

AN INVESTIGATION INTO TECHNIQUES FOR ISOLATING NOISE IN OBSERVATORY DATA

C. Turbitt, O. Baillie, D. Kerridge, and E. Clarke*

British Geological Survey, West Mains Road, Edinburgh EH9 3LA, United Kingdom

** Email: cwtu@bgs.ac.uk*

ABSTRACT

In this study, using one-minute definitive data published by a number of INTERMAGNET observatories, we apply a number of time- and frequency-domain techniques to characterise the global, natural geomagnetic signal and isolate the artificial noise at an individual observatory. With the aim of developing an analytical tool that can be used to identify observatory noise against the natural signal, we report on the suitability of these techniques to detect common observatory noise types.

1 INTRODUCTION

The demand for rapid access to high-quality, high-time-resolution data from magnetic observatories is becoming more and more apparent (e.g., space weather applications and internal field studies in combination with data from the upcoming SWARM mission). Observatory operators are meticulous in their efforts to ensure that their observatories deliver such data and apply various quality control procedures to achieve this. However, ensuring that observatory data accurately represent natural field variations in the long-term is an increasing challenge.

Firstly, a number of long-running observatories were built in magnetically quiet locations, but urban expansion is now bringing sources of man-made disturbance within their range of detection. Secondly, the ability to resolve low-amplitude higher frequency natural signals, which is the intention in producing one-second data, may be compromised by artificial noise.

The limiting factor on any digital measurement is noise, which can come from a variety of sources and be measured by a number of methods. Every observatory will have background man-made noise to some degree, including localised disturbance, instrument noise, and noise introduced during data processing. Components of the natural signal (e.g., localised current systems resulting from induction effects) vary in amplitude from observatory to observatory and may be considered as noise for studies or applications based on one part of the signal, such as representation of the main field for general navigation.

Some methods for noise detection are routinely employed by observatories. Where an observatory operates more than one magnetometer, comparisons of data between instruments are used to detect instrument noise, processing errors, and localised disturbance. Comparisons between two or more instruments can be used to identify which instrument is affected. However other noise sources, such as non-localised disturbance and systematic processing errors, are not effectively detected by such comparisons.

Another useful method for detecting noise is to compare data from two nearby observatories, and such a tool is provided on the INTERMAGNET CD viewing software. Inter-observatory comparisons are particularly useful for detecting spikes, steps, etc. in a time series but rely on an observatory being available nearby that is known to be of good quality. Given that even the closest of observatories are typically separated by hundreds of kilometres, the difference in the natural signal between observatories is often sufficient to dominate comparison plots such that low amplitude noise is difficult to detect. This is particularly true for higher latitude observatories.

2 FIRST DIFFERENCE OBSERVATORY COMPARISONS

In a previous study, Love (2006) investigated whether noise at a particular observatory could be isolated by looking at a global set of observatory data. This study supposed that observatories at a similar geomagnetic latitude have natural signals of broadly similar characteristics (although the time series may have noticeable differences). The signal amplitude at an observatory on a particular day was characterised by the standard deviation (SD) of the first differences of the one-minute time series. Following Love (2006), observatory amplitudes were compared by plotting the SD values against geomagnetic latitude, as shown in Figure 1 for a selected quiet day in 2004.

