

Management and quality assurance of observational data recorded by geomagnetic observatories in Egypt

Emad M. H. Takla^{*,1}, Taha Rabeh¹, Aalaa Samy¹

⁽¹⁾ National Research Institute of Astronomy and Geophysics, Cairo, Egypt

Article history: received July 24, 2024; accepted August 08, 2024

Abstract

The National Research Institute of Astronomy and Geophysics (NRIAG) is an Egyptian research institution that operates two Geomagnetic Observatories (GMOs), namely the Misallat (MLT), and Abu-Simbel (ABS) observatories. Data availability, reliability and quality are the basis for any scientific research. So, it is essential to check and manage the quality of Geomagnetic Data (GMD) and prepare them in a proper Data Format (DF) before publishing them for research studies. This paper focuses on the management and quality assurance of the GMD recorded by the MLT and ABS observatories, and presents the current efforts deployed for managing, examining and processing them as well as converting their original DF into international DF. The raw data has been processed mathematically using MATLAB scripts. Results show that GMD from the MLT and ABS observatories suffer from various non-natural disturbances. Electric power lines, industrial and human activities nearby the MLT observatory as well as some internal problems with the measuring devices, might have caused these disturbances. In addition, the study examined the impact of these disturbances on the quality of recorded data. The comparison of raw data from the two observatories and other INTERMAGNET data showed a good correlation between them before 2015. From 2015 until October 2022, the noise level mostly at the MLT became very high and the natural Ultra Low Frequency (ULF) emissions at the MLT data were completely hidden by the background noise. The study recommends that the Magnetically Clean Environment (MCE) must be resorted again as well as fixing the measuring instruments especially at the MLT observatory to provide high quality data that can be suitable for all aspects of scientific research and satisfy the INTERMAGNET requirements.

Keywords: Geomagnetic data; Geomagnetic observatories; Data management; Quality control; Background noise; Data format

1. Introduction

The Geomagnetic Field (GMF) is an invisible shield around the Earth, and it protects humans and other living organisms from the dangerous effects of space radiation and particles. The intensity of this field is inconstant and numerous sources can cause variations in this field. These variations have periods ranging from seconds to millions of years on timescales. Short-timescale variations are generally related to external sources, while longer-timescale variations are mainly linked with internal sources.

Geomagnetic Observatories (GMOs) are main locations for acquiring Geomagnetic Data (GMD), so they continuously measure the strength and direction of the Earth's magnetic field (Macmillan, 2007). Karl Friedrich Gauss in the 1830s organized a network of scientific observers in Germany to observe and measure variations in the GMF, and that was the first international effort in the field of geomagnetic observations (Garland, 1979). In the late 1960s, GMOs started using the digital electronic instruments in measuring and recording the GMF variations. The transition from simple analogue magneto-grams to digital recordings along with the development of technological capabilities have significantly increased the availability and accessibility of the observational data and a real revolution in data types was started. In addition, metadata (information about the data) also had its own development to follow the fast and large spreading of the various data sets. Most significantly, the observational GMD are now recorded, handled, stored, and distributed using computers and Internet (Reay et al., 2011). GMD obtained by ground-based stations are essential sources for investigating the processes occurring in the Earth's interior and near-Earth environment. Moreover, the GMD provided by various platforms can give a picture of the structure of the Earth's magnetic field and its interaction with geospace. So, the reliability and precision of data are essential not only for scientific research but also for commercial applications.

Monitoring and investigating the acquired GMD need special facilities, which are classified as Magnetically Clean Environments (MCEs). The MCE means that no external geomagnetic disturbances exist. In this case, the recorded data are expected to reflect only the natural processes, and not the effects of man-made activities (Pirjola et al., 2007). The man-made and external geomagnetic disturbances can be generated due to the presence of electric power lines, metal pipelines, factories, and vehicles or ferromagnetic bodies within the Earth's crust (Love, 2009). In addition, the GMOs should be in an area where no large geomagnetic gradients observed due to the magnetized crustal rocks (Jankowski and Sucksdorff, 1996; Neska et al., 2012). Moreover, an observatory must produce no artificial noise by its own activities. Therefore, any activities that occur inside or nearby the observatory must be carefully managed to ensure that magnetometer measurements are undisturbed (Lowes, 2009).

An important task of a GMO is to monitor and track both short and long-term geomagnetic variations (Zhang et al., 2016). Therefore, GMOs are intended to run over very long periods of time (decades). Despite remarkable achievements in the manufacturing of new instruments for geomagnetic measurements, the mathematical techniques for data processing and data quality are mainly depending on the actual circumstances and conditions in which the geomagnetic measurements are performed in the GMOs (Jankowski and Sucksdorff, 1996).

2. Motivation

It is obvious to state that a MCE is a significant factor for the success of a GMO in providing reliable GMD, but this situation is not easy to accomplish in practice. As it was previously mentioned, disturbances in the geomagnetic measurements can be resulted from instrument failure, power supply failure, environmental disturbance and human activities. The production of high-quality and long-term data series is a challenge that includes management, Quality Control (QC) methods and the process for publishing datasets for the scientific community.

In fact, there is no real QC strategy for both the MLT and ABS observatories data. So, based on the experience gained during handling long-term GMD at both the National Research Institute of Astronomy and Geophysics (NRIAG) and the International Center of Space Weather Science and Education (ICSWSE) at Kyushu University, Japan that formerly known as Space Environment Research Center (SERC), the current study proposes a management and QC strategy of the actual practice for data check, managements and noise elimination as well as converting the original Data Format (DF) [text format] into an international DF [IAGA2002 format]. However, these procedures are tough processes, but they are essential to maintain a high level of GMD quality.

3. Instrumentation and data acquisition

NRIAG has a very long history of astronomical and geophysical observations, so it is one of the oldest research institutions in Africa and one of the world heritage spots in science and technology. NRIAG owns and operates different types of astronomical and geophysical observatories allocated all over the country to satisfy the mission of NRIAG and among them are two GMOs: the Misallat (MLT), and Abu-Simbel (ABS) observatories.

The data used in the current study were collected during regular geomagnetic measurements at the MLT and ABS observatories. The Misallat GMO is currently the primary GMO of NRIAG. It is located about 70 km to the

southwest of Cairo in Fayuom governorate at 29.515°N and 30.889°E coordinates. The Misallat GMO was established in 1960 as a substitute for the geomagnetic observations in Helwan observatory due to the electrification of the railways that made accurate geomagnetic measurements impossible in Helwan area. Currently, the observatory uses triaxial fluxgate magnetometer (Magson) for the X, Y, and Z variation measurements every one second. Moreover, absolute measurements are made using the DI flux and Overhauser magnetometer. The observatory IAGA code is MLT (currently the MLT observatory is not a member of INTERMAGNET network). It is important to mention that the MLT observatory is currently surrounded by industrial and urban areas (only a few hundred meters away from the observation site). Therefore, there are several sources of potential noise, such as factories, houses, electricity power lines and highway road, which can influence the geomagnetic measurements.

