The strategy to calculate impedance using RMHHT estimator for continuous and discontinuous broadband magnetotelluric time series data

Hao Chen (chenhaomagnetic@gmail.com)
Kyushu University https://orcid.org/0000-0002-1944-2605
Hideki Mizunaga
Kyushu University - Ito Campus: Kyushu Daigaku
Toshiaki Toshiaki Tanaka
Kyushu University - Ito Campus: Kyushu Daigaku
Gang Wang
Energy and Deep Earth Exploration Laboratory

Research Article

Keywords: impedance, magnetotelluric, Hilbert-Huang transform

Posted Date: February 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-233183/v1

The strategy to calculate impedance using RMHHT estimator for continuous and discontinuous broadband magnetotelluric time-series data

Hao Chen[1)], Hideki Mizunaga[2)], Toshiaki Tanaka[2)] and Gang Wang[3)]

1) Department of Earth Resources Engineering, Graduate School of Engineering, Kyushu University, Fukuoka 819-0395, Japan

2) Department of Earth Resources Engineering, Faculty of Engineering, Kyushu University, Fukuoka 819-0395, Japan

3) Energy and Deep Earth Exploration Laboratory, Institute of Geophysical and Geochemical Exploration, China Geological Survey, Langfang 065000, China

Hao Chen: chenhaomagnetic@gmail.com

Hideki Mizunaga: mizunaga@mine.kyushu-u.ac.jp

Toshiaki Tanaka: tanaka@mine.kyushu-u.ac.jp

Gang Wang: jiaowulou214@163.com

Abstract: MT method is widely used in exploration surveys worldwide but hardly applicable in urban areas because of the large amount of artificial electromagnetic noise. The MT time-series data at night is much quieter than in the daytime because most of the electric equipments are shut down for several hours during nighttime. Therefore, we focused on calculating the MT impedance using the quiet time-series data. In this research, the data observed by the Phoenix System was used. The data contain continuous and discontinuous time-series data. As an alternative way, we introduced a robust impedance estimator based on the Hilbert-Huang transform (RMHHT). We indicated that this technique needs 4-hour data to get a reliable impedance up to 1,000-second in the numerical simulation. This short measurement time makes it possible to carry out MT surveys in urban areas with strong noises. This paper demonstrated a strategy to process the broadband MT time-series data properly by the RMHHT estimator. Finally,

a successful case study, using the nighttime data to get reliable impedance tensor in the areas contaminated by heavy noises, was demonstrated.

Keywords: impedance; magnetotelluric; Hilbert-Huang transform

1 INTRODUCTION

The Magnetotelluric (MT) method is an electromagnetic geophysical method for inferring the earth's subsurface electrical conductivity from measurements of natural geomagnetic and geoelectric field variation at the earth's surface. MT method is widely used in exploration surveys around the world. Still, along with urban constructions, artificial disturbances to the electromagnetic observations are becoming more serious. MT exploration is hardly applicable in urban areas with a large amount of artificial electromagnetic noise.

The field data consists of natural sources and local noise. Szarka(1988) and Junge(1996) summarized active and passive noise sources observed in MT measurements. When the data segment contains heavy noise, it is difficult to differentiate between the MT signals coming from the natural sources and the noise signals produced by the local environment (e.g., electric transmission lines, electric fences, trains). If most of the time-series data contain strong noise, any impedance estimator will fail to get reliable results. But in one day, there might be several hours' time window at nighttime that most of the electric equipment is shutdown. The time-series data in the period is much quieter than in the daytime. In this paper, we consider using the quiet period data only to calculate the impedance tensor.

The first step in MT data processing is to estimate the frequency domain's impedance tensor from the measured time-series data. In conventional MT data processing routines, the original time-series is divided into short segments and then transforms the segments into the frequency domain using the Fourier Transform. Here we should know one segment is one dataset when doing the impedance estimation in the frequency domain. MT time-series data occasionally contain transient noise, which is not suitable for the basic requirements of conventional methods

based on the Fourier transform. In the Fourier transform, a part of the time-series data contaminated by transient noise will influence the whole spectrum.

The purpose of impedance estimation is to get the impedance tensor that depends on frequency. We also can transform the time-series into the frequency domain data by the time-frequency transform technique like HHT. Huang et al. (1998) introduced empirical mode decomposition (EMD) in the framework of the Hilbert-Huang Transform (HHT). This novel time-series analysis tool is data-adaptive and suitable for nonlinear and nonstationary data. The time-series data is decomposed into intrinsic mode functions (IMFs) that can be represented as a mono-component, containing a single frequency. Huang et al. (1998) argued that the IMF allowed for a meaningful computation of its instantaneous parameter (IP) with the Hilbert Transform. Chen et al. (2012) used the IPs for impedance estimation. Neukirch and Garcia (2013) firstly proved that IPs could be used directly for impedance estimation in theory. Neukirch et al. (2014) also showed that when the signal source is nonstationary (NS), the method based on the HHT performs better than the traditional method based on windowed FFT. The estimator based on Hilbert-Huang Transform can estimate MT response functions even in NS signal and transient noise.

In this paper, the data sets observed by the MTU-A Geophysical Instruments was used. The time-series data are stored in three files. Two of them store the high and middle-frequency bands (2,400 and 150 Hz) discontinuously. The other file stores the low-frequency data (15 Hz) continuously. As an alternative, we introduced a robust impedance estimator based on the Hilbert-Huang transform (RMHHT) and demonstrate a proper strategy based on RMHHT to calculate the full-scale broadband MT time-series data, including the continuous and discontinuous data. The property of the HHT, which could express the spectrum variation in the time-frequency domain well, makes it possible to directly connect the discontinuous time-series and process the combined discontinuous time-series in the same way as the continuous time-series. The biggest problem of the estimator based on HHT is time-consuming. By this strategy, the computation time can be shortened much.

In the synthetic data test, we have shown that there is continuous 4-hour noise-free time-series data. We could get a reliable impedance until 1,000-second. This strategy is suitable to get a reliable result by selecting the data in the time domain. That makes it possible to use the nighttime data only to get a reliable impedance.

Finally, we compared the performance of RMHHT with the conventional method Bounded Influence Remote Reference Processing (Chave and Thomson, 2004) and the SSMT 2000 program and showed the successful case study using the nighttime data to get a reliable result from the areas contaminated by heavy noisy.

2 METHOD

In this section, three different impedance tensor estimators, BIRRP, SSMT-2000, and RMHHT, are introduced to compare them using the field data in subsection 4.3.

2.1 BIRRP

Bounded Influence Remote Reference Processing (Chave and Thomson, 2004) is a typical conventional robust estimator to calculate the impedance tensor based on windowed FFT.

The first estimator (least-squares estimator, Sims et al., 1971) divides the original time-series into the short segments and then transforms the segments into the frequency domain using Fourier Transform. Finally, the impedance tensor at a specific frequency can be calculated as follow:

$$\begin{pmatrix} E_x(\omega) \\ E_y(\omega) \end{pmatrix} = \begin{pmatrix} Z_{xx}(\omega)\ Z_{xy}(\omega) \\ Z_{yx}(\omega)\ Z_{yy}(\omega) \end{pmatrix} \begin{pmatrix} H_x(\omega) \\ H_y(\omega) \end{pmatrix}. \quad (1)$$

The objective is to minimize the squares residual sum by the least-squares theory, and the solution of impedance tensor at a specific frequency is given by:

$$\mathbf{Z}=(\mathbf{H}^{\dagger}\mathbf{H})^{-1}(\mathbf{H}^{\dagger}\mathbf{E}) , \quad (2)$$

where **E** and **H** are the horizontal electric and magnetic field components at a specific frequency, and **Z** is the MT impedance tensor. The superscript † denotes the complex conjugate transpose.

The M-estimator (Egbert, 1986) gives a weight w based Huber weight function to the outlier depending on the residual between the output (electric field) of the least-square estimator and the observation data. The solution of M-estimator at a specific frequency is given by:

$$\mathbf{Z}=(\mathbf{H}^{\dagger}\mathbf{w}\mathbf{H})^{-1}(\mathbf{H}^{\dagger}\mathbf{w}\mathbf{E}). \quad (3)$$

M-estimator can reduce the influence of unusual data (outliers) in the electric field but are not sensitive to exceptional input (magnetic field) data, which are termed leverage points. The hat matrix is widely used to detect unusual input data. The bounded influence (BI) estimator combines the robust M-estimator with leverage weight (**v**) based on the statistics of the hat matrix diagonal element (Chave et al., 2004), and the solution of BI-estimator at a specific frequency is given by:

$$\mathbf{Z}=(\mathbf{H}^{\dagger}\mathbf{w}\mathbf{v}\mathbf{H})^{-1}(\mathbf{H}^{\dagger}\,\mathbf{w}\mathbf{v}\mathbf{E}). \quad (4)$$

The detail of the M-estimator and BI estimator is described by Chave (2012).

The remote reference processing can improve the estimator's performance using the cross-spectral instead of the auto-spectral when performing the regression based on the least-squares estimator (Gamble et al., 1979), and the solution of remote reference based on BI-estimator at a specific frequency is given by:

$$\mathbf{Z}=(\mathbf{H}_r{}^{\dagger}\mathbf{w}\mathbf{v}\mathbf{H})^{-1}(\mathbf{H}_r{}^{\dagger}\mathbf{w}\mathbf{v}\mathbf{E}), \quad (5)$$

where $\mathbf{H}_r$ denotes the remote reference magnetic field data. BIRRP is an abbreviation for the Bounded Influence Remote Reference Processing method.

The error bar is calculated by jackknife at a specified frequency. The jackknife is a resampling method used to estimate the bias of a large population. It involves a leave-one-out strategy to estimate the error bar for the data set.

2.2 SSMT-2000

SSMT-2000 is one of the standard Phoenix software set. This program takes the raw time-series

files, calibration files, and site parameter files as inputs. It produces Fourier coefficients in an intermediate step to calculate the impedance tensor using the robust routine.

Before impedance estimation, coherency and resistivity variance are used to filter out the noisy data. Resistivity variance gives a larger weight to data points with smaller error bars. Ordinary coherency gives a larger weight to data points with good coherency between E and H channels. This process reduces the size of the error bars and smooths the curves in plots of apparent resistivity. It gives error estimates that are systematically smaller than other methods, especially at high frequencies.

2.3 RMHHT

The purpose of impedance estimation is to get the impedance tensor depends on frequency. We also can transform the time-series into the frequency domain data by the time-frequency transform technique like HHT.

The impedance tensor derived from Eq. 1 is defined purely in the frequency domain and cannot be used for the instantaneous spectra obtained by the HHT. Neukirch (2013) showed that the instantaneous parameters (IPs) obtained by the HHT could be used to calculate the impedance directly. The impedance is calculated in the time-frequency domain as follows:

$$\begin{pmatrix} \sum_{n=0}^{N} E_{xh}(t_n) H_{xh}^{\dagger}(t_n) \\ \sum_{n=0}^{N} E_{yh}(t_n) H_{xh}^{\dagger}(t_n) \end{pmatrix} = \begin{pmatrix} Z_{xx}(\omega)\ Z_{xy}(\omega) \\ Z_{yx}(\omega)\ Z_{yy}(\omega) \end{pmatrix} \begin{pmatrix} \sum_{n=0}^{N} H_{xh}(t_n) H_{xh}^{\dagger}(t_n) \\ \sum_{n=0}^{N} H_{yh}(t_n) H_{xh}^{\dagger}(t_n) \end{pmatrix}, \tag{6}$$

and

$$\begin{pmatrix} \sum_{n=0}^{N} E_{xh}(t_n) H_{yh}^{\dagger}(t_n) \\ \sum_{n=0}^{N} E_{yh}(t_n) H_{yh}^{\dagger}(t_n) \end{pmatrix} = \begin{pmatrix} Z_{xx}(\omega)\ Z_{xy}(\omega) \\ Z_{yx}(\omega)\ Z_{yy}(\omega) \end{pmatrix} \begin{pmatrix} \sum_{n=0}^{N} H_{xh}(t_n) H_{\mathrm{yh}}^{\dagger}(t_n) \\ \sum_{n=0}^{N} H_{yh}(t_n) H_{yh}^{\dagger}(t_n) \end{pmatrix}, \tag{7}$$

where $E_{xh}(t_n)$,$E_{yh}(t_n)$,$H_{xh}(t_n)$,$H_{yh}(t_n)$ denote the instantaneous values of the analytic signal corresponding to the electromagnetic field, which has the same instantaneous frequency ω at the same time point. t_n denotes the n^{th} time point. The symbol † denotes the complex conjugate transpose. N is the total number of sampling points. The procedure of RMHHT method is as

follows:

(1) Transforming the time-series data into the frequency domain data by HHT

(2) The least-squares regression with the Huber weight function is repeated to estimate MT impedances using the IPs obtained by HHT.