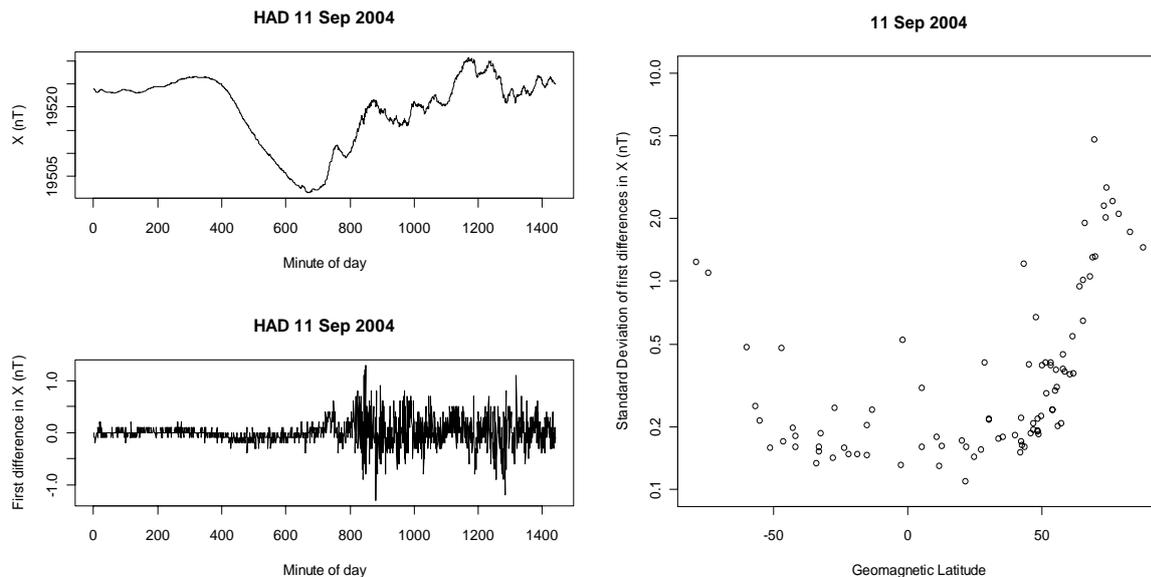


Figure 1. HAD X-component magnetogram (top left), first difference plot of the magnetogram (bottom left), and SD of first differences against geomagnetic latitude for a set of 97 INTERMAGNET observatories (right).

As expected, the SD of the first differences is lowest for low-latitude observatories although there are some anomalies, such as the equatorial observatory Huancayo, which has high amplitude due to the enhanced daily variation signal resulting from the equatorial electrojet. The observatory with the lowest amplitude is Honolulu, and this is assumed to be due to the attenuation of the external field variations caused by the surrounding deep ocean. Otherwise, it is generally possible to compare observatories by how well they sit on this W-shaped curve, giving a technique with which to identify an observatory exhibiting higher than average levels of noise at its geomagnetic latitude.

First differences have been used in this method to remove DC and harmonics of the daily variation, which would otherwise dominate the signal amplitude representation. The effect of taking first differences is to apply a high-pass filter that, for one-minute data, has a -3dB point at a period of 4 minutes and a roll-off of 8 dB/octave at 10 minutes. This poor roll-off and fixed-frequency cut-off point means that a first-difference filter is ineffective at both isolating noise at a particular frequency and is also limited in attenuating the low-frequency, large amplitude harmonics of the daily variation.

3 FREQUENCY-DOMAIN OBSERVATORY COMPARISONS

A more effective method to isolate noise from natural signal in observatory data would be to examine data in the frequency-domain rather than the time-domain.

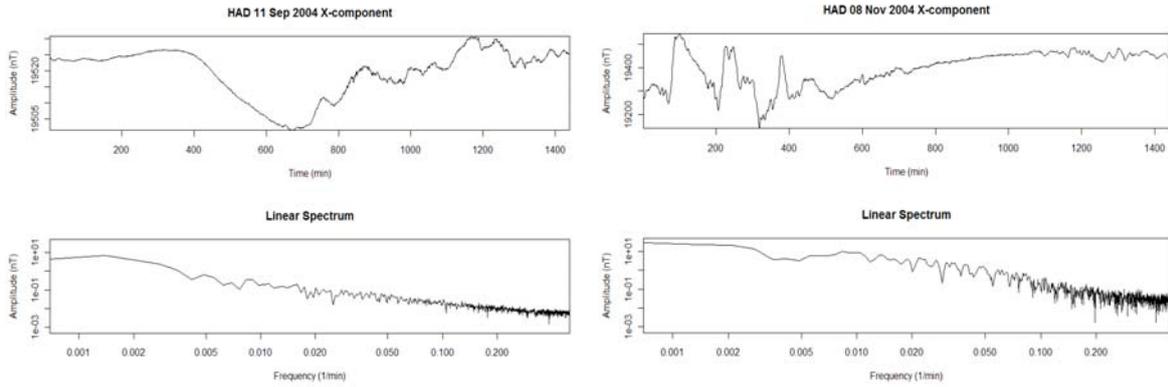


Figure 2. Time-series (top) and Fourier transform (bottom) of a quiet day (left) and active day (right) at HAD

Figure 2 shows data from Hartland (HAD) Observatory during two days in 2004, along with the linear spectra of the time-series. It is evident from both the quiet day and the active day that the amplitude of the natural signal diminishes smoothly with increasing frequency, with the amplitude higher at all frequencies on the active day.