The ABS observatory is the second GMO in Egypt and its installation was completed in 2007 and became operational by the end of 2008. The observatory is located about 15 km to the north of Abu-Simbel city in Aswan governorate in south Egypt at 22.48°N and 31.54°E coordinates. The observatory uses Magson magnetometer for the X, Y, and Z variation measurements at one-second sampling rate. Currently, the magnetometer provides high-resolution data (five readings per second). The ABS observatory is located very far from urban and industrial areas (in the desert), but it is worth pointing out that the basement rocks almost appear at the surface in the area where ABS observatory is located. Moreover, extreme climate conditions at the ABS observatory are of great importance where this area is characterized by a sharp desert climate with big seasonal temperature changes. To avoid or minimize any temperature drift in the recorded data, both observatories were constructed as double-wall buildings. In addition, at the ABS observatory, the sensor hut was constructed at about one meter below the ground surface (inside the double-wall building) to minimize the temperature drift as much as possible.

4. Data management and quality control strategy

The continuous growth of geomagnetic observations requires not only good measurement practice but also adequate and significant methods for their management, processing, and analysis (Linthe et al., 2012). Modern infrastructures sometimes produce growing levels of perturbation on a number of GMOs. As a result, several observatories have been moved to new places with sufficiently quiet conditions for the long-term geomagnetic measurements. It is worth mentioning that moving (relocating) a GMO is not only very expensive and time consuming issue but also the measured data can no longer be considered as continuous series. If moving the observatory to a new suitable location is not possible, it is necessary to find a way to correct the noisy or disturbed data. To correct artificially disturbed GMD series, some mathematical techniques and methods can be developed for their cleaning and noise elimination (Linthe et al., 2012).

High-quality, well-documented datasets are necessary for successful scientific research. So, the purpose of QC process is to examine the consistency and precision of the recorded GMD. As a result, the current study presents a management, processing, and QC strategy for handling the recorded data at both observatories. This strategy begins with data recording and collection from the observatory site. Then, the collected data need to be examined and processed to ensure sufficient data quality; in other words, to be convenient and useful for scientific research works. A number of MATLAB scripts were prepared for the data analysis and processing such as plotting, noise detection and elimination, band-pass filtering and power spectra density.

It is worth mentioning that every trial to correct (denoising) GMD may produce unwished secondary effects in the result (i.e. erasing or destroying natural ULF signals in the processed data). So, the processed data should be compared with original data to ensure that the reconstructed data show the same properties of the original one. For this reason, the filtering, denoising and formatting techniques applied in the current study were designed carefully to avoid any damages or erasing of the natural characteristics of the recorded data. Stages of data management and QC strategy that need to be applied for the MLT and ABS observatories are described below as well as their efficiency.

4.1 Data compilation and continuity

Before the initial data check, the collected data undergoes an important step, which is the data compilation (compiling short interval data records). For example, the data recorded at the MLT observatory are saved into

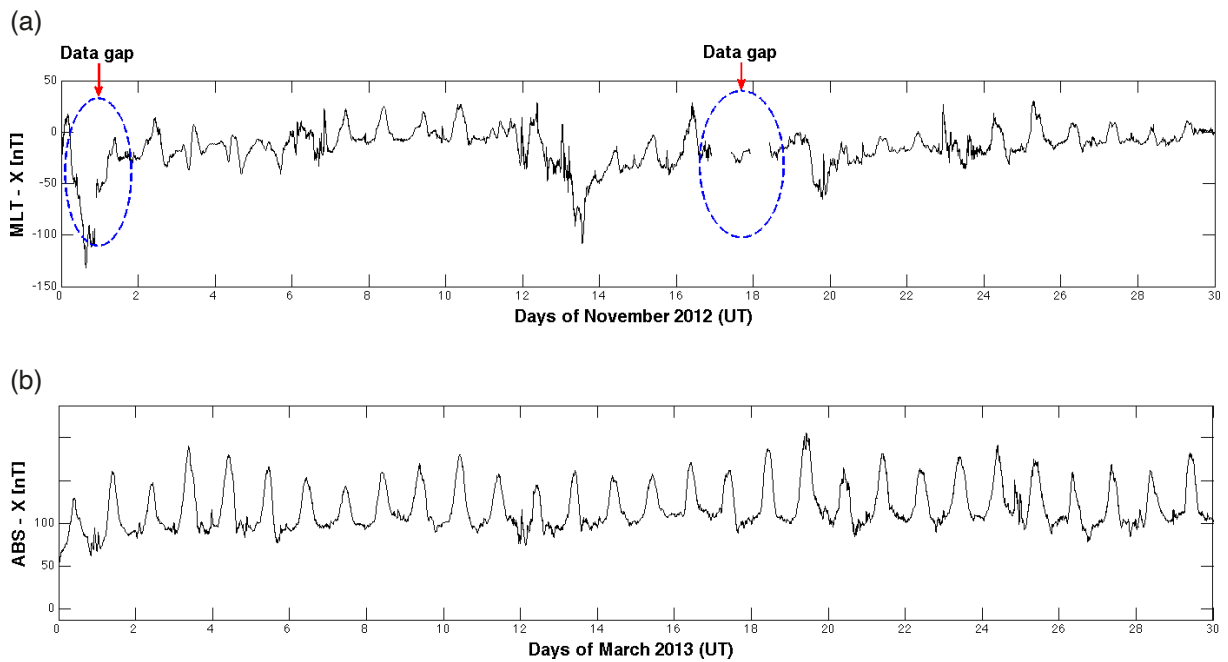


Figure 1. Monthly plot of the X-component data recorded at the MLT and ABS observatories.

one-hour data files, so it is important to compile these one-hour data files (24 files per day) into one data file per day (daily data files) for processing and analysis of data quality review.

After data compilation, the daily and monthly data will be plotted to examine the reliability and continuity of the recorded geomagnetic components (X-, Y-, and Z-components). Two examples of the monthly data plotting are presented in Fig. 1 from the MLT (a) and ABS (b) observatories to check the continuity of the recorded data. From Fig. 1, we notice that there are some data gaps in the MLT data that can be related to problems in the power source at this observatory. On the other hand, recorded data from the ABS observatory are continuous without such data gaps (for the checked data). One main problem concerning the ABS observatory is the absence of a regular (real time) check of the system operation. Sometimes data recordings stop for a long period of time (several months) due to system failure. As an example of such problem, there was no data from the ABS observatory during the first half of 2022, which emphasizes the importance of data QC.