The detail of the RMHHT is as follow:

Step1: Time-frequency transformation

Huang et al. (1998) proposed the Hilbert-Huang transform (HHT). It is the combination of the empirical mode decomposition (EMD) and the Hilbert spectral analysis.

EMD is a technique to adaptively decompose a signal into a finite number of intrinsic mode functions (IMFs), with a trend using a process called the sifting algorithm. IMFs are time-varying mono-component functions, which contain a single frequency. To obtain meaningful IPs, the IMFs have to satisfy the following two conditions (Huang et al., 1998):

(1) in the whole dataset, the number of extrema and the number of zero-crossings must either equal or differ at most by one

(2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero

Based on these two conditions, a signal x(t) is decomposed into IMFs through an iterative process termed sifting (Huang et al., 1998) that is described as follows:

(1) Identify the local extrema(maxima and minima) of x(t).

(2) Interpolate all local maxima and local minima by cubic splines to obtain upper and lower envelopes, respectively.

(3) Subtract the mean m(t) of the upper and lower envelopes from x(t) to obtain the signal h(t) = x(t) − m(t).

(4) If h(t) fulfils the two conditions of IMF, the first IMF $c_1(t) = h(t)$. Otherwise, steps 1–3 are applied on h(t) (instead of x(t)) and are repeated i-1 times until $h_i = h_{(i-1)}(t) - m_i(t)$ fulfills

the conditions of IMF. $c_1(t) = h_i(t)$ is considered as the first IMF.

(5) Subtract the first IMF $c_1(t)$ from x(t) and treat the obtained signal as a new time-series $\tilde{x}(t) = x(t) - c_1(t)$.

(6) Steps 1 to 5 are repeated to determine the subsequent IMFs from the remaining signal $\tilde{x}(t)$. The decomposition process is stopped after n iterations when the residue becomes a monotonic function, which can not be decomposed further. The final signal can then be expressed as a sum of all IMFs and the residual.

The second step of HHT is the Hilbert transform, the complex conjugate y(t) of any real-valued function x(t) can be determined as:

$$y(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau\ , \tag{8}$$

Then the time complex number can be created as:

$$z(t)=x(t)+i*y(t)=a(t)*e^{i\theta(t)}\ , \tag{9}$$

where $a(t)=\sqrt{x(t)^2+y(t)^2}$, $\theta(t)=\arctan(\frac{y(t)}{x(t)})$. Here, a(t) is the instantaneous amplitude, and θ is the instantaneous phase function. The instantaneous frequency is simply defined as:

$$f(t)=\frac{1}{2\pi}\frac{d\theta(t)}{dt}\ . \tag{10}$$

Finally, we apply the Hilbert transform to each IMFs, and we can get the IPs in the time-frequency domain.

Huang et al. (1998) presented the HHT to univariate data, but MT data consists of at least four data channels depending on each other. Rehman and Mandic (2009) developed a new scheme, multivariate empirical mode decomposition (MEMD), to analyze multivariate channels and compute the IMFs of each channel. They maintain a correlation in their time scale as much as possible. Rehman and Mandic also developed the Matlab function "memd.m" to carry out the processing. MEMD provides a set of IMFs for each channel and retains the most similar time scale for all channels. Neukirch (2014) showed that the instantaneous frequencies (IFs) obtained

from IMFs with a similar scale can be considered the same. We will extract the instantaneous parameter in the same frequency band based on the IFs later.

There are different ways to derive instantaneous frequencies (IF) and instantaneous amplitudes (IA) from IMFs. Chen et al. (2012) discussed applying the direct quadrature method to MT data and showed that this method is more accurate than Hilbert transform. A detailed description of the method is given by Huang et al. (2009). The direct quadrature method is implemented by an empirical AM-FM decomposition based on a spline fitted normalization scheme to produce a unique and smooth empiric envelope(AM) and a unity-valued carrier(FM).

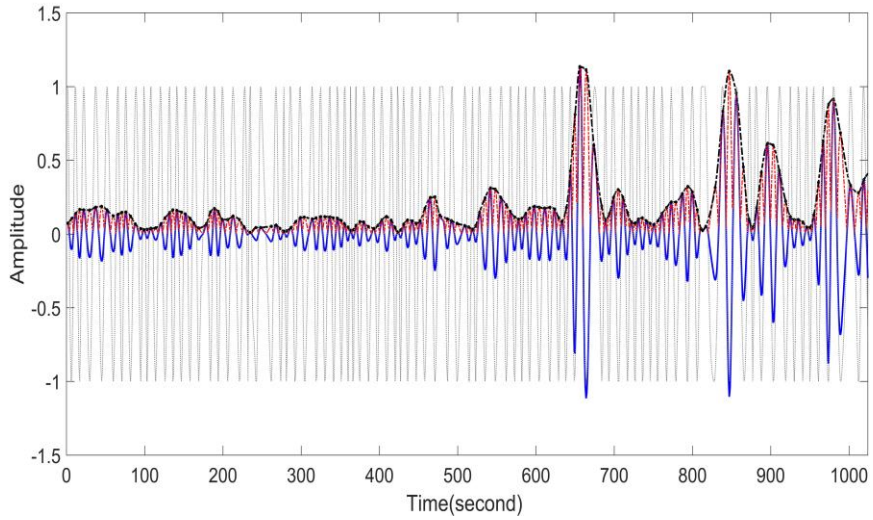

Fig. 1 The empirical AM-FM decomposition. The blue line denotes the original data, and the red line denotes the data's absolute values. The black line indicates the envelope of the maxima of the data's absolute values (AM part). The grey line is the normalized data (FM part).

Fig. 1 illustrates a mono-component time-series, x(t), and the empirical AM-FM decomposition. The blue line denotes the original data. The red line denotes the data's absolute values. At first, the empiric envelope(en(t)) is obtained from the maxima of the data's absolute values by spline fitting (black line in Fig. 1), the AM part is defined first. It is exactly A(t) = e(t), and then the

envelope is used to normalize the data by F(t) = x(t)/en(t), where F(t) is the normalized data. This procedure is repeated until all of the F(t) extrema are unity value (grey line in fig. 1). F(t) is then designated as the empirical FM part of the data. The mono-component signal can be expressed as $x(t) = A(t) * F(t) = A(t) \cos\theta(t)$. Its quadrature part is simply defined as $x^q(t) = A(t) \sin\varphi(t)$; and $\sin\theta(t) = \sqrt{1 - F^2(t)}$. The phase angle is computed by $\theta(t) = \arctan (F(t)/ (1 - F^2(t)))$. The instantaneous frequency can be calculated for each IMF by eq. (10). For the complex time-series $C(t) = x(t) + i* x^q(t)$. We will use the instantaneous parameters to calculate the impedance tensor.

Step 2: Extract the IPs

In the paper, we depend on the EMD's (Neukirch et al. 2014) spectrum estimation and extraction. Before the regression, the independent instantaneous parameters calculated by HHT were gathered, and the same frequency band's IPs were extracted to calculate the impedance tensor. Neukirch et al. (2014) describe how to extract the IPs for the different frequency bands in detail.

We divide the frequency into different frequency bands. As the impedance tensor changes smoothly with frequency, we can group the IPs based on the IFs in the same frequency band to do further processing at different frequency bands.

Step 3: Selection scheme by hat matrix

Before the estimation of impedance, the leverage point should be removed. The hat matrix is N by N matrix and defined as follows:

$$\mathbf{H}_{hat} = \mathbf{H}(\mathbf{H}^{\dagger}\mathbf{H})^{-1}\mathbf{H}^{\dagger}. \quad (11)$$

Chave (2003) suggested that the hat matrix's diagonal element, which is more than several times 2/N, is problematic. As the noisy data is energetic, the hat matrix's diagonal element's statistical analysis is useful to detect the noisy data. N denotes the number of data. Before the regression,

the data where the hat matrix's diagonal element is more than two times 2/N should be removed by the selection strategy to remove the leverage point.

Step 4: Estimation of the impedance tensor

The regression step is similar to the method that Neukirch et al. (2014) and Campanya et al. (2014) described. The impedance tensor is calculated as follows :

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{R}_x + \boldsymbol{\sigma}_x, \tag{12}$$

$$\boldsymbol{Z} = \left(\boldsymbol{R}_{hx}, \boldsymbol{R}_{hy}\right)^{-1}\left(\boldsymbol{R}_{ex}, \boldsymbol{R}_{ey}\right), \tag{13}$$

where $\boldsymbol{y}$ denotes the N by one matrix E_x,E_y,H_x,H_y,H_z; $\boldsymbol{x}$ denotes the N by two matrices, it can be local $[H_x, H_y]$ or remote reference $[H_{xr}, H_{yr}]$, $\boldsymbol{R}_x$ is a two by one matrix and denotes the interstation transfer function. $\boldsymbol{R}_{ex}$ denotes when the **y** equal E_x in equation 12, it is the same to $\boldsymbol{R}_{hx}$, $\boldsymbol{R}_{hy}$, $\boldsymbol{R}_{ey}$.

When $\boldsymbol{x}$ equals the local N by two matrices $[H_x, H_y]$, the $(\boldsymbol{R}_{hx}, \boldsymbol{R}_{hy})$ equal the identity matrix; and the $\mathbf{Z}=\left(\boldsymbol{R}_{ex}, \boldsymbol{R}_{ey}\right)$; in fact, there is no difference between Eqs. 6 and 7. The solution to $\boldsymbol{R}_x$ is a bivariate complex linear regression problem. When $\boldsymbol{y}$ equals E_x, $\boldsymbol{R}_x$ equal to the two by one matrix $\begin{pmatrix} Z_{xx} \\ Z_{xy} \end{pmatrix}$. Equation 12 equal to E_x=$H_x Z_{xx}$+$H_y Z_{xy}$, extend the function to complex form:

$$E_{xr_n} + jE_{xi_n} = \left(H_{xr_n} + jH_{xi_n}\right)(Z_{xxr} + jZ_{xxi}) + \left(H_{yr_n} + jH_{yi_n}\right)\left(Z_{xyr} + jZ_{xyi}\right), \tag{14}$$

where the n=1,2,…, N denotes the n^{th} number of the data, H_{xr_n} denote the n^{th} real part of H_x, H_{xi_n} denote the n^{th} imaginary part of H_x; it is the same to E_x, H_y. Z_{xxr} denotes the real part of Z_{xx}, Z_{xxi} denotes the imaginary part of Z_{xx}; it is the same as Z_{xy}. The symbol j denotes the imaginary number unit. The total equations for all data can be transformed into matrix form as follows:

$$\begin{pmatrix} E_{xr_1} \\ E_{xr_2} \\ \vdots \\ E_{xr_n} \\ E_{xi_1} \\ E_{xi_2} \\ \vdots \\ E_{xi_n} \end{pmatrix} = \begin{pmatrix} H_{xr_1} & -H_{xi_1} & H_{yr_1} & -H_{yi_1} \\ H_{xr_2} & -H_{xi_2} & H_{yr_2} & -H_{yi_2} \\ \vdots & \vdots & \vdots & \vdots \\ H_{xr_n} & -H_{xi_n} & H_{yr_n} & -H_{yi_n} \\ H_{xi_1} & H_{xr_1} & H_{yi_1} & H_{yr_1} \\ H_{xi_2} & H_{xr_2} & H_{yi_2} & H_{yr_2} \\ \vdots & \vdots & \vdots & \vdots \\ H_{xi_n} & H_{xr_n} & H_{yi_n} & H_{yr_n} \end{pmatrix} \begin{pmatrix} Z_{xxr} \\ Z_{xxi} \\ Z_{xyr} \\ Z_{xyi} \end{pmatrix}. \quad (15)$$

The bivariate complex linear regression problem converts to the real multivariate linear regression problem. The Matlab intrinsic function "robustfit.m" can easily solve the regression problem, and Huber weight is adopted. We also can solve the $\boldsymbol{R}_x$ by Eq. 3, following the M-estimator. Both methods give a weight $\mathbf{w}$ to reject the outlier in the electric field. They are the same in theory.