To obtain a single value representative of the signal amplitude from the Fourier transform, the time-domain signal is Hanning windowed to minimise spectral leakage prior to a fast Fourier transform (FFT). The FFT is then windowed to attenuate all frequencies outside the band of interest. The result is scaled and the power averaged over the band to produce a mean linear spectral density.

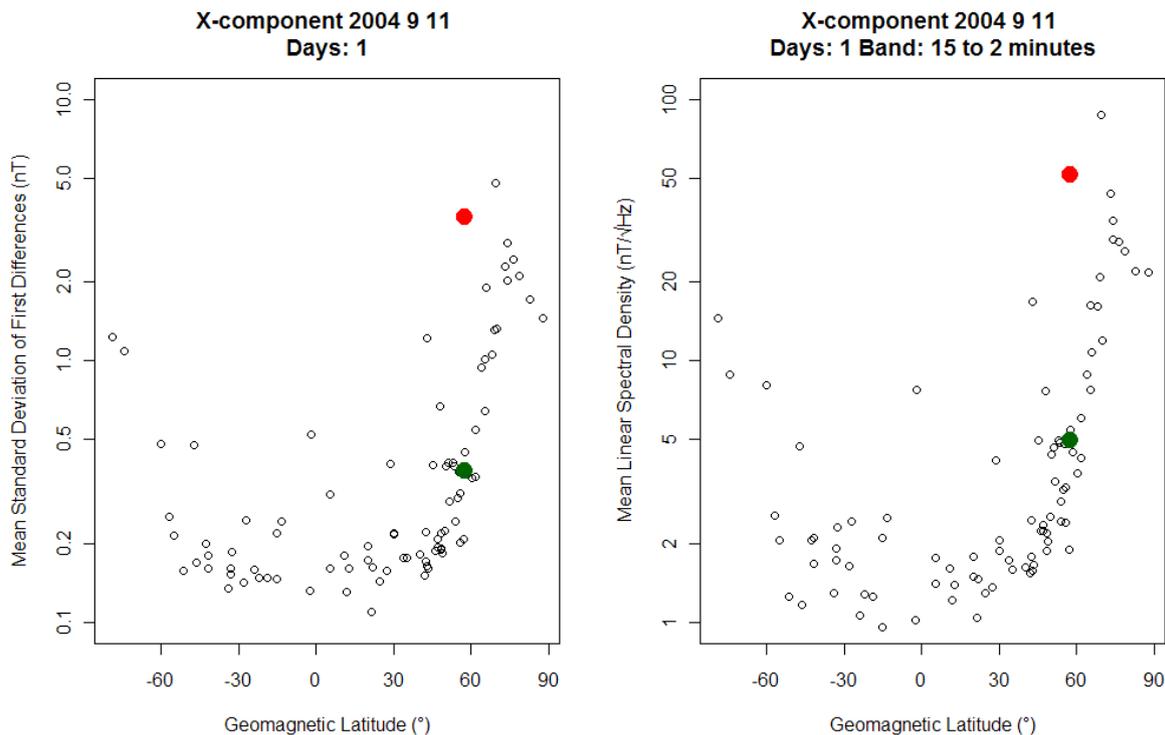


Figure 3. The SD first difference method (left) and frequency-domain method over a limited band of 15 to 2 minutes (right) for 97 INTERMAGNET observatories on a quiet day. Observatory ESK is highlighted (green) and also with artificial noise of 5nT, 6-minute period added (red).

As would be expected, where the pass-band for this frequency-domain technique is set to approximate the pass-band of the first-difference filter, the results of the two techniques are comparable as shown in Figure 3. This is made clear in the figure with the addition of artificial noise to Eskdalemuir (ESK) Observatory data. The

separation of the clean (green) and noisy (red) data are equally distinguishable above the W-shaped curve by both the SD first difference method and the frequency-domain method.