4.2 Noise identification and elimination

This stage includes identification of any noise forms in the recorded data and their elimination. To develop a mechanism to deal with noises effectively, it is necessary to recognize their nature. In general, technical problems or failures in measuring equipment or interference related to devices operating near by a magnetometer can generate regular noises. Identifying the noise form is an important part in solving noise problems. The other part is to use an efficient method to eliminate these noises (Khomutov et al., 2017). The detected noise forms in the MLT and ABS data are described below.

4.2.1 Noise spikes

Spikes are one of the most common forms of contaminating signals in the recorded GMD and usually they are not related to natural sources or processes. They can be considered as short period signals (usually with duration of few seconds or measurements) but with significant amplitudes. Spikes are generally characterized by well-defined sharp leading and back edges that are similar in amplitude (Curto et al., 2009 and Khomutov et al., 2017). In addition, spikes with duration of one measurement cycle are most likely to be associated with hardware problems or interferences from nearby sources (Khomutov et al., 2016). The spikes amplitude is an important characteristic element in

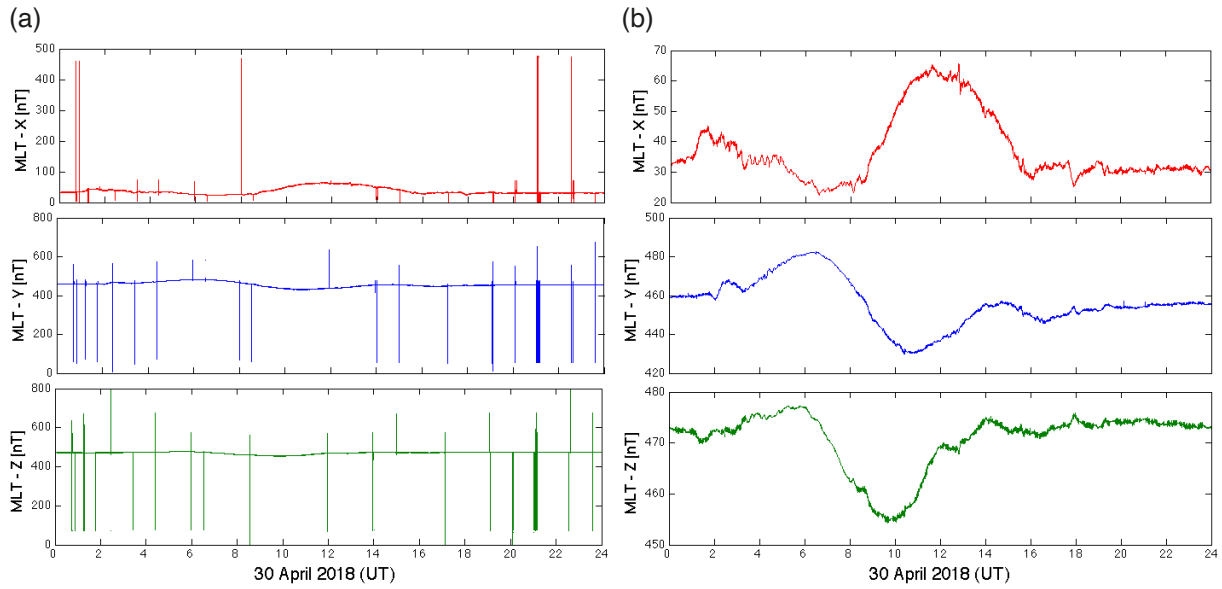


Figure 2. An example of spikes in GMD recorded at the MLT observatory. (a) Raw data with spikes (vertical lines). (b) Filtered data after spikes removal.

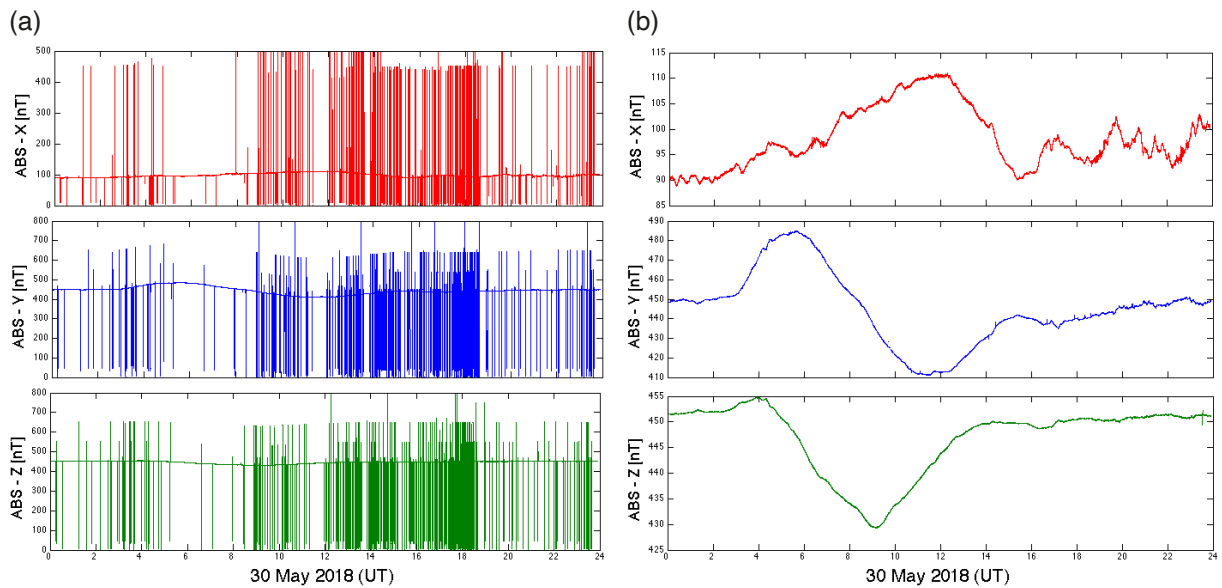


Figure 3. An example of spikes in GMD recorded at the ABS observatory. (a) Raw data with spikes (vertical lines). (b) Filtered data after spikes removal.

recognizing them. Spikes with amplitudes of tens of nanotesla or more have a low probability of being caused by natural sources. The Fig. 2a shows one second of recorded data from the MLT observatory on 30 April 2018. As we can clearly see, there are irregular and frequent spike noises in the X, Y and Z components with amplitudes of more than 300 nT. These spikes have large amplitudes and produce significant noise in the one-second GMD. A similar picture of interference was also observed in the raw data recorded at the ABS observatory on 30 May 2018 as shown in Fig. 3a. It should be noted that the interference is irregular, as it appears in GMD in different days with different occurrence times and amplitudes. The duration of these spikes does not exceed one measurement. The most probable reasons for this noise are instrumental failure and human activities (some constructions occurred during this period inside the observatory area). When averaging one-second data containing such asymmetric noisy signals to get one minute-data, the obtained mean values are also biased. Since these spikes have big amplitudes, it is necessary to eliminate them. There are different approaches to eliminate these spikes. If a file with raw data has

text format, it is possible to remove the unreliable records manually. This option is practically very hard to apply due to the huge data sets. Another possible approach is to create an intermediate file, usually in text format, in which noisy data are manually or semi-automatically replaced by the values that indicate unreliable measurements. This option is particularly useful if the initial measurements are recorded in files with binary format as applied in MAGDAS network. In this case special converter software is usually used, which allows an operator to encode the required data as unreliable. The third approach is to apply a digital filter to remove such spikes. In the current study, depending on the above-mentioned spike characters, a special filter was designed using MATLAB script for the detection and removal of such noises. The idea of this filter is to determine the position and amplitude of these spikes in the raw data then replace them with special values NaN (Not-a-Number). This filter is useful for eliminating short periods noise signals such spikes. An example of the outcome of this clearing process is shown in right side panels (b) of Fig. 2 and 3 from the MLT and ABS observatories respectively. The results show high reliability of this filter for detecting and removing both low and large amplitude spikes.