When $\boldsymbol{x}$ equals the remote reference N by two matrices $[H_{xr},H_{yr}]$, the interstation transfer function is calculated at first as follow:

$$\boldsymbol{R}_e = (\mathbf{H}_r{}^{\dagger}\mathbf{H}_r)^{-1}(\mathbf{H}_r{}^{\dagger}\,\mathbf{E}),$$

$$\boldsymbol{R}_h = (\mathbf{H}_r{}^{\dagger}\mathbf{H}_r)^{-1}(\mathbf{H}_r{}^{\dagger}\mathbf{H}). \quad (16)$$

Then impedance tensor is calculated using the relationship between the local transfer function and interstation transfer function as follow:

$$\mathbf{Z} = \boldsymbol{R}_h{}^{-1}\,\boldsymbol{R}_e = (\mathbf{H}_r{}^{\dagger}\mathbf{H}_r)^{-1}\,(\mathbf{H}_r{}^{\dagger}\,\mathbf{E})\,. \quad (17)$$

This strategy can be categorized into one kind of two-stage processing strategy. In this way, separating the calculation into the regression between the remote reference with all channels can avoid a direct effect of coherent noise between local channels. If the remote magnetic field does not contain correlated noise with the local site, it will work well.

Finally, the robust regression was bootstrapped to compute the data-dependent distribution of impedance values and estimate the intrinsic data errors. Empirically, 1,024 iterations were repeated, considering a sufficient trade-off between accuracy and computation time to evaluate results' uncertainty.

3 CHARACTERISTICS OF RMHHT

3.1 The Property of Hilbert-Huang Transform

In the first test, the HHT was applied to synthetic data combining the following equations to show the property of Hilbert-Huang Transform;

$$x_1(t)=\sin(2\pi\cdot 50t)+\sin(2\pi\cdot 200t),$$

$$x_2(t)=\sin(2\pi\cdot 25t)+\sin(2\pi\cdot 100t)+\sin(2\pi\cdot 250t). \quad (18)$$

The sampling rate is 1,000 Hz, and each segment is 4 seconds. Both of the time segments are added with 10% Gaussian noises. The combination of the two segments is shown in Fig. 2. The HHT spectrum is shown in Fig. 3. The color denotes the value of $10\cdot\log10(amp/max)$, where "amp" denotes the amplitude of the IPs, "max" denotes the maximum amplitude of the IPs. Fig.3 showed that the HHT could express the spectrum variation in the time-frequency domain very well.

In this section, we showed the capability of the HHT to express the spectrum variation in the time-frequency domain. Therefore, we could combine the two-segment time-series directly, which are in a different property, and get a reliable spectrum by HHT. Because of this kind of property, we could combine the discontinuous time-series data directly and process the combined discontinuous time-series data in the same way as the continuous time-series data when we compute the impedance by RMHHT.

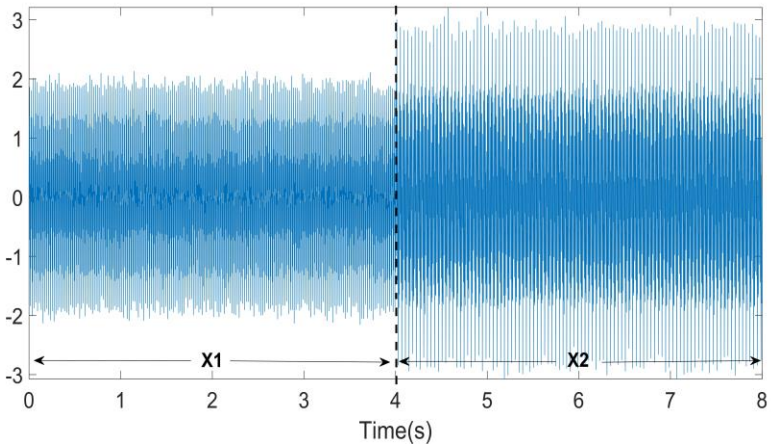

Fig. 2 The time-series x_1(t)=sin(2π·50t)+sin(2π·200t) and x_2(t)=sin(2π·25t)+sin(2π·100t) +sin(2π·250t). Both of the segments are added with 10% Gaussian noises.

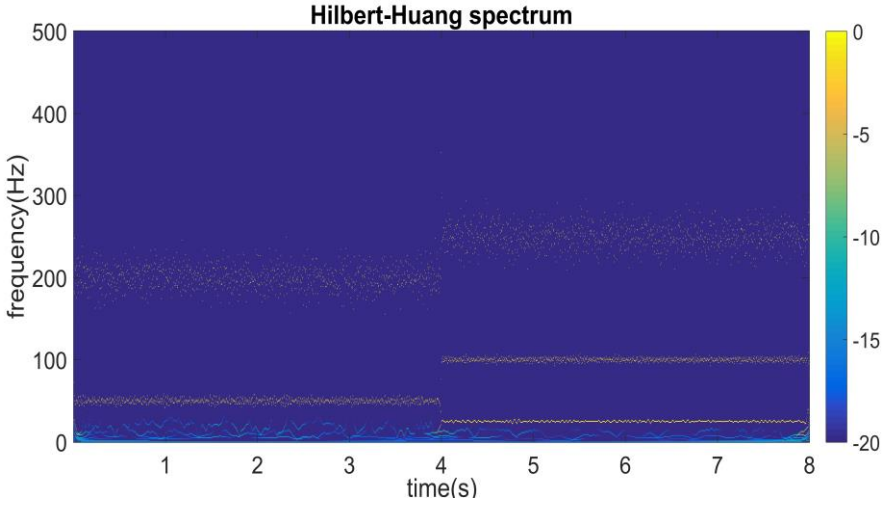

Fig. 3 The HHT spectrum of time-series data, which is shown in Fig. 1. The color denotes the value of 10·log10(amp/max), where "amp" denotes the amplitude of the IPs, "max" denotes the maximum of the amplitude of the IPs.

3.2 The Test with Different Length of Synthetic Time-series data

In this subsection, we tried to compare the performance of RMHHT by different lengths of synthetic MT time-series data. The way to create synthetic time-series is similar to the method

proposed by Chen (2012). The 1-day magnetic time-series data from Memambetsu (MMB) station was used as the synthetic magnetic time-series data. MMB is one of the Magnetic Observatory stations where geomagnetic and geoelectric observations are performed in Japan. The sampling rate is 1 second, and its unit is nT. At first, the time-series magnetic field h(t) is transferred to the frequency domain using the Fourier transform. Then, the impedance $\mathbf{Z}(\omega)$, calculated from the simple 1-D model, is multiplied with $\mathbf{H}(\omega)$ to determine $\mathbf{E}(\omega)$. Subsequently, the electric field spectra are transformed back into the time domain to obtain the electric time-series $e(t)$.

We tested the effectiveness of data-length using 1-hour, 2-hour, 3-hour, 4-hour, 5-hour, 6-hour, 12-hour, and 24-hour synthetic MT time-series data. The 1-hour time-series data can only get the maximum reliable period is 150 seconds. The maximum reliable period is determined by the error bar and the impedance result. 4-hour time-series data is relatively good enough to get the maximum reliable period of about 1,000-second. Fig. 4 shows the results calculated using 2-hour and 4-hour synthetic MT time-series data. The light blue lines are the true model, which are calculated using a simple 1-D model. The 2-hour data set was used to calculate the dark blue circles. The 4-hour data set was used to calculate the red circles. The 2-hour time-series data can get the maximum reliable period is about 300-seconds. Below 300-seconds the result coincides with the 4-hour's result, and the 2-hour's result (blue circles) is covered by the 4-hour's result (red circles).

The XY component calculated using the different lengths of time-series at 874-second is shown in Fig.5. The red dotted lines denote the true model, which was calculated using a 1-D model. The horizontal axis represents the length of the time-series data. The error bar becomes smaller by extending the length of the time-series. We think 4 hours is good enough to get a reliable result.

From this result, we confirmed that if there were 2-hour noise-free time-series data, we could get reliable results at least with a period of 300-second. If there are approximately 4-hour noise-

free time-series data, we could get reliable data with a period of 1,000-second by RMHHT.

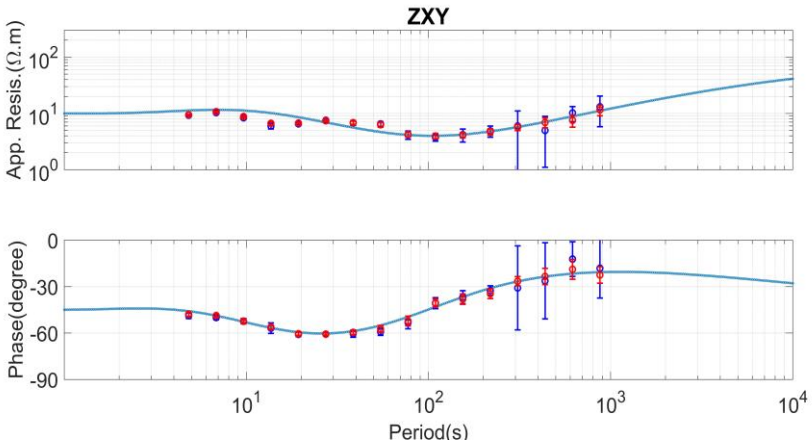

Fig. 4 The comparison of the impedance curves calculated by the 2-hour and 4-hour time-series dataset. The light blue curves are the true model curves, which were calculated from the simple 1-D model. The 2-hour dataset was used to calculate the dark blue circles, and the 4-hour dataset was used to calculate the red circles. The upper figures show the apparent resistivity, and the lower figures show the impedance phase.

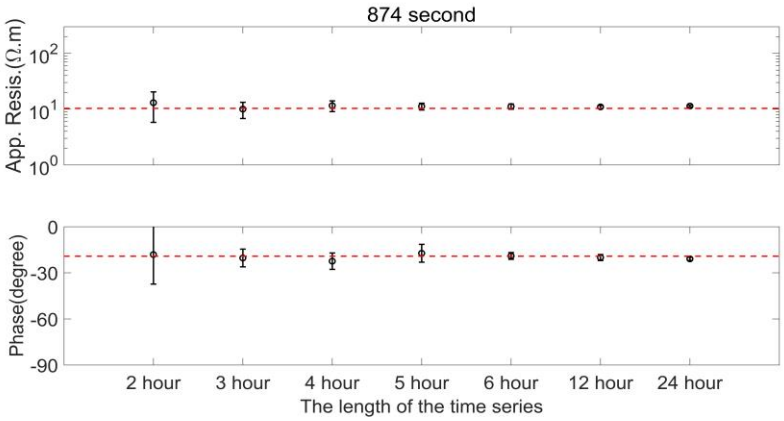

Fig. 5 The comparison of the TE mode impedance result calculated by the different lengths of time-series at the period of 874-second. The red dot line denotes the true model curves, which were calculated by the forward modeling. The horizontal axis indicates the length of the time-series.

4 APPLICATIONS OF RMHHT TO MT DATA

4.1 The Characteristic of the Field MT Data Observed by Phoneix Equipment

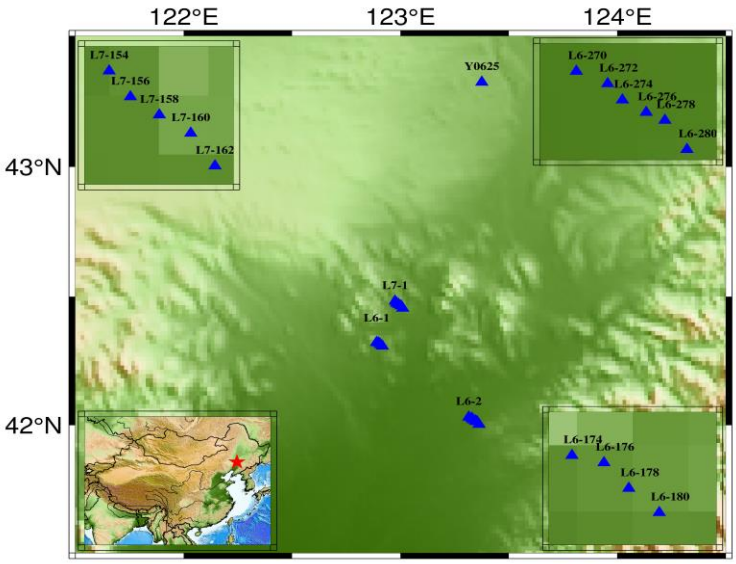

Fig. 6 The location map of the study area. Y0625 is the remote reference site, and L6-1, L6-2, L7-1 in the middle is the local survey line. The lower left inset map shows the survey area in China. The red star denotes the research location. The upper left insert map shows the detail of survey line L7-1; The lower right insert map shows the detail of survey line L6-1; The upper right insert map shows the detail of survey line L6-2.