Where the frequency-domain method is of benefit is in being able to select the frequency band by modifying the linear spectral density window to a band of interest. Figure 4 is an example where the artificial noise on the ESK Observatory data has been set to a lower frequency, such that the SD first difference method is unable to distinguish the noise from the natural signal. The band of the frequency-domain method has been extended to 120-minutes, increasing the natural signal power as a result; however, the low-frequency noise is clearly evident above the W-shaped curve.

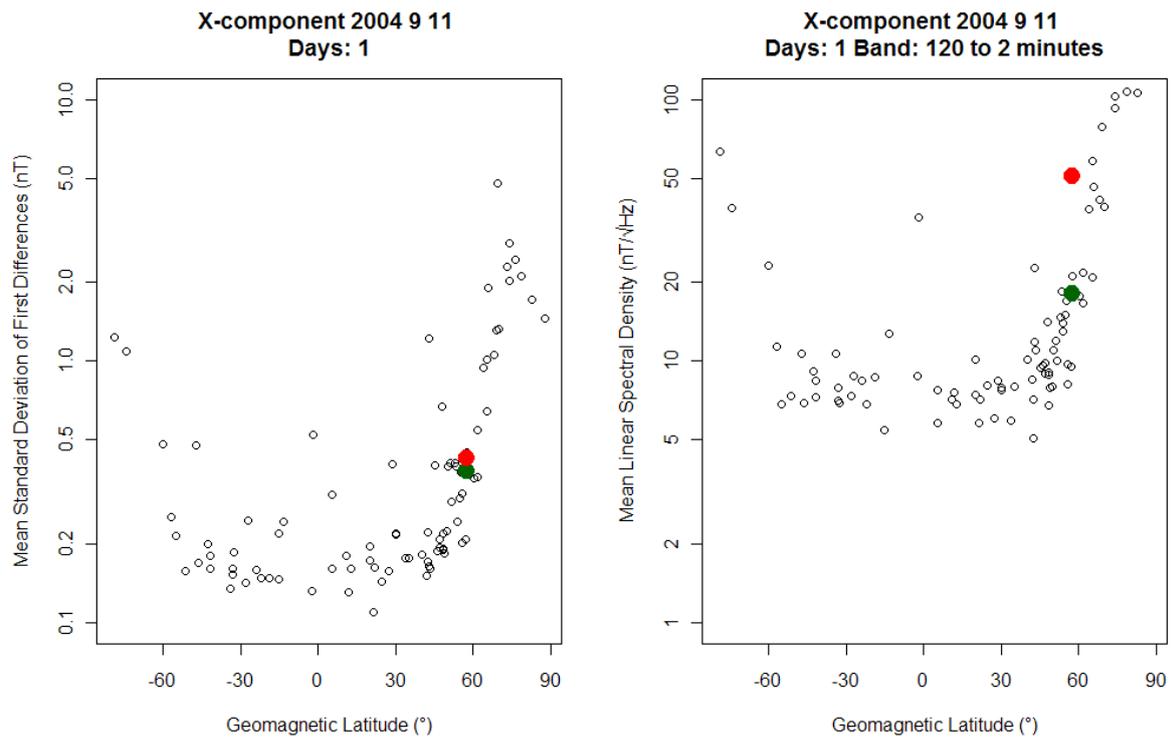


Figure 4. The SD first difference method (left) and frequency-domain method over a limited band of 120 to 2 minutes (right). Observatory ESK is highlighted (green) and also with artificial noise of 5nT, 120-minute period added (red).

Figure 5 shows a similar result for high frequencies. In this case, the input noise is of lower amplitude (0.5nT) but at a period of 3 minutes. Due to the low amplitude, the noisy signal for ESK in the SD first difference plot is not well differentiated from the W-shaped curve representing the natural signal amplitude. However, with the frequency-domain method, the band has been limited to 3.1-2.9 minutes, which lowers the amplitude of the natural signal and as a consequence allows the input noise to become detectable.