4.2.2 Jumps

Jumps in the recorded data can be seen as spike noises characterizing by continuous measurements between their leading and back edges. If data jumps have significant sharp edges (sharp intensity increase or decrease), it will be easy to identify them. Changing the instrument parameters or technical operations can be the main source of data jumps (Khomutov et al., 2017).

An example of jumps observed in the ABS observatory data is shown in Fig. 4. The effect is visible in the geomagnetic components, which recorded on 21 February 2014 as shown in the upper panel of Fig. 4 (for X-component). This example represents the situation that is very widespread in GMD. Mathematical approaches of pattern recognition can be a suitable solution to overcome such noise problem. Detection and solving processes require information about measurement conditions and undisturbed geomagnetic records as references (in this case, MLT and HER data were the reference data). So, we were able to solve this issue effectively as shown in the lower panel of Fig. 4 by a tool designed using MATLAB script depending on the pattern recognition technique (Chow and Huang, 2008).

4.2.3 Background noise level

Background noise level is a term referring to the total noise that is irrelevant to natural signals. It is a significant indicator for the quality and reliability of the GMD. In general, the background noise is not connected with a certain frequency range or time domain. This type of noise can be generated either by external sources at the observatory site or by problems with the magnetometer or its sensitivity. The background noise problem can be solved by two ways. The first way is the mathematical solution that can be used to reduce background noise (filtering technique). The second way is an organizational solution that can be done by moving the magnetometer to a MCE or replacing magnetometer by another one that is less sensitive to such external noises (Linthe et al., 2012; Khomutov et al., 2017).

For examining the background noise, it is very useful to compare the recorded data with other data sets from nearby GMOs as we mentioned before. Background noise levels are usually measured at times when the GMF is comparatively quiet. The most common technique for estimating and examining noise level at a certain observatory is to analyze the rate of change of the measured field during low geomagnetic activity (Love 2006). The Fig. 5 shows the background noise for the X-component during 9 February 2021 at the MLT and ABS observatories compared with the noise level for data recorded at the Hermanus (HER) GMO (as a reference point). The noise levels for the X-component recorded at the MLT and ABS observatories are about ± 1 and ± 0.15 nT respectively, see Fig. 5. On the other hand, HER observatory has a very low background noise level, which is less than ± 0.1 nT (lower panel of Fig. 5). This means that the noise level at the MLT observatory is about 10 times higher than that of the HER observatory, while the noise level at the ABS observatory is about 1.5 times higher than that of the HER observatory. It is important here to mention that, applying a filter to reduce the background noise can reduce the noise level and make the data smooth but on the other hand it can destroy the small natural geomagnetic variations such as the magnetic pulsations (Ultra Low Frequency emissions). So, dealing with raw data (1-second

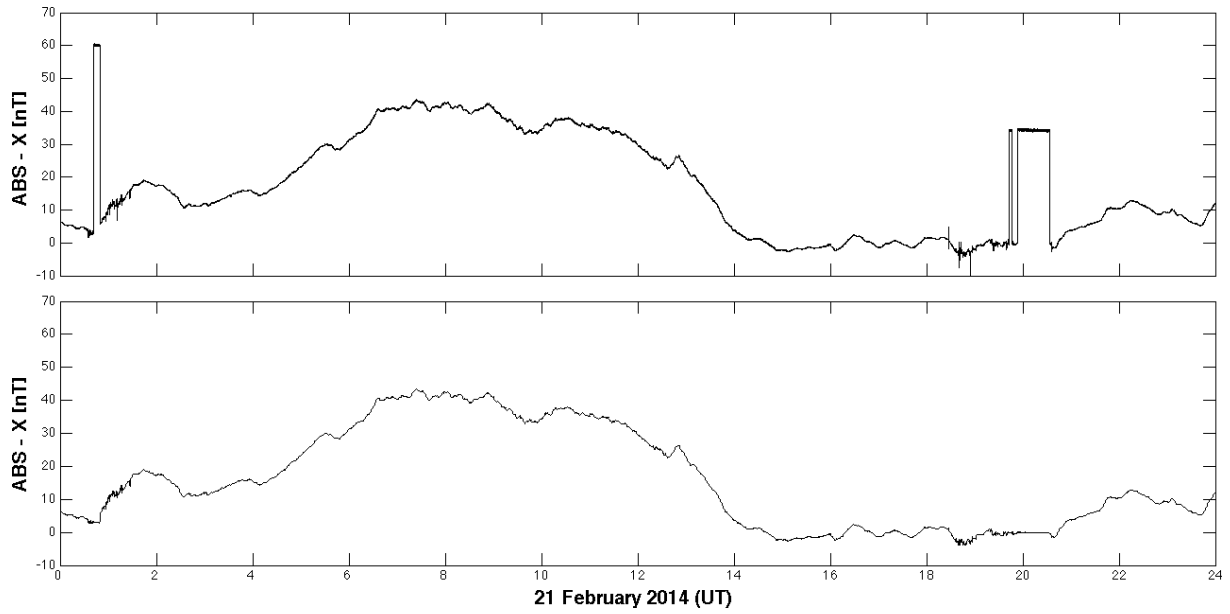


Figure 4. The upper panel shows an example of jumps in the GMD recorded at the ABS observatory on 21 February 2014 between 19:00-21:00 UT, while the lower panel shows the processed data after fixing the jumps.

data) to reduce the background noise level is a very critical issue and it needs a well-trained and experienced magnetologist. In case of the MLT observatory data, random background noises (with a wide frequency range) are superimposed on the original natural signals. As a result, it is impossible to remove these background noises or reduce their amplitude without destroying or eliminating the natural geomagnetic fluctuations (geomagnetic pulsations).

The noise level at the MLT observatory is always high during both night and day times. Moreover, it can be observed from the upper panel of Fig. 5 that between 14.00 and 16.30 (UT), the noise level is low compared with other time intervals. This period is between 4.00 and 6:30 pm (Local time) which is still a period with high human activities in the industrial area nearby the MLT observatory. So, it can be concluded that the main part of the noise level at the MLT observatory is highly related to an instrument failure rather than the nearby human activities.