In this paper, the broadband-frequency MT time-series data observed by the Phoenix MTU-A Geophysical Instruments is used. This data belongs to the Institute of Geophysical and Geochemical Exploration, China Geological Survey. In this data set, the high-frequency band (2,400 Hz) sampled 1 second at intervals of 2 minutes from the beginning of the minute, the middle-frequency band (150 Hz) sampled 8 seconds at intervals of 2 minutes from the beginning of the minute, and the low-frequency band (15 Hz) sampled continuously. Fig.6 shows the location map of sites in the study area. The local survey lines are L6-1, L6-2 and L7-1, and the Y0625 is set as the remote reference site. The lower left inset map shows the survey area in China. The upper left insert map shows the detail of survey line L7-1; The lower right insert map shows

the detail of survey line L6-1; The upper right insert map shows the detail of survey line L6-2. The observation period is shown in Table 1. In this research, we used the data observed on June 26, 2016.

Table 1 The observation date of each site.

Observation Date	Site Names
2016.06.25	Y0625, L6174, L6176, L6278, L6280, L7154, L7156
2016.06.26	Y0625, L6178, L6274, L6276, L7158
2016.06.27	Y0625, L6180, L6270, L6272, L7160, L7162

The field data consists of natural sources and local noise. The natural MT source coming from the magnetosphere or global lighting is far enough from the local site following the plane wave's assumption. All other parts of the measured electric and magnetic fields are considered as noise. We can subdivide the noise and signals into three parts: the correlated noise (^{CN}), where the noise is coherent between **E** and **H** field, uncorrelated noise (^{UN}), and MT signal (^{MT}). With this subdivision, we can rewrite the fields **E** and **H** as follows:

$$\mathbf{E} = \mathbf{E}^{MT} + \mathbf{E}^{CN} + \mathbf{E}^{UN},$$
$$\mathbf{H} = \mathbf{H}^{MT} + \mathbf{H}^{CN} + \mathbf{H}^{UN}. \quad (19)$$

Szarka (1988) and Junge (1996) summarized the active and passive noise sources observed in MT measurements. Most of the artificial noise is caused by earthing currents originated from the electrical equipment. Usually, the remote reference magnetic field has a high correlation with the local magnetic field. Therefore, the correlation between the local magnetic field and remote reference magnetic field is a good indicator of whether the noise is strong or not.

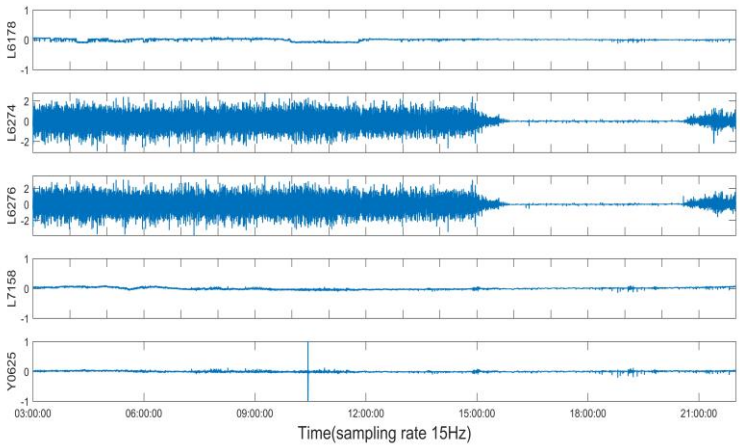

Fig. 7 The time variations of the Hx component whose sampling rate is 15 Hz. They were measured simultaneously from 3:00:00 to 22:00:00 in UCT time. The unit of the magnetic field is nT.

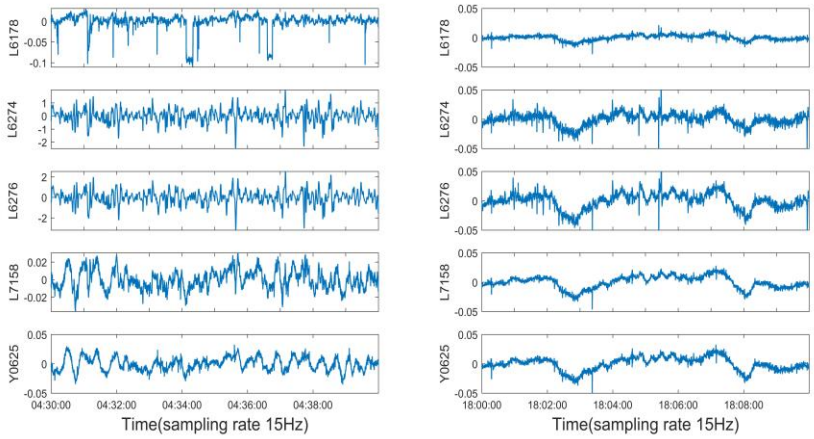

Fig. 8 The time variations of the Hx components whose sampling rate is 15 Hz. The left figures show the daytime data, and the right figures show the night data. Each data segment is for ten minutes. The unit of the magnetic field is nT.

Fig. 7 shows the Hx component time-series measured simultaneously from 3:00:00 to 22:00:00 on June 26 in different sites. The sampling rate is 15 Hz. The variation in the daytime is much stronger than that at nighttime at site L6274 and L6276. The variation of the time-series at other

sites is almost keeping the same scale. Fig. 8 shows the Hx component, whose sampling rate is 15 Hz, measured simultaneously in different sites. The left figures show the data in the daytime, and the right figures show the data at nighttime. Each segment is ten minutes. The unit of the magnetic field is nT. The variation at site L6274 and L6276 is about 100 times the variation at other sites in the daytime and similar at nighttime. That means the noise to signal ratio is 100 in the daytime; this will bias the impedance tensor a lot.

Fig. 9 shows the correlation variation between the local and remote magnetic field (Hx) at L6274 and Y0625 from 3:00:00 to 22:00:00 in UCT time on June 26. The correlation is based on the Pearson correlation using the time-series segment, and each segment length is 10 minutes. The blue denotes a negative correlation, and the red indicates a positive correlation. Due to local noise, the correlation becomes low or negative in the daytime, and the correlation becomes higher at nighttime. It means that the data in the local night is much quiet.

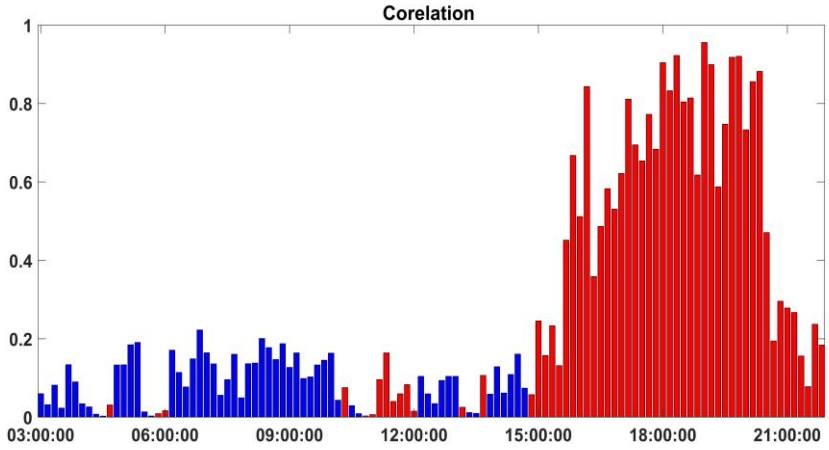

Fig. 9 The correlation variation between the local and remote magnetic field (Hx) at L6274 and Y0625 from 3:00:00 to 22:00:00 in UCT time on June 26. The sampling rate is 15 Hz. Each segment is 10 minutes. Blue denotes a negative correlation, and red indicates a positive correlation.

Moreover, we also introduce another parameter, polarization directions, to estimate the background noise in a different frequency band. The polarization directions of the electric field (α_E) and magnetic field (α_H) (Fowler et al., 1967) at a specific frequency are defined as:

$$\alpha_E = arctan\frac{2Re[E_x E_y{}^{\dagger}]}{[E_x E_x{}^{\dagger}]-[E_y E_y{}^{\dagger}]} \tag{20}$$

$$\alpha_H = arctan\frac{2Re[H_x H_y{}^{\dagger}]}{[H_x H_x{}^{\dagger}]-[H_y H_y{}^{\dagger}]} \tag{21}$$

The polarization directions describe the time-harmonic variation of two orthogonal field components with constant phasing. A variety of sources generates a natural magnetic signal. These sources generate magnetic fields that vary in their incidence directions, and thus there is no preferred polarization direction for the magnetic field. However, according to a given conductivity distribution of the subsurface, there might be a preferred polarization direction of the induced electric field (Weckmann et al., 2005).

In this research, we analyzed the polarization directions at different frequencies for all sites. As the signal strength at the dead band (1-10s) is very low, it is easy to influence the local noise. The typical polarization directions for the dead band is shown in Fig.10. This figure shows the result of polarization directions calculated by the site L7158 at 6-second. The polarization directions of the magnetic field have a preferred direction from 3:00:00 to 15:00:00. That means the data in that period is contaminated by strong noise. A similar situation occurs at all sites, including the remote site.

This data set is contaminated by typical industrial noise from a manufacturing plant from the analysis above. The correlation of time-series between the local site and remote site, at site L6274 and L6276, is much higher at nighttime. The magnetic field's polarization has a preferred direction in the dead-band at all sites in the daytime. That means the data is contaminated by strong noise in the dead band at all sites in the daytime. Garcia et al. (2002) also found that the high-frequency band's MT signal was stronger at nighttime. The signal to noise ratio is higher at nighttime than in the daytime at all frequency bands in this data set. Therefore, we consider using

the data at nighttime only to calculate the impedance tensor.

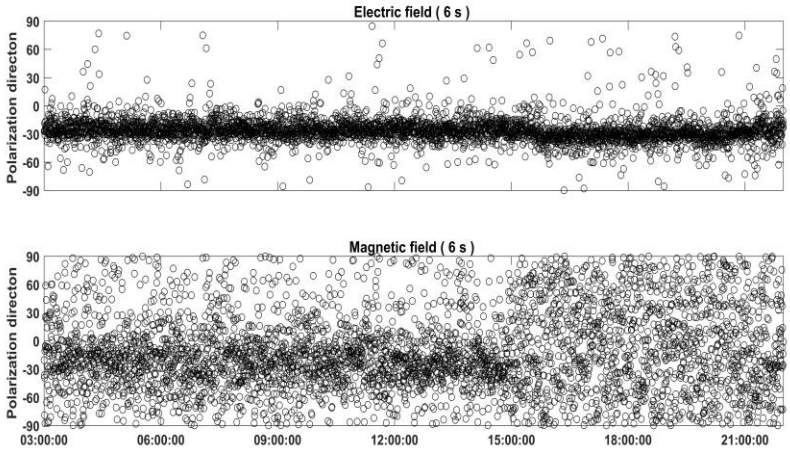

Fig. 10 The polarization direction for a period of 6-second at site L7158 from 3:00:00 to 22:00:00 in UCT time. The upper figure shows the electric field's polarization directions, and the lower is the polarization direction of the magnetic field.

4.2 The Strategy to Calculate the Wideband Time-series Data Based on RMHHT

The processing of broadband MT time-series data by RMHHT is demonstrated in this section. The data set observed using the MTU-A geophysical instruments was used. The time-series data are stored in three files. The high-frequency band (2,400 Hz) and the middle-frequency band (150 Hz) were sampled intermittently. The length of each segment is 8 seconds and 1 second separately. The low-frequency band (15 Hz) was sampled continuously.