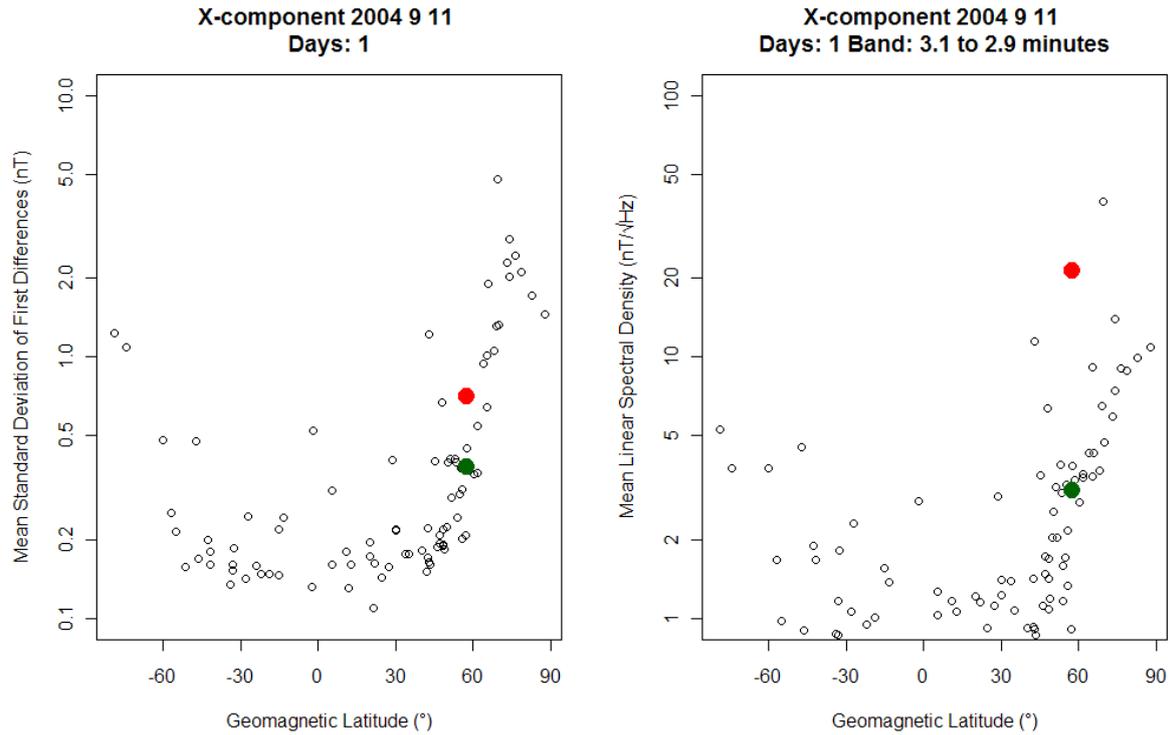


Figure 5. The SD first difference method (left) and Fourier transform method over a limited band (right). Observatory ESK is highlighted (green) and also with artificial noise of 0.5nT, 3-minute period added (red).

Since the natural signal amplitude shown in Figure 5 has a dependence on geomagnetic latitude, the W-shaped curve can be modelled and removed to improve the noise differentiation. As an example, the data in the right-hand plot in Figure 5 has been modelled using a spline function and the residuals plotted in Figure 6. The chosen spline function is subjective; however, the figure shows that the correlation between the amplitude of the natural signal and the geomagnetic latitude is sufficient that it can be well represented by that function. Any additional signal (noise) is readily identified from it, suggesting that this method has the potential to be employed in an automated data quality monitoring system.