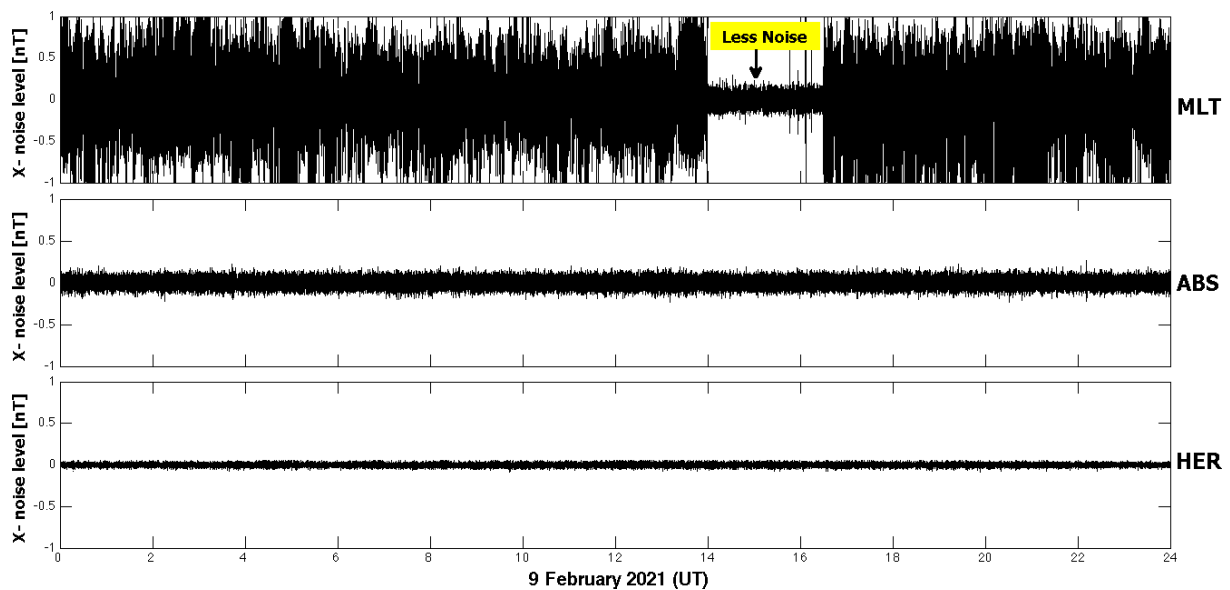


Figure 5. Background noise of the X-components recorded at the MLT and ABS observatories compared with a reference observatory data (HER).

4.3 Data quality check

One-minute data is important for studying geomagnetic variations caused by the magnetospheric and ionospheric currents, in particular studying daily variations of the GMF and geomagnetic storms and substorms. Therefore, the geomagnetic X-components (1-minute data) from some GMOs are used to examine the quality of the processed data (MLT and ABS) during the storm-time and the disturbance in the daily geomagnetic variations at the MLT and ABS observatories compared with the X-components from some INTERMAGNET observatories as reference points (nearly with the same local times). The Fig. 6 shows the variations of X-component (1-minute data) obtained from KIV, SUA, MLT, ABS, HBK and HER observatories on 19 February 2014. It can be seen from this figure that there is a correlation between all of them, which indicates that the averaged (after filtering) 1-minute data from the MLT and ABS observatories can be suitable for studies that require 1-minute data (not high-resolution data).

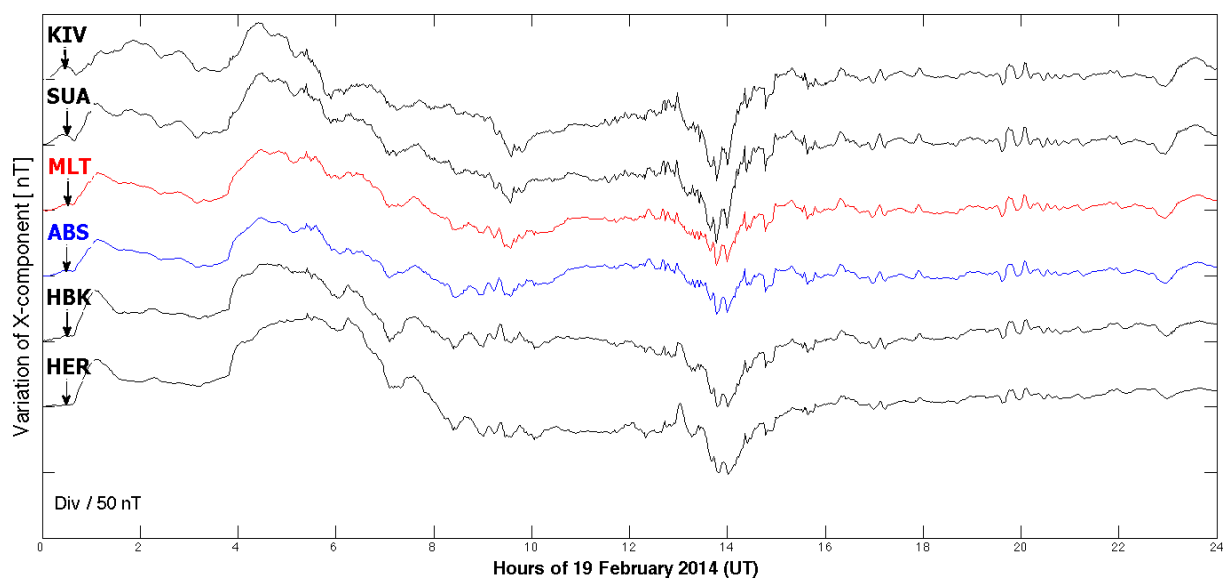


Figure 6. One-minute data of the X-component recorded at both the MLT (red curve) and ABS (blue curve) observatories compared with those recorded at some INTERMAGNET observatories (black curves) on 19 February 2014 (UT).

In addition, high-resolution data with sampling rate of 1-second is very important for several research fields such as the Ultra-Low Frequency (ULF) studies (geomagnetic pulsations). Geomagnetic pulsations are naturally occurring Magneto-Hydrodynamic waves in the Earth's magnetosphere. The occurrence and strength of these pulsations are mainly linked with the interplanetary magnetic field (IMF) and the solar wind speed (SW) as revealed from many geomagnetic observations (Greenstadt et al., 1979).

The Fig. 7 shows the total magnitude of the IMF, the solar wind speed, the spectrograms of the X-component recorded at the MLT, and the ABS respectively on 27 February 2014 (UT). The comparison between the spectrograms of the X-component at both the MLT and ABS observatories with the magnitude of IMF and solar wind speed aims to show the correlation between them during intense geomagnetic storms which indicates that source of these pulsations is an external one and linked with the solar parameters. Also, this figure indicates that the MLT and ABS observatories data (recorded before 2015) can be suitable for studies that required high-resolution data.