Table 2 shows the strategy for the discontinuous time-series. For the windowed FFT, we should transform each segment into the frequency domain and group the target frequency data in the frequency domain. In contrast, the HHT can express the spectrum variation well in the time-frequency domain, as shown in subsection 3.1. We can combine the intermittent time-series data directly and process the discontinuous data in the same way as the continuous time-series data. Considering the accuracy of the complex coefficient data, we set the effective frequency to 4

times of sampling rate. Using the high-frequency (2,400 Hz) data, calculate the period between 4/2,400 to 4/150 seconds. One continuous segment(1 second) contains about 37 windows of the maxima period. There are 72,000 data points in 1-hour, and 100,000 points (about one and a half hours) are used to calculate the impedance. The 150 Hz data was used to calculate the period between 4/150 to 4/15 seconds. One continuous segment(8 seconds) contains about 30 windows of the maxima period. There are 36,000 data points in 1-hour, and 100,000 points (about 3 hours) are used to calculate. For the high-frequency band (2,400 Hz) and the middle-frequency band (150 Hz), we suggest that 2-hour time-series data is enough to get a reliable result.

Table 2 Strategy for the discontinuous time-series data.

Sampling rate (Hz)	The length of one segment (s)	Calculated period (s)	Data points for 1 hour	Total data points
2400	1	4/2400 to 4/150	72,000	100,000
150	8	4/150 to 4/15	36,000	100,000

Table 3 shows the strategy for continuous time-series. As the computation of HHT is time-consuming. We design a strategy that can decrease the computation time much. The 15-Hz data was used to calculate the period data between 4/15 to 5 seconds. There are 54,000 data points in 1-hour, and 100,000 points (about 2 hours) are used to calculate the impedance, and we downsampled the continuous 15-Hz data to 1-Hz to compute the period larger than 5 seconds. As shown in the simulation, 4-hour noise-free time-series data can get a relatively reliable result until 1,000-second. We can calculate the full scall period from high frequency to 1,000 seconds by four hours by combining all the results.

This strategy allows us to select data in the time domain. There may be several hours at nighttime that most of the electric equipment is shutdown. The time-series data in these periods are much quiet. This strategy is suitable to use only the quiet period time-series data to calculate

the impedance.

Table 3 Strategy for continuous time-series data.

Sampling rate (Hz)	Calculated period (s)	Data points for 1 hour	Total data points
15	4/15 to 5	54,000	100,000
15 --> 1	> 5	3,600	14,400

4.3 The Application to the Field Data

In this section, we will compare the performance of the three different estimators by the field data. The upper figures of all the results show the apparent resistivity, and the lower figures show the impedance phase.

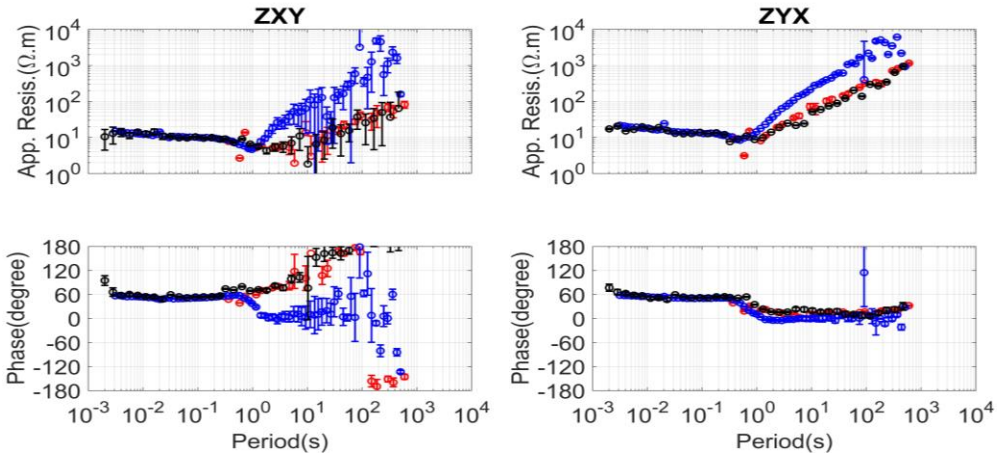

Fig. 11 The magnetotelluric sounding curves calculated from the data at site L6274. SSMT- 2000 was used to calculate the blue curves using one-day data; the BIRRP was used to calculate the red curves. And the RMHHT was used to calculate the black curves. Both BIRRP and RMHHT results were computed using nighttime data from 12:00:00 to 16:00:00 in UTC on June 26. The upper figures show the apparent resistivity curves, and the lower figures show the impedance phase curves.

Fig.11 shows the result calculated by site L6274. The RMHHT estimator calculated the black curves. The result was calculated by the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26. This calculation took about 16 hours. It is also possible to calculate the different frequency band data simultaneously, and the computation time can be limit to around 3 hours. The BIRRP code calculated the red curves. We only used the BIRRP code to process continuous time-series. The time-bandwidth 3 for the Slepian sequence taper is applied to each section. The result was also calculated by the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26. The SSMT-2000 calculated the blue curves. The results were calculated from one-day data.

In the SSMT-2000 result, the XY component phase was close to 0º, and the apparent resistivity act as a straight line in the log scale in the long period. That is the phenomenon of artificial noise (Zonge and Hughes,1987). The daytime and nighttime data are quite different for long periods, so the RMHHT and BIRRP results using nighttime data are statistically different from the SSMT-2000 results calculated by one-day data. Unfortunately, we failed to get a reliable result in the long period (1 to 1,000 seconds) by the BIRRP code and RMHHT as the nighttime data is also noisy. On the other hand, the RMHHT and SSMT-2000 results are consistent and reliable at high-frequency periods.

Fig.12 shows the result derived from site Y0625. SSMT-2000 calculates the blue curves from one-day data, BIRRP calculates the black curves using daytime data from 12:00:00 to 16:00:00 in UTC on June 26. BIRRP calculates the red curves using nighttime data from 16:00:00 to 20:00:00 in UTC on June 26. The SSMT-2000 and BIRRP results calculated using the one-day data have unnatural humps between 1 and 10 s in the apparent resistivities. On the other hands, BIRRP result calculated by the nighttime data become smooth and change naturally between 1 and 10 s in the apparent resistivities and phase. This result reflects the different properties between the daytime and nighttime in the dead band and is consistent with the background noise analysis in section 4.1.

Fig.13 shows the result derived by site Y0625. SSMT-2000 calculates the blue curves from

one-day data. BIRRP calculates the red curves by nighttime data from 16:00:00 to 20:00:00 in UTC on June 26. It is the same as shown in Fig.12. RMHHT calculates the black curves using the nighttime data from 16:00:00 to 20:00:00 in UTC using L7158 as the remote site. By comparing the full-scale result, both RMHHT and SSMT-2000 results coincide with each other. The same problem occurs in the dead band. RMHHT and BIRRP results coincide with each other using the nighttime data. But in the TM mode, the phase of RMHHT is changed unnaturally. We think it may be introduced by the magnetic field noise (at site L7158) as the remote site. In the full-scale, the RMHHT perform well comparing with the BIRRP and SSMT-2000 result.

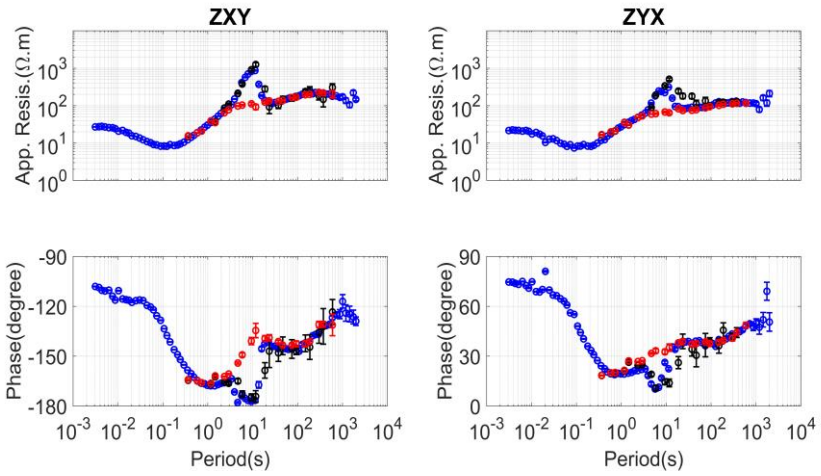

Fig. 12 The magnetotelluric sounding curves calculated using the data at site Y0626. SSMT-2000 was used to calculate the blue curves from one-day data, BIRRP was used to calculate the black curves using daytime data from 12:00:00 to 16:00:00 in UTC on June 26. BIRRP was used to calculate the red curves from nighttime data from 16:00:00 to 20:00:00 in UTC on June 26.

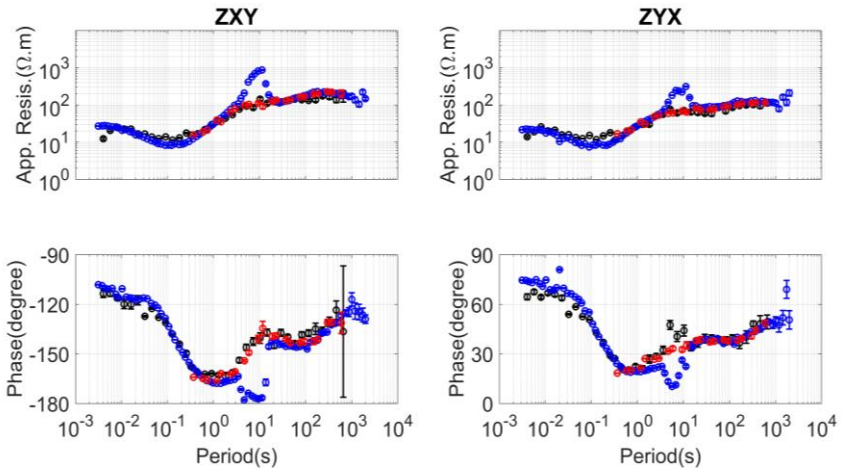

Fig. 13 The magnetotelluric sounding curves calculated using the data at site Y0626. SSMT-2000 was used to calculate the blue curves using one-day data, RMHHT was used to calculate the black curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26 at the remote site L7158. BIRRP was used to calculate the red curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26.

Similar results were obtained at the L6178 and L7158. Fig.14 and Fig.15 show the result calculated using the data at site L6178 and L7158, respectively. The RMHHT result performs well at the full-scale. In the dead band, as we used the relatively quiet data at the Y0625 as a remote site, it performs as well as BIRRP, but in the long period of around 1,000-second, the error bar calculated by RMHHT becomes lagger. Suppose there is another quiet nighttime data; we can combine them directly. In this way, we can get more data at the impedance estimation step in the frequency domain, and the result will become more stable.

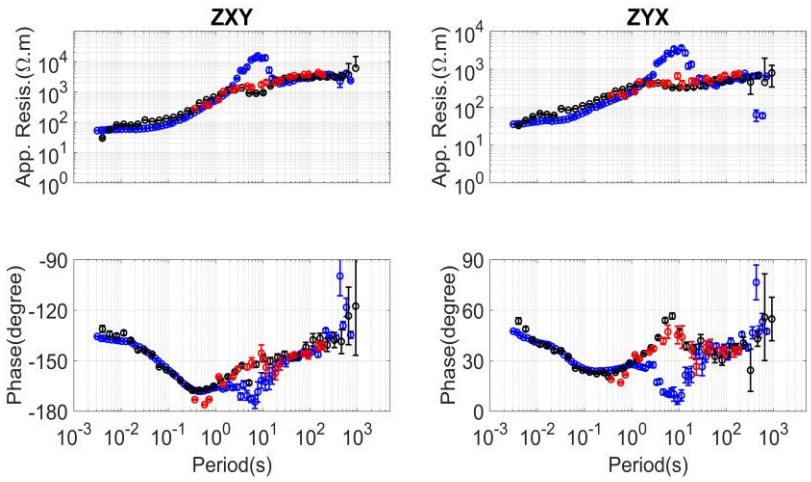

Fig. 14 The magnetotelluric sounding curves calculated using the data at site L6178. SSMT-2000 was used to calculate the blue curves using one-day data, RMHHT was used to calculate the black curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26 at the remote site Y0625. BIRRP calculates the red curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26.