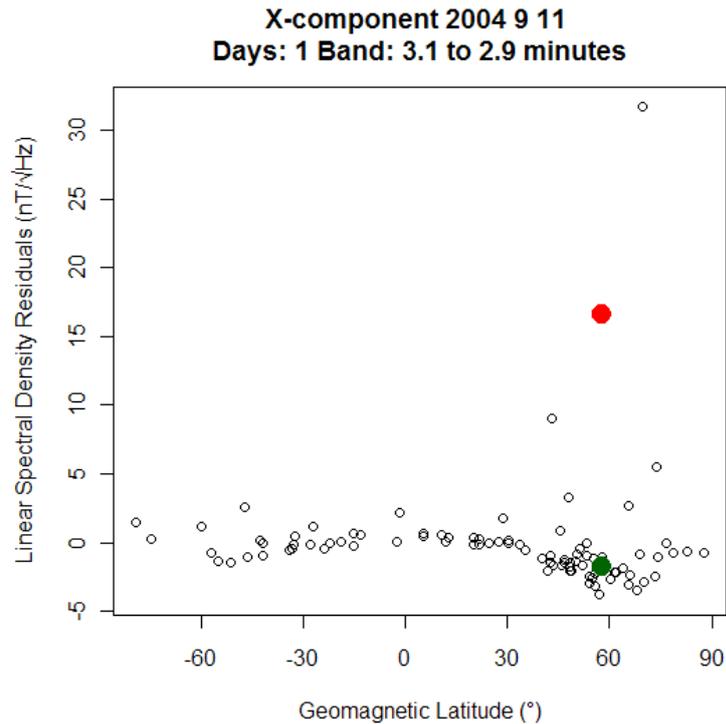


Figure 6. The Fourier transform method over a limited band. Observatory ESK is highlighted (green) and also with artificial noise of 0.5nT, 3-minute period added (red).

4 APPLICATION OF FREQUENCY-DOMAIN OBSERVATORY COMPARISONS

In previous examples, noise has been added artificially to investigate the limits to which the use of frequency-domain analysis can be used to compare observatories in order to identify non-natural signal. Here, we show how the technique can be employed to investigate existing noise in observatory data. The daily magnetogram of an INTERMAGNET observatory (nominally OBS) contains no discernable noise even on a quiet day (Figure 7), and the spectral density indicates that the amplitude of the signal diminishes smoothly with increasing frequency in the same way as HAD data in Figure 2. Hence, the time-series and the frequency spectrum in isolation are insufficient to identify any noise.

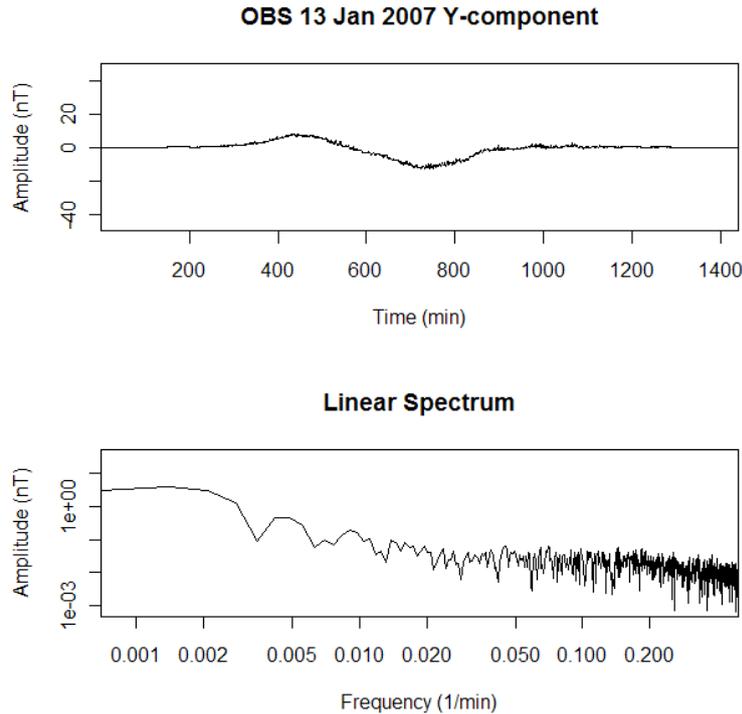


Figure 7. Time series and linear spectrum for an INTERMAGNET observatory (OBS) on a quiet day in 2007.

Using the frequency-domain technique with residuals to compare the signal amplitude of observatory OBS in the band 120-minutes to 10-minutes (Figure 8, top-left) again does not show any discernable signal amplitude over the natural signal as measured by observatories of a similar latitude since the natural signal amplitude is relatively large. However, in the higher frequency band of 10-minutes to 2-minutes (Figure 8, top-right), where the natural signal amplitude is lower, the signal amplitude of OBS is distinctively higher than comparable observatories.