On the other hand, Fig. 8 shows the spectrograms (10-100 sec) at the MLT and ABS observatories on 25 February 2021 and comparing them with those for two INTERMAGNET observatories (KMH and HER) as reference points. It is clear that all spectrograms are showing a high correlation except the MLT observatory where its spectrogram reflects that the MLT data are very contaminated (very noisy in a wide frequency range).

To get a clearer idea, the data from these observatories were band pass filtered in the period range 20-30 sec (Pc3 range) then the obtained wave trains were compared to examine the harmony and the coherency between them as shown in Fig. 9. It is very clear that the wave train for the MLT data (red color in Fig. 9) is totally different

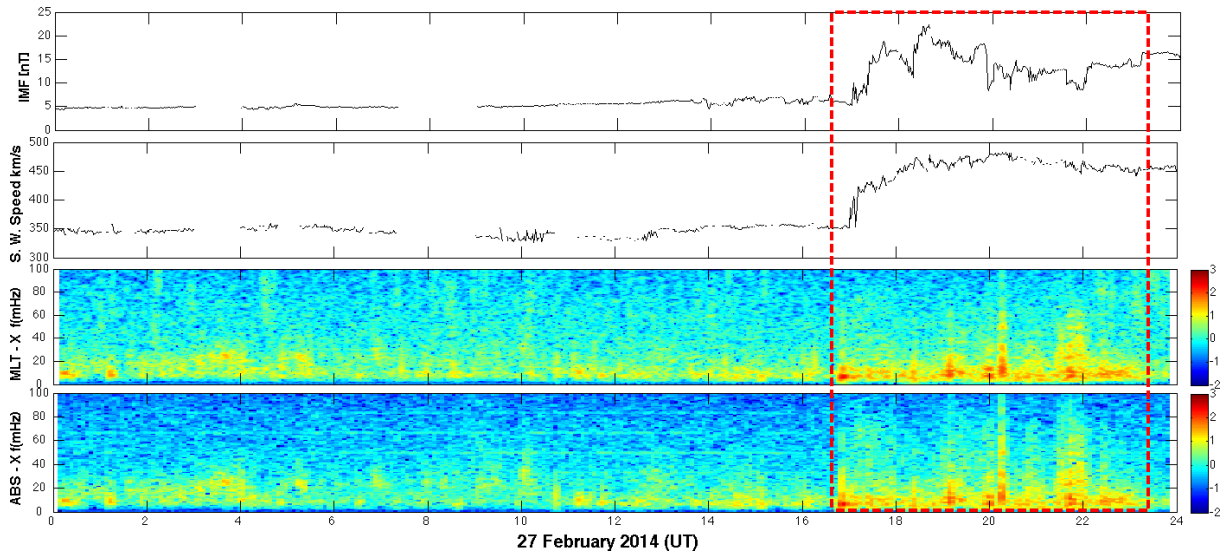


Figure 7. The panels from top to bottom show the total magnitude of IMF, the solar wind speed, the spectrogram of the X-component at the MLT, and the ABS respectively on 27 February 2014 (UT). The red rectangle indicates a good correlation between the pulsation activity observed at the MLT and ABS observatories with the IMF and SW fluctuations.

in shape and magnitude compared with those from the ABS and reference observatories. Thus, the MLT observatory data were contaminated by artificial noises that spoiled them (considering the natural small amplitude geomagnetic fluctuations as the ULF emissions). These artificial noises started gradually from 2015 until October 2022 (data are examined up to this date). The existence of such artificial noises for this very long period of time indicates that no QC strategy was applied for the MLT observatory data and here we can recognize the importance of data management and QC.

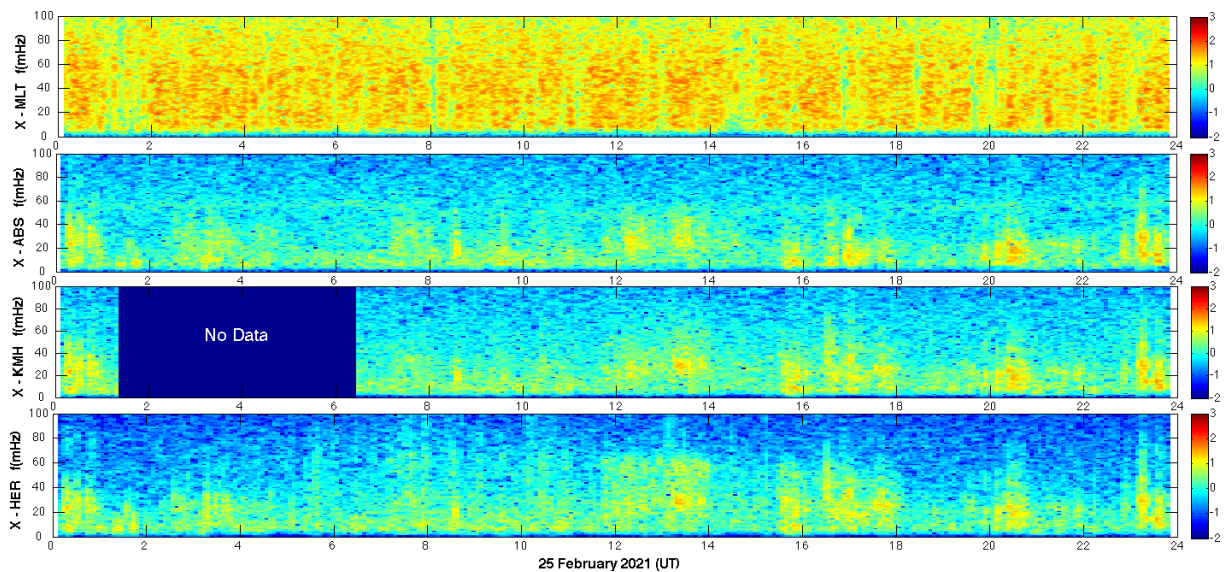


Figure 8. Spectrograms of the X-component at the MLT, ABS, KMH and HER GMOs on 25 February 2021(UT). It is clear from the upper panel that the MLT data is very noisy in a wide range of frequency. As a result, the natural ULF signals are hidden by this artificial noise. On the other hand, data from the other three observatories show good correlations.