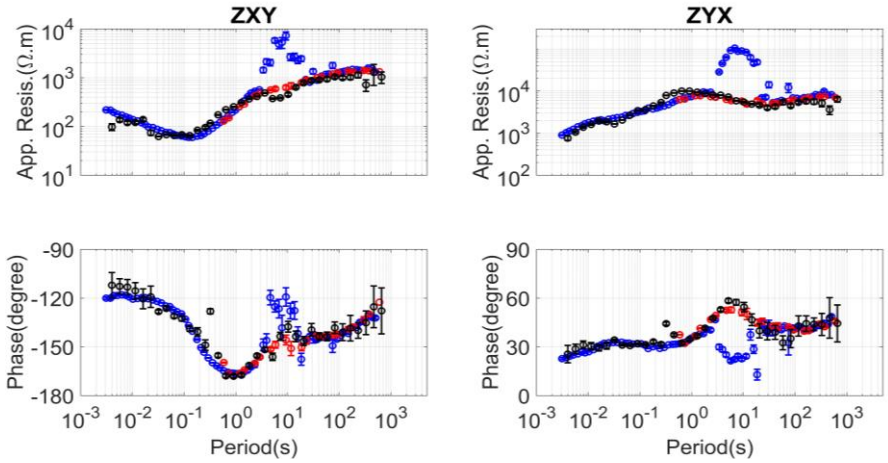

Fig. 15 The magnetotelluric sounding curves calculated using the data at site L7158. SSMT-2000 was used to calculate the blue curves using one-day data, RMHHT was used to calculate the black curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26 at the remote site Y0625. BIRRP was used to calculate the red curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26.

5 CONCLUSIONS AND DISCUSSION

At nighttime, most of the electrical equipment is shutdown. Usually, the data is much quieter than in the daytime. We tried to use the quiet period time-series data only to calculate the impedance tensor. We introduced a robust impedance estimator based on the HHT (RMHHT) and demonstrated a strategy to process the broadband MT time-series data properly. It has been shown that we could get a reliable result until 1,000-second from continuous 4-hour noise-free time-series data in the simulation. This short measurement time makes it possible to do MT exploration in strong electromagnetic noise areas. This paper also demonstrated that a time window of 4-hour length was almost noise-free during the local nighttime at site L7158, L6178, and Y0625. We used the data from this quiet period alone to estimate impedances and get a reliable result.

Moreover, if there are several nighttime time-series data, we can combine them directly. In this way, we can increase the number of data sets at the impedance estimation step. The result will become more stable.

This paper showed that RMHHT could get a similar result with the conventional result based on windowed FFT from the quiet data set. We have not shown the method's superiority based on the HHT than the conventional method based on FFT. Any method needs the normal data at the last step of impedance estimation. When transforming the time-series data into the frequency domain, a part of the time-series data, which is contaminated by the transient noise, will influence the whole spectrum calculated by the Fourier transform. In the worst case, every segment contains transient noises. The method based on the FFT is impossible to get a reasonable impedance.

In contrast, HHT can express the spectrum variation in the time-frequency domain well. The transient noises will affect the signal contaminated time-span, and the other parts are not affected. We can remove the IPs in the contaminated period in the frequency domain before the impedance

estimation and get a reliable result. The method based HHT can reject the noisy data maximum in the time-frequency domain. We think it is promising to pay more attention to identifying the noise data in the time-frequency domain based on the HHT.

Declarations

Availability of data and materials

The magnetic time-series data used to construct the synthetic MT data was downloaded from the INTERMAGNET (International Real-time Magnetic Observatory Network). The data used in this study belong to the Institute of Geophysical and Geochemical Exploration, China Geological Survey under the regional geophysical survey technology research and information system platform construction, DD20190030-05. Alan Chave provided the BIRRP code. Maik Neukirch provided the EMT code.

Competing interests

There are no conflicts of interest associated with this publication.

Funding

The data used in this study belong to the Institute of Geophysical and Geochemical Exploration, China Geological Survey under the regional geophysical survey technology research and information system platform construction, DD20190030-05.

Authors' contributions

Hao Chen wrote the code, processed the time-series data by BIRRP and RMHHT estimator. Hao Chen contributes about 60%. Hideki Mizunaga reviewed the manuscript and contributed about 30%; Gang Wang processed the time-series data by SSMT-2000 software. Gang Wang and Toshiaki Tanaka contribute about 10% to this work.

Acknowledgments

We thank Maik Neukirch for his constructive comments and for providing the EMT code to investigate. Finally, we also thank the reviewer Alan Chave who gave us constructive comments for improving the manuscript.

REFERENCES

Berdichevsky, M. N. and Bezruk I. A. (1973): Magnetotelluric Sounding with the Use of Mathematical Filters, Izv. Akad. Nauk SSSR.

Campanya, J., Ledo, J., Queralt, P., Marcuello, A., and Jones, A. G. (2014): A new methodology to estimate magnetotelluric (MT) tensor relationships: Estimation of Local transfer-functIons by Combining Interstation Transfer-functions (ELICIT), *Geophys. J. Int*., **198**, 484-494.

Chave, A. D. and Thomson, D. J. (1989): Some comments on magnetotelluric response function estimation, *J.Geophys.Res.*, **94**, 14215–14225.

Chave, A. D. and Thomson, D. J. (2003): A bounded influence regression estimator based on the statistics of the hat matrix, *J. R. Stat. Soc., Ser. C,* **52**, 307–322.

Chave, A. D. and Thomson, D. J. (2004): Bounded influence estimation of magnetotelluric response functions, *Geophys. J. Int*, **157**, 988–1006.

Chave, A. D. and Jones, A. G. eds. (2012): The magnetotelluric method: Theory and practice. Cambridge University Press.

Chen, J., Heincke, B., Jegen, M. and Moorkamp, M. (2012): Using empirical mode decomposition to process marine magnetotelluric data, *Geophys. J. Int.*, **190**, 293–309.

Fowler, R.A., Kotick, B.J., Elliott, R.D., (1967): Polarization analysis of natural and artificially induced geomagnetic micropulsations. *J. Geophys. Res*. **72**, 2871–2883.

Gamble, T. D., Goubau, W. M. and Clarke, J. (1979): Magnetotellurics with a remote reference,

Geophysics, **44**, 53–68.

Garcia, X. and Jones, A. G. (2002): Atmospheric sources for audio-magnetotelluric (AMT) sounding, *Geophysics*, **67**, 448-458.

Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., and Chen, X. (2009): On instantaneous frequency, *Adv. Adapt. Data Anal.*, **1**, 77.

Junge, A. (1996): Characterization of and correction for cultural noise, *Surv. Geophys.*, **17**, 361–391

Neukirch, M. and Garcia, X. (2013): Nonstationary Time-series Convolution: On the Relation Between the Hilbert–Huang and Fourier Transform. Advances in Adaptive Data Analysis, **5**, 1350004.

Neukirch, M. and Garcia, X. (2014): Nonstationary magnetotelluric data processing with the instantaneous parameter, *J. Geophys. Res. Solid Earth*, **119**, 1634–1654.

Rehman, N., and Mandic D. P. (2009): Multivariate empirical mode decomposition, Proc. Phys. Soc. London, Sect. A, **466**, 1291–1302.

Sims, W. E., Bostick, F. X., and Smith, H. W. (1971): The estimation of magnetotelluric impedance tensor elements from measured data, *Geophysics*, **36**, 938–942.

Szarka, L. (1988): Geophysical aspects of man-made electromagnetic noise in the Earth - A review, *Surveys in Geophysics*, **9**, 287-318.

Weckmann, U., Magunia, A., Ritter, O., (2005): Effective noise separation for magnetotelluric single site data processing using a frequency domain selection scheme, *Geophys. J. Int*., **161**, 635–652.

Zonge, K. L. and Hughes, L. J. (1987): Controlled source audio-frequency magnetotellurics, Electromagnetic methods in applied geophysics, Applications, 713–809, ed. Nabighian, M.N., SEG, Tulsa.

Figures

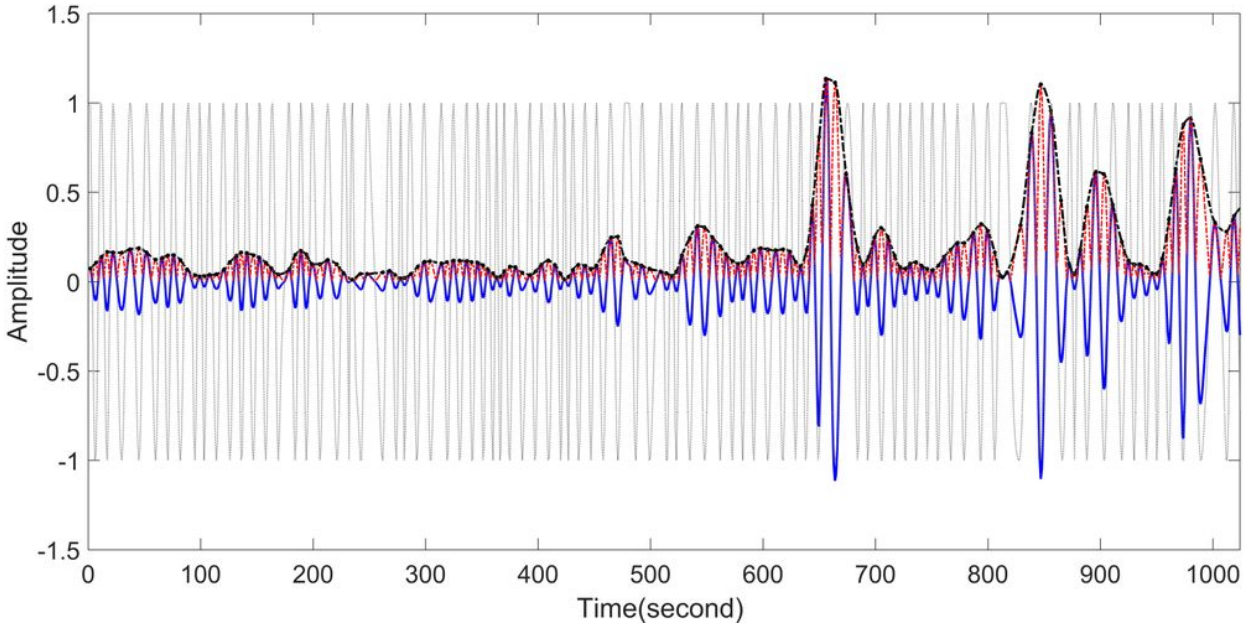

Figure 1

The empirical AM-FM decomposition. The blue line denotes the original data, and the red line denotes the data's absolute values. The black line indicates the envelope of the maxima of the data's absolute values (AM part). The grey line is the normalized data (FM part).

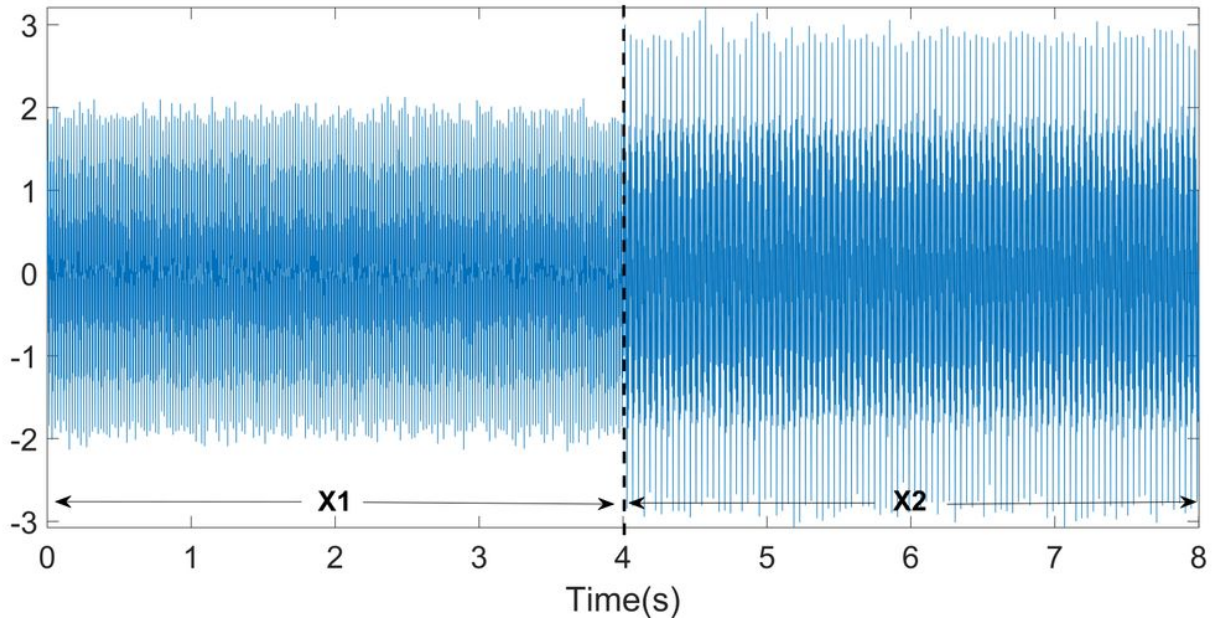

Figure 2

The time-series x1(t)=sin(2π·50t)+sin(2π·200t) and x2(t)=sin(2π·25t)+sin(2π·100t) +sin(2π·250t). Both of the segments are added with 10% Gaussian noises.