The lower two plots in Figure 8 split the high frequency band further in an attempt to identify the frequency of the noise source, but the signal amplitude can be seen to remain relatively constant at all frequencies. Therefore, it can be concluded that in this real example, the observatory data contain noise spread evenly across all measured frequencies, i.e., analogous to white noise. The level of the noise is not such that the observatory fails to meet the INTERMAGNET one-minute specifications, as can be seen in the time-series in Figure 7, but the analysis suggests that the observatory has a problem with artificial noise that is maybe a concern for reported one-second data.

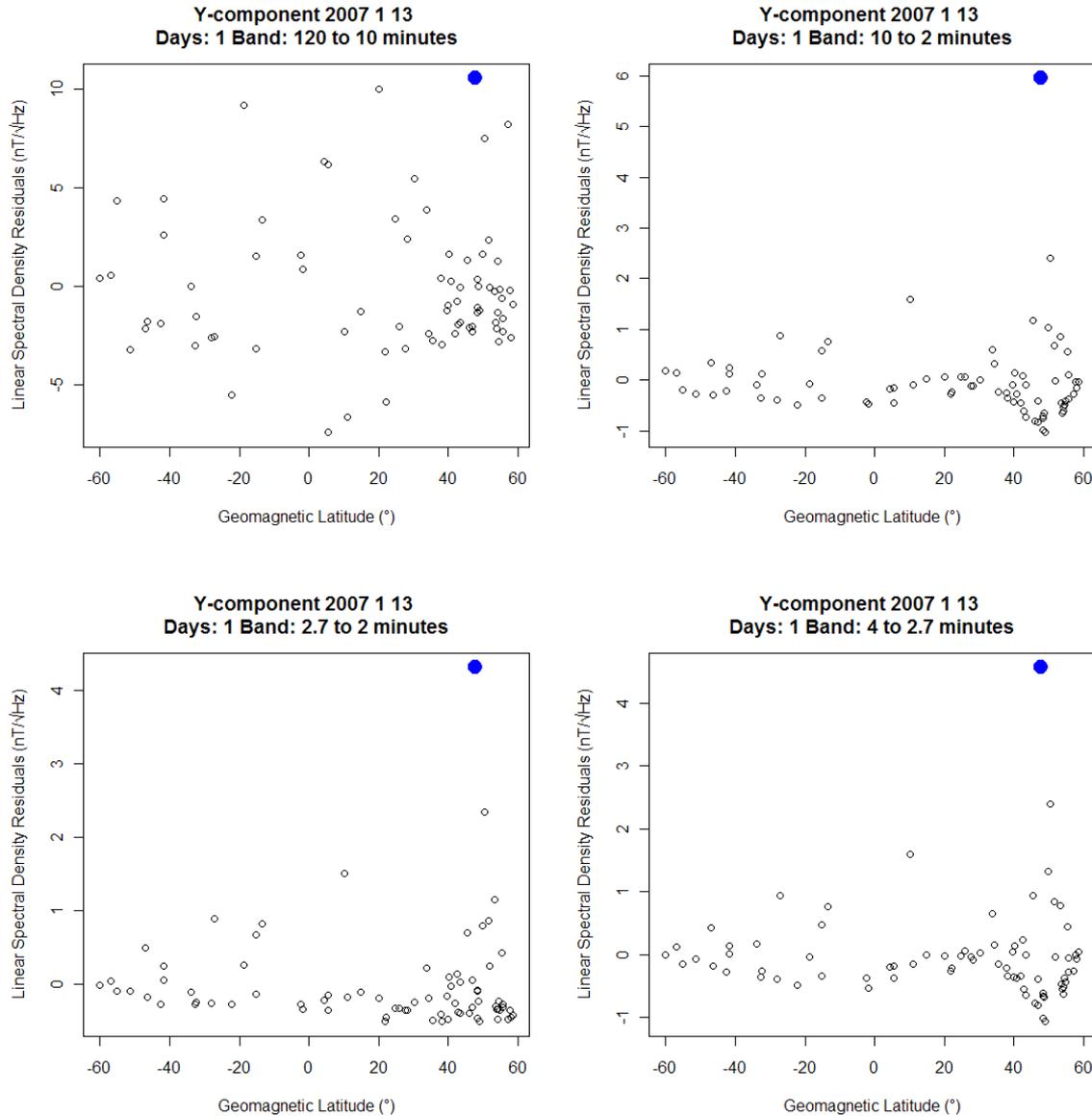


Figure 8. Residual plots of linear spectra for all INTERMAGNET observatories on a quiet day in 2007 at selected frequency bands. The highlighted observatory (blue) evidently has a higher amplitude signal than observatories at similar geomagnetic latitude.