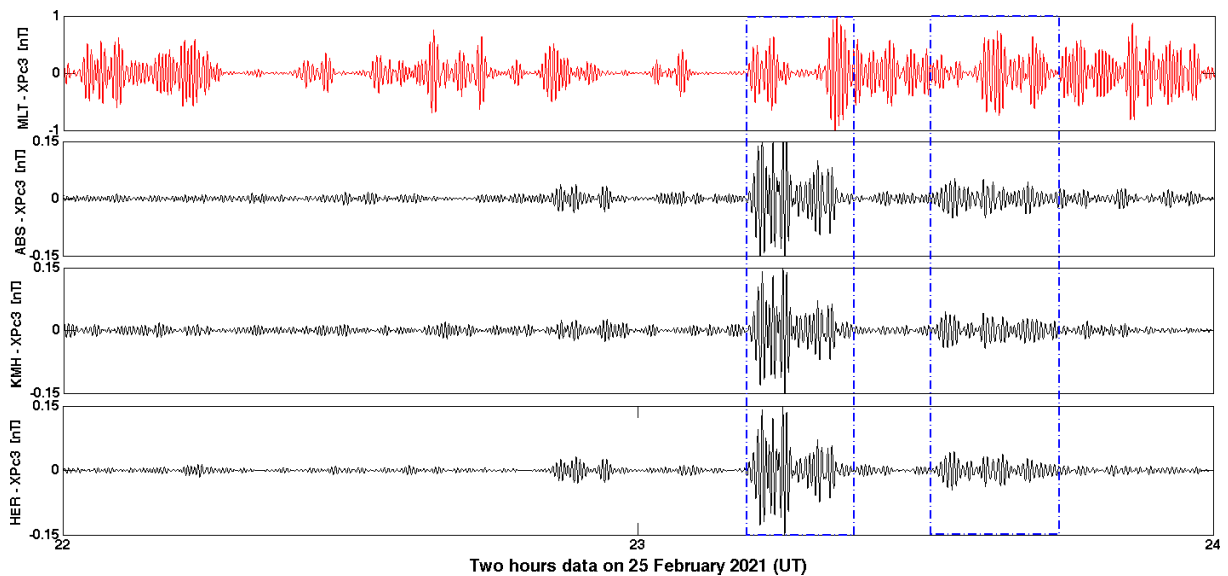


Figure 9. The comparison of the Pc3 of X-component recorded at the MLT, ABS, KMH and HER GMOs for two hours (23–24 UT) on 25 February 2021 (UT). The Pc3 wave train obtained from the MLT observatory data (red) shows no correlation with the Pc3 wave trains obtained from other observatories data (black).

4.4 Data format and metadata

The GMD should be published in an international DF so that researchers can use them easily in their research works. The most common GMD format is the IAGA2002 format. Formats could be developed for one-second, one-minute, one-hour, and so on. Metadata are an increasingly important aspect of GMD provision. Metadata are information about data sets. They describe the content, quality, originator, and other characteristics of a data set that help users understand the nature of that data set. As time passes, personnel, instrumentation, and data processing procedures may change, so it is important that current and future generations have sufficient information about the data to enable them to independently understand and use them. The issue of metadata was first introduced at the IAGA conference in Sopron (Reay et al., 2011). In addition, Reay et al. (2011) emphasized the importance of good Metadata provenance that helps in the scientific inspection and investigation of GMD sets.

In fact, one of the significant drawbacks of GMD recorded at both the MLT and ABS observatories is that the data files don't include a header portion of the data records. Generally, the header part ensures that those unfamiliar with the data set will have the best opportunity to be well informed about the nature of the data and any issues relating to their quality. Another drawback is that the nomenclature of data files doesn't refer to the observatory where they were recorded. Names of data files recorded at the MLT observatory (hourly data files) start with three letters refer to the month then followed by two digits refer to the day and after that two digits refer to the year. Finally, the extension of the file will be two digits referring to the hour at which these data was recorded. For example, the file name 'FEB1218.00' means that this file contains data recorded at the first hour of 12 February 2018 (UT). On the other hand, names of data files recorded at ABS observatory (daily data files) start with two digits refer to the year then followed by two digits refer to the month and ended with two digits refer to the day at which the data was recorded. For example, the file name '181105.txt' means that this file contains data recorded on 5 November 2018 (UT). This is certainly not an easy matter to resolve specially for the MLT data, but effort needs to be made to standardize the formats of GMD as much as possible in order to minimize the need for multiple data access software programs, minimize the potential for error in retrieving these data, and maximize the availability and ease of use. Thus, it was very important to convert original data text format into an international DF. Therefore, to maintain the quality level required by scientific research, GMD recorded at both the MLT and ABS observatories were converted to the IAGA2002 format by using some MATLAB scripts that we prepared for this aim. The original data files were converted from text format into IAGA2002 format (1-second and 1-minute data files). The nomenclature of the converted data files follows the rules provided by the IAGA2002 format. For example the name of converted data file starts with three letters refer to the observatory code, four digit represent the year, two digit represent the month,

two digit for the day of month, one letter for the type of data, three digits for data interval (mon-monthly, day-daily, hor-hourly, min-minute, and sec-second), after Manual (2012) and Turbitt et al. (2014). For example, the file name 'ABS20210520vsec.sec' means that this file contains one-second data from the ABS observatory recorded on 20 May 2021. A detailed example of original and converted data files from the MLT observatory is presented in Fig. 10.

It is worth mentioning that there are two basic principles that must be followed in data QC. Firstly, the natural geomagnetic variations cannot be changed or modified (Zhang et al., 2016). To identify the natural geomagnetic signals and also the quality of the processed data, the data must be compared with those of another GMO to ensure that the processed data still contain the natural characteristics of the recorded data especially the small fluctuations (geomagnetic pulsations). Secondly, a backup of raw data and processed data should be stored in different database storages.

