre for Geomagnetism, Edinburgh, Data Science Journal, 12, WDS47-WDS51, 2013.

Reda, J., Fouassier, D., Isac, A., Linthe, H. J., Matzka, J., and Turbitt, C. W.: Improvements in Geomagnetic Observatory Data Quality, in: Geomagnetic Observations and Models (Vol. 5), edited by: Mandea, M. and Korte, M., Springer Sciencee+Business Media,127-148, 2011.

https://doi.org/10.5194/essd-2022-203
Preprint. Discussion started: 27 June 2022

Sergeyeva, N., Gvishiani, A., Soloviev, A., Zabarinskaya, L., Krylova, T., Nisilevich, M., and Krasnoperov, R.: Historical K index data collection of Soviet magnetic observatories, 1957-1992, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-270, 2020.

Thomson, A. W. P.: Geomagnetism Review 2019, British Geological Survey Open Report, OR/20/00852 pp, 2020.

Xu, W. Y.: Physics of Electromagnetic Phenomena of the Earth , Hefei: University of Science and Technology of China press, 2009.(in Chinese).

Zhang, S. Q., Fu, C. H., He, Y. F., Yang, D. M., Li, Q., Zhao, X. D., and Wang, J. J.: Quality Control of Observation Data by the Geomagnetic Network of China, Data Science Journal, 15: 15, pp. 1–12, DOI: 10.5334/dsj-2016-015, 2016.

Zhang, S. Q., Yang, D. M., Li, Q., and Zhao, Y. F.: The 1991 and 1999 jerks in China. Earthquake Research in China， 24(3): 253–260, 2008. (in Chinese).

Zhang, S. Q., Yang, D. M., Li, Q., and Zhao, Y. F.: The consistence analysis of IGRF model value and annual mean value of some geomagnetic observatories in China, Seismological and Geomagnetic Observation and Research, 29(2), 42-29, 2008. (in Chinese).

Zhang, S. Q., Zhu, G. H., Wang, J. J., and Chen, C. H.: Quality-controlled geomagnetic hourly mean values datasets of Sheshan observatory from 1933 to 2019 [Data set], Zenodo, https://doi.org/10.5281/zenodo.6584285, 2022.

Zhao, X. D., He, Y. F., Chen, J., Zhang, S. Q., Li, Q., and Yuan, Y. R.: The distribution of ring current and field-aligned current during storms based on ground observatory data. Chinese Journal of Geophysics, 62(9): 3209-3222, doi: 10.6038/cjg2019M0268, 2019. (in Chinese).