5 CONCLUSION

With increasing urbanisation, geomagnetic observatories around the world are subject to rising levels of artificial noise from sources such as transportation, communication, and power distribution. Conversely, science is demanding that there be better coverage of geomagnetic observatories and that those observatories be capable of measuring higher frequency components of the natural signal with better resolution. Observers consistently employ best practice to ensure recorded data accurately represent natural variations and that artificial signal is minimised. This study has shown that existing analytical methods can be improved upon to identify artificial signal of relatively low amplitude. The application of the frequency-domain technique to a selected INTERMAGNET observatory shows that this technique is capable of identifying an artificial noise signal where other techniques failed to do so and the technique can also be used, to an extent, to describe the nature of that noise signal by isolating the dominant frequency.

The frequency-domain technique has advantages over conventional techniques, such as instrument or two-observatory time-domain comparisons; it is more sensitive to localised noise and, by making use of a global data set, is more effective in removing natural signal. The SD first difference method also uses a large data set but has limited frequency resolution as a result of the poor response of the first difference filter compared to the frequency-domain method. The synthetic noise test performed on the frequency-domain technique shows that it can be used to detect noise at lower frequencies and also, by limiting the band, detect low amplitude noise at high frequencies.

The frequency-domain technique, however, is limited to detecting periodic noise of amplitude greater than the natural signal amplitude. Hence this technique is insensitive to transient errors in the data, such as spikes or steps, and, since it uses comparisons of observatory signal amplitudes rather than phase, problems such as timing errors will not be detected. Thus, data errors such as spikes, steps, and timing errors continue to be best detected in the time domain using instrument and/or inter-observatory comparisons. All of the examples in this study used data from selected quiet days since this technique relies on low natural signal amplitude to differentiate the natural signal from noise. Another consequence of the reliance on low amplitude natural signal is that the technique becomes less effective with increasing latitude as demonstrated by the W-shaped curve in Figure 3 and is most effective on quiet days for observatories of geomagnetic latitude between -60° and $+60^{\circ}$.

For brevity, this study has only made use of one-minute data over periods of one day, but the technique is readily adapted to investigating lower frequency signals over longer time periods and also to higher frequency data such as the proposed one-second data standard.

Since the principle has been established, a further aim of the study will be to develop an analytical tool that can be incorporated into the British Geological Survey's daily processing routine to identify, investigate, and minimise artificial noise, thus improving data quality in anticipation of higher frequency data products. The technique as described here is not immediately suited to automatic processing, requiring judgements to be made over the band of interest and the modelling function. Additional work is required to identify a function that reliably models the W-shaped curve at all activity levels such that the residuals can be automatically generated. Residuals would then most effectively be presented in the form of a spectrogram where observatories with unusually high signal amplitudes across one or more frequency bands can be clearly identified.

In developing the tools for this analysis, the authors have constructed a library of software functions in R for reading INTERMAGNET data, performing frequency-domain analysis, and plotting the results. These can be made available to other institutes wishing to conduct similar noise evaluations.

6 ACKNOWLEDGEMENTS

The results presented in this paper rely on data collected at magnetic observatories. We thank the national institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice (www.intermagnet.org).

This study made use of the freeware mathematical & plotting library provided by the R Project for Statistical Computing (www.r-project.org).

This paper is published with the permission of the director of the British Geological Survey (Natural Environment Research Council).

7 REFERENCES

Love, J.J. (2006) Examination of Intermagnet Observatory Noise. *INTERMAGNET Meeting Minutes, Polish Academy of Sciences, Warsaw26 – 28 June 2006*, Warsaw, Poland.