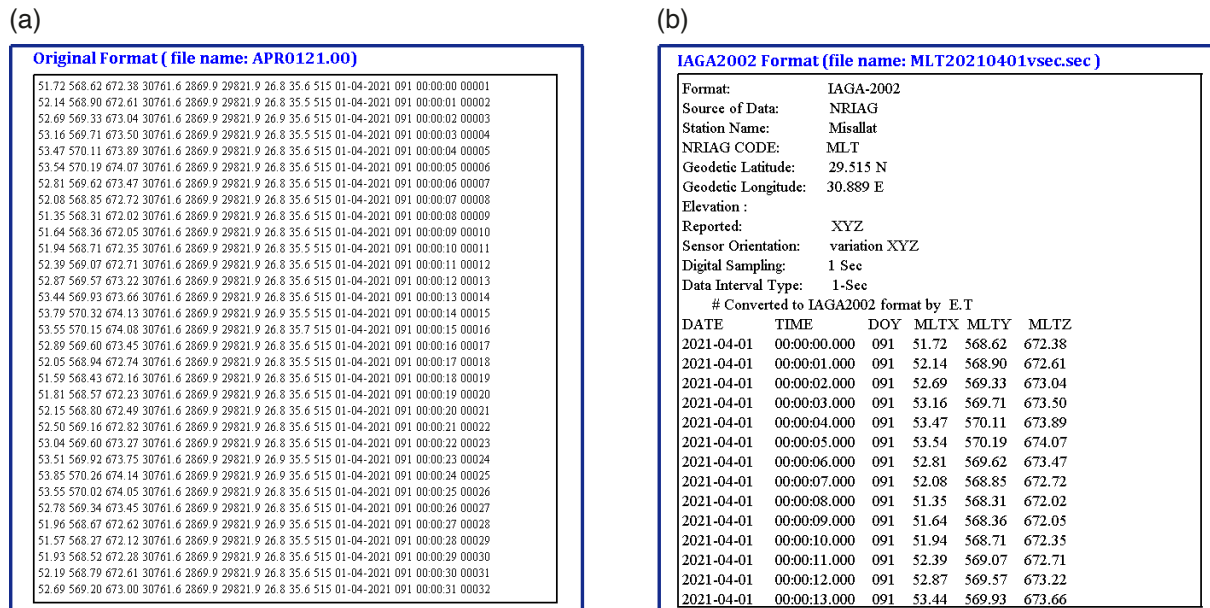


Figure 10. An example of the original data format (a) and the IAGA2020 data format (b) for one day at the MLT observatory.

5. Summary and conclusion

Valuable outcomes can be only achieved with precise and faultless data, so high quality observatory data are highly appreciated by the research community. Therefore, the current study presents a management and QC strategy including noises identification and elimination of their influence on the final data as well as converting the original DF into an international DF. The obtained results showed that the data recorded at the MLT and ABS observatories have different forms of artificial noises.

Since the best way to learn about the quality of data is to use them for scientific research, the processed data was compared with other data from some INTERMAGNET observatories for the examination of daily geomagnetic variations and ULF emissions. This comparison showed that the MLT and ABS data were in a good correlation with the reference INTERMAGNET data before 2015. On the other hand, the background noise especially at the MLT observatory was very high starting from 2015 till October 2022. These artificial noises completely hide the natural ULF signals at the MLT observatory. As some of the noise forms encountered at the MLT observatory are random fluctuations with the same frequency range of natural ULF emissions (0.002-5Hz), they are very hard or impossible to be cleaned, see Fig. 8 and 9. As a result, any trial to eliminate such noises will remove the natural ULF signals in the GMD.

It worth emphasizing that, the external artificial noise is extremely disturbing for the successful operation of any GMO. The main recommendation from this research work is that we should keep the protection and preservation of the MCE within its legislative framework at the two GMOs, particularly at the MLT observatory in order to obtain high quality data.

Acknowledgments. The Misallat and Abu-Simbel geomagnetic field data were provided by the Geomagnetism Laboratory, NRIAG (<http://www.nriag.sci.eg/>), while the solar parameters data were downloaded from the OMNI Website (<https://omniweb.gsfc.nasa.gov/>). The results presented in this paper rely on data collected at geomagnetic observatories, so the authors thank the national institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice (www.intermagnet.org).

References

- Chow, T. W. S. and D. Huang (2008). Data Reduction for Pattern Recognition and Data Analysis, in Computational Intelligence: A Compendium, Studies in Computational Intelligence, J. Fulcher and L. C. Jain (Eds.), 115, Springer, Berlin, Heidelberg, doi:10.1007/978-3-540-78293-3_2.
- Curto, J. J., S. Marsal, J. M. Torta and E. Sanclement (2009). Removing spikes from magnetic disturbances caused by trains at Ebro Observatory, in Proceedings of XIII IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition and Processing, J. J. Love (Ed.), U.S. Geological Survey Open File Report 2009-1226, U.S. Geological Survey, Reston, VA, USA, 2009, 60-66.
- Greenstadt, W., H. J. Singer, C. T. Russel and J. V. Olson (1979). IMF orientation, solar wind velocity and Pc3-4 signals, A joint distribution, *J. Geophys. Res.*, 84, 527-532, doi:10.1029/JA084iA02p00527.
- Jankowski, J. and C. Sucksdorff (1996). Guide for magnetic measurements and observatory practice, Warsaw, 235, ISBN: 0-9650686-2-5.
- Khomutov, S., O. Mandrikova, E. Budilova, A. Kusumita et al. (2017). Noise in raw data from magnetic observatories, *Geoscientific Instrumentation, Methods and Data Systems*, 6, 329-343, doi:10.5194/gi-6-329-2017, 2017.
- Khomutov, S., V. Sapunov, A. Denisov, D. Savelyev et al. (2016). Overhauser vector magnetometer POS-4: Results of continuous measurements during 2015-2016 at geophysical observatory Paratunka of IKIR FEB RAS, Kamchatka, Russia, *E3S Web Conf.*, 11, doi:10.1051/e3sconf/20161100007.
- Linthe, H. J., J. Reda, A. Isac, J. Matzka et al. (2012). Observatory Data Quality Control the instrument to ensure valuable research, in XVth IAGA Workshop on Geomagnetic Observatory Instruments and Data Processing, Cadiz, Spain, 173-177.
- Love, J. J. (2009). Missing data and the accuracy of magnetic observatory hour means, *Ann. Geophys.*, 27, 3601-3610, doi:10.5194/angeo-27-3601-2009.
- Lowes, F. J. (2009). DC railways and the magnetic fields they produce the geomagnetic context, *Earth Planets Space*, 61, i-xv, doi:10.1186/BF03352944.
- Macmillan, S. (2007). Observatories: an overview, in *Encyclopedia of Geomagnetism and Paleomagnetism*, D. Gubbins and E. Herrero-Bervera (Eds.), Netherlands, Springer, 708-711, doi:10.1007/978-1-4020-4423-6_224.
- Neska, A., J. Reda, M. Neska and Y. Sumaruk (2012). On the influence of DC railway noise on variation data from Belsk and Lviv geomagnetic observatories, *Acta Geophys.*, 61, doi:10.2478/s11600-012-0058-0.
- Pirjola, R., L. Newitt, D. Boteler, L. Trichtchenko et al. (2007). Modelling the disturbance caused by a dc-electrified railway to geomagnetic measurements, *Earth Planets Space*, 59, 943-949, doi:10.1186/BF03352033.
- Reay, S. J., D. C. Herzog, S. Alex, E. Kharin et al. (2011). Magnetic Observatory Data and Metadata: Types and Availability, in *Geomagnetic Observations and Models*, M. Manda and M. Korte (Eds), IAGA Special Sopron Book Series, 5, 149-181, doi:10.1007/978-90-481-9858-0_7.
- Zhang, S., F. Changhua, H. Yufei, Y. Dongmei et al. (2016). Quality Control of Observation Data by the Geomagnetic Network of China, *Data Sci. J.*, 15, doi:10.5334/dsj-2016-015.

***CORRESPONDING AUTHOR: Emad M. H. TAKLA,**

National Research Institute of Astronomy and Geophysics (NRIAG),
11421 Helwan, Cairo, Egypt
e-mail: emad.takla@nriag.sci.eg