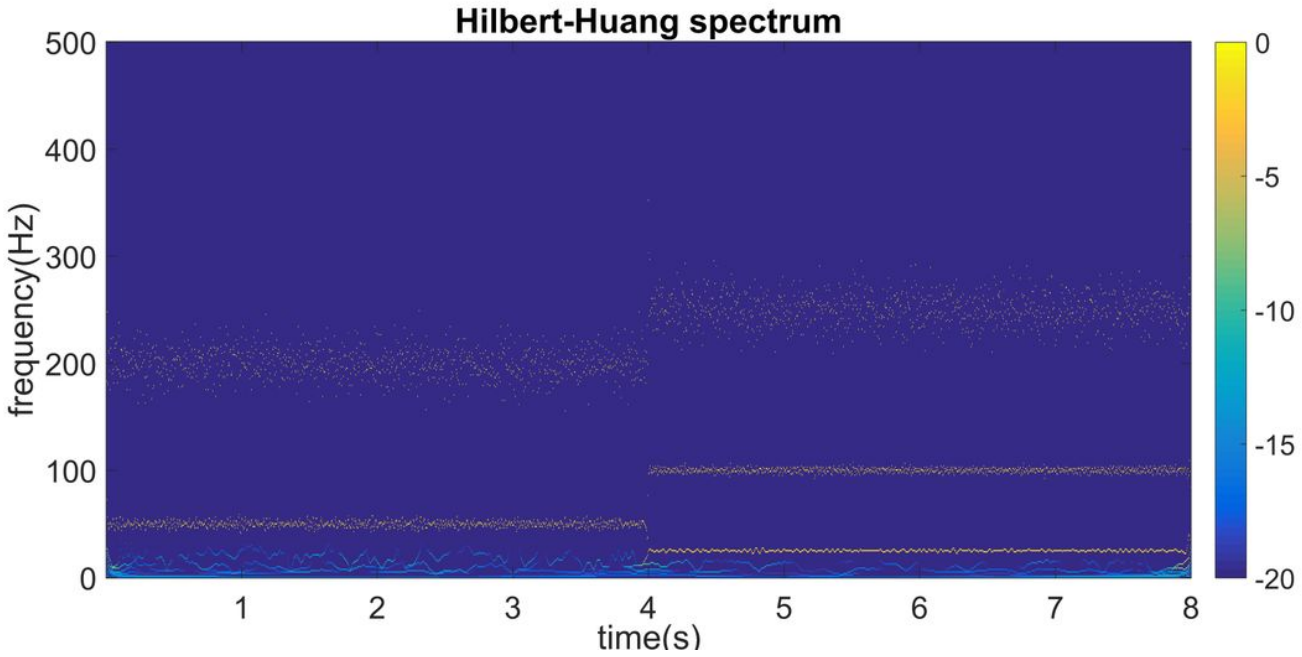

Figure 3

The HHT spectrum of time-series data, which is shown in Fig. 1. The color denotes the value of 10·log10(amp/max), where "amp" denotes the amplitude of the IPs, "max" denotes the maximum of the amplitude of the IPs.

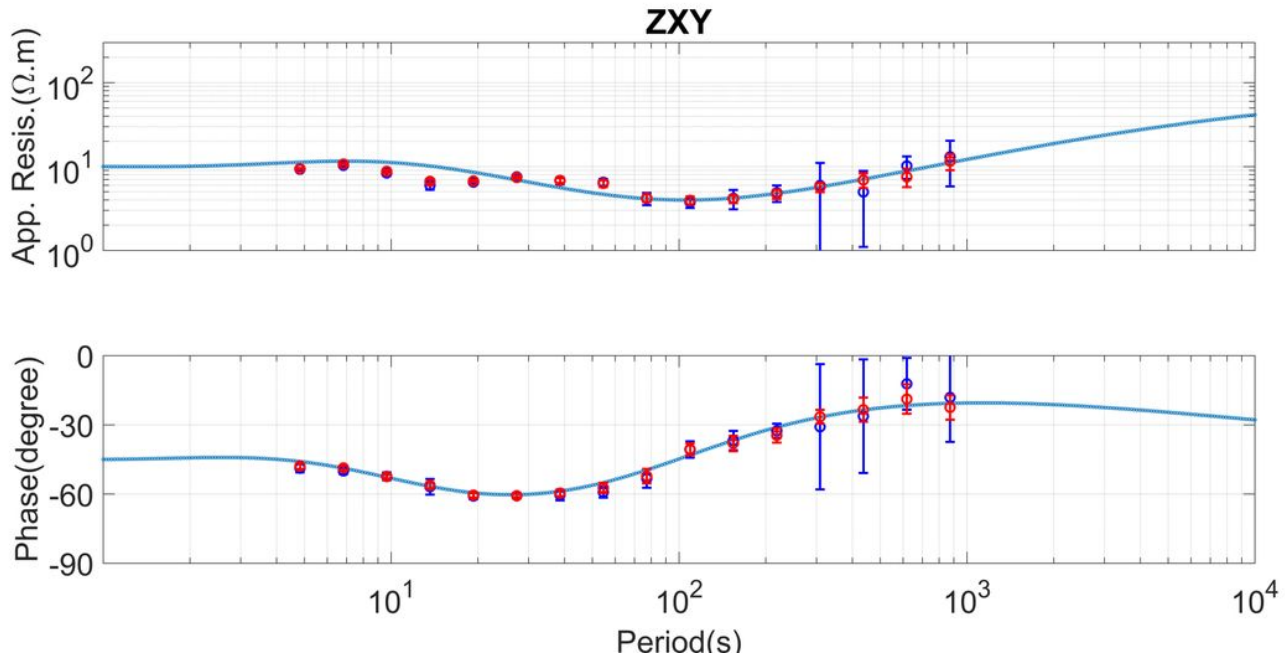

Figure 4

The comparison of the impedance curves calculated by the 2-hour and 4-hour time-series dataset. The light blue curves are the true model curves, which were calculated from the simple 1-D model. The 2-hour dataset was used to calculate the dark blue circles, and the 4-hour dataset was used to calculate the red circles. The upper figures show the apparent resistivity, and the lower figures show the impedance phase.

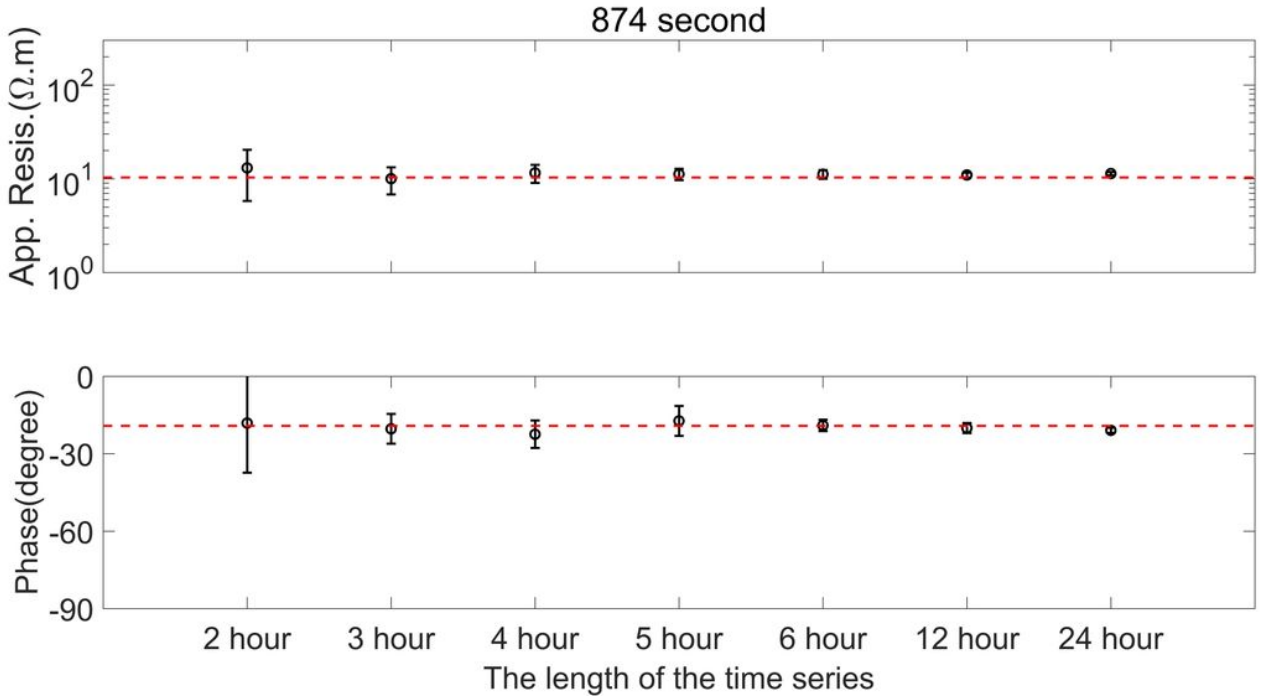

Figure 5

The comparison of the TE mode impedance result calculated by the different lengths of time-series at the period of 874-second. The red dot line denotes the true model curves, which were calculated by the forward modeling. The horizontal axis indicates the length of the time-series.

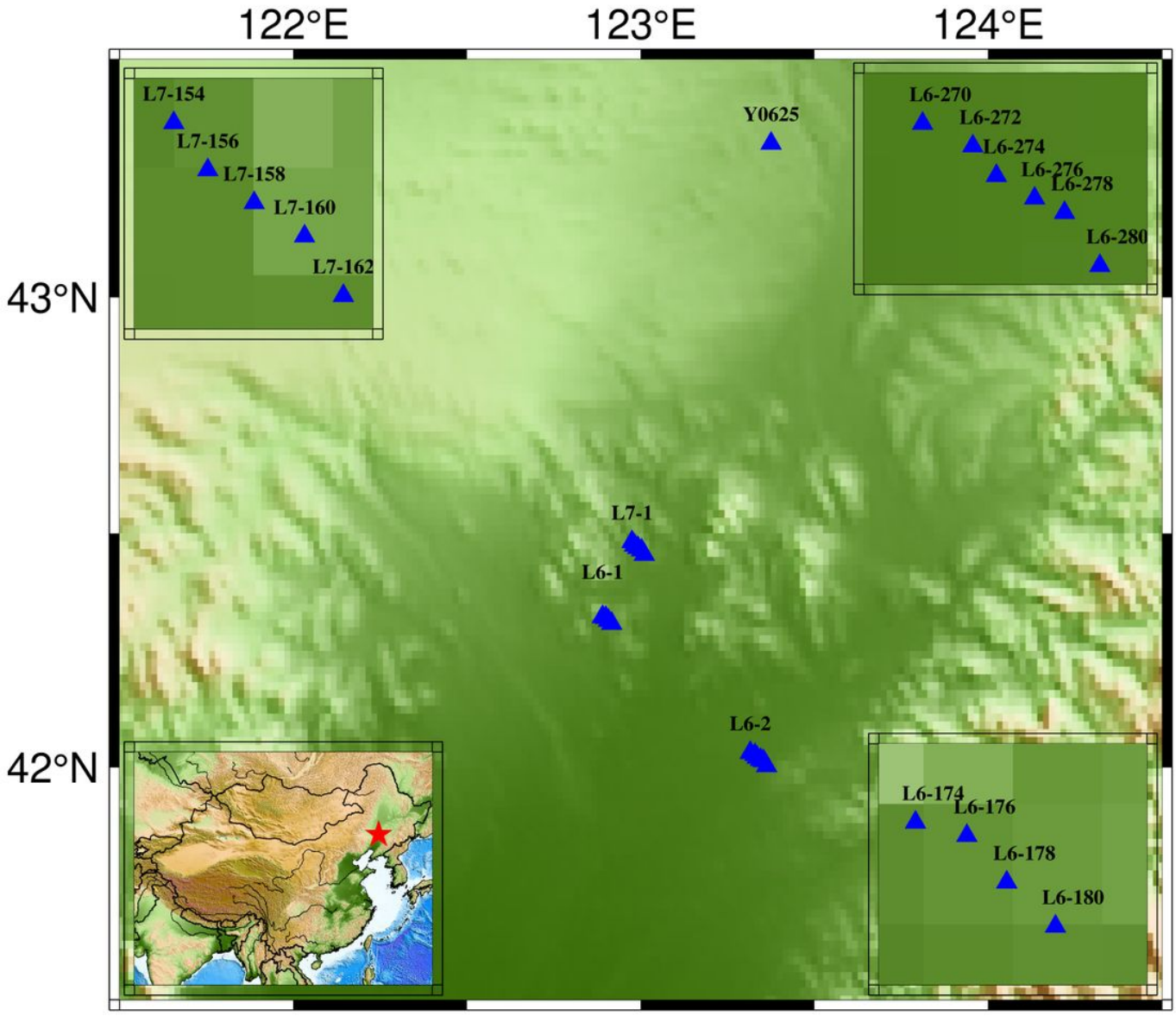

Figure 6

The location map of the study area. Y0625 is the remote reference site, and L6-1, L6-2, L7-1 in the middle is the local survey line. The lower left inset map shows the survey area in China. The red star denotes the research location. The upper left insert map shows the detail of survey line L7-1; The lower right insert map shows the detail of survey line L6-1; The upper right insert map shows the detail of survey line L6-2. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

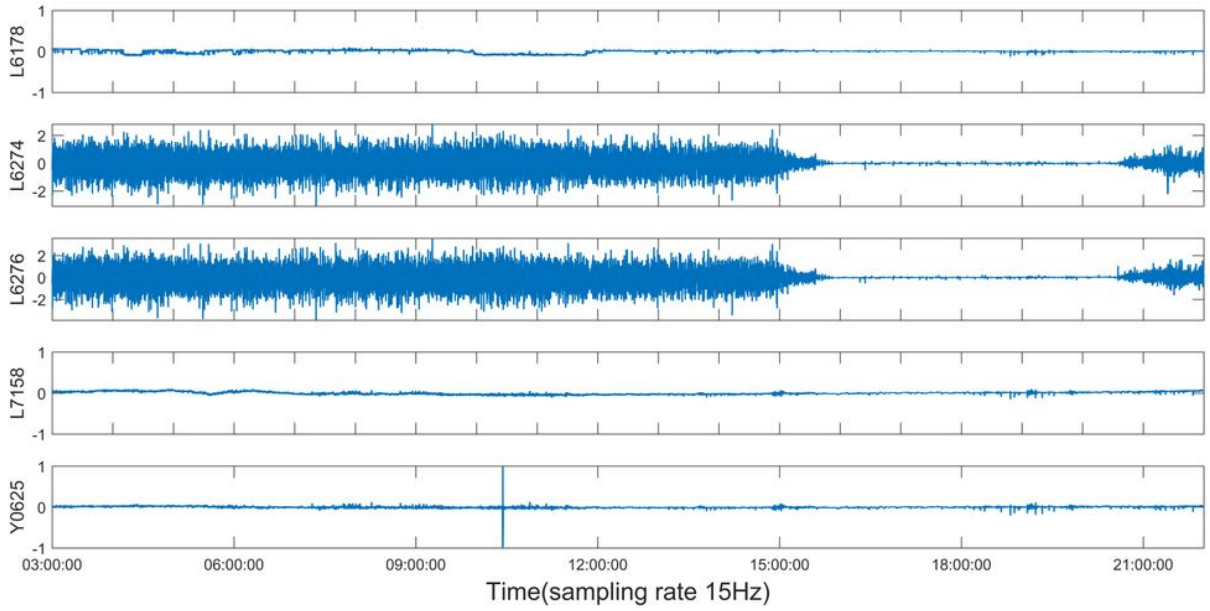

Figure 7

The time variations of the Hx component whose sampling rate is 15 Hz. They were measured simultaneously from 3:00:00 to 22:00:00 in UCT time. The unit of the magnetic field is nT.

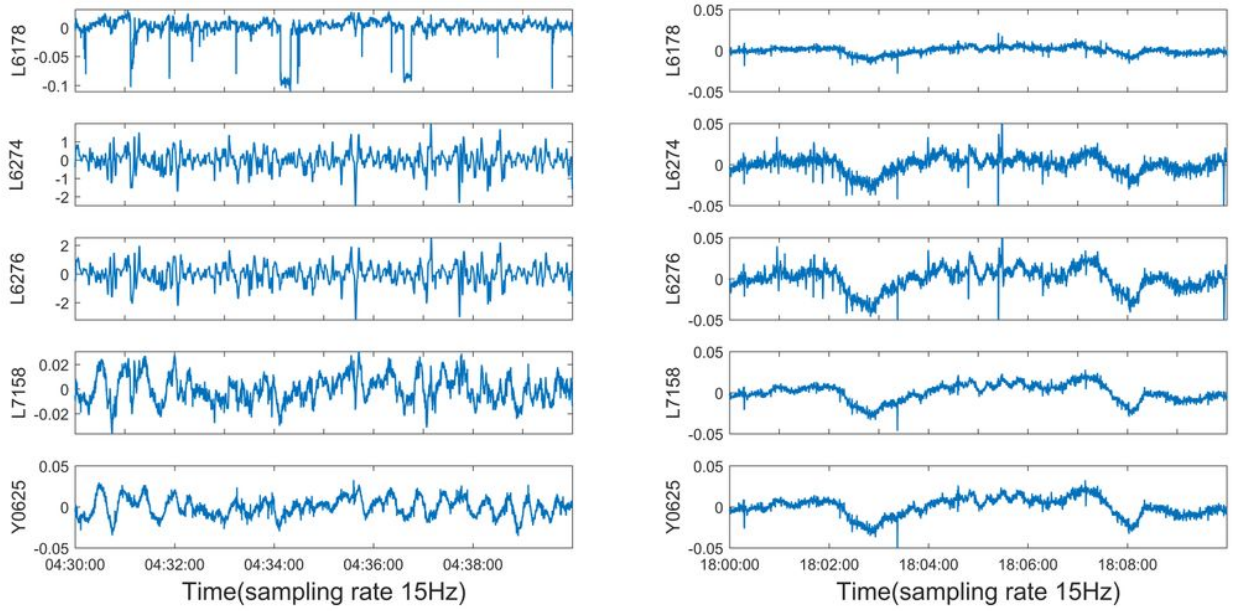

Figure 8

The time variations of the Hx components whose sampling rate is 15 Hz. The left figures show the daytime data, and the right figures show the night data. Each data segment is for ten minutes. The unit of the magnetic field is nT.

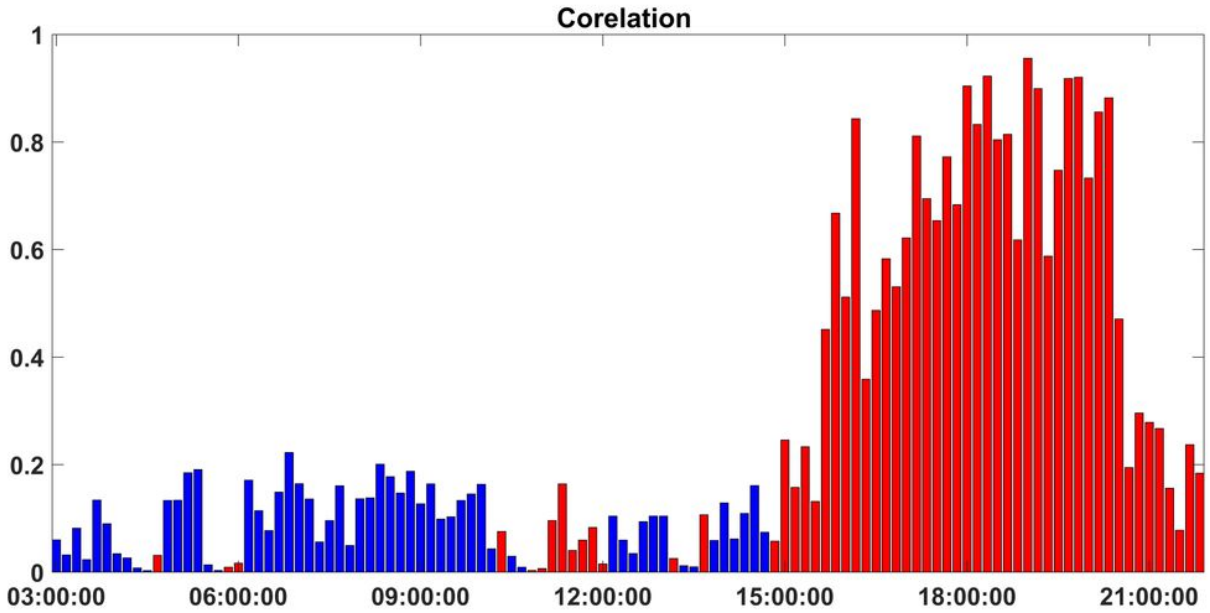

Figure 9

The correlation variation between the local and remote magnetic field (Hx) at L6274 and Y0625 from 3:00:00 to 22:00:00 in UCT time on June 26. The sampling rate is 15 Hz. Each segment is 10 minutes. Blue denotes a negative correlation, and red indicates a positive correlation.

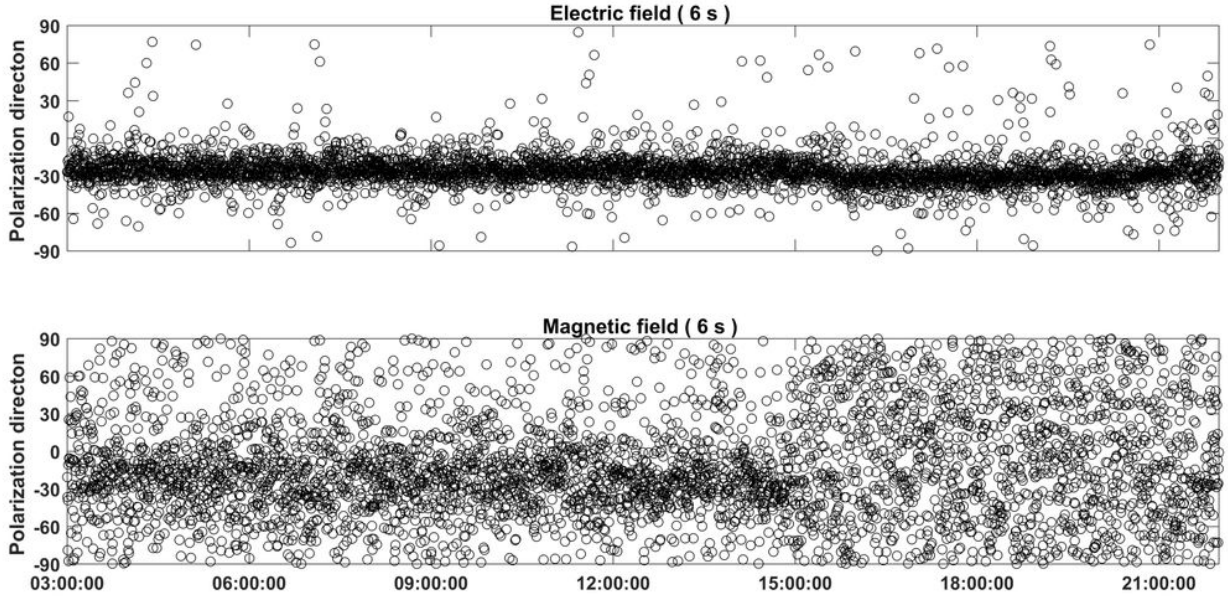

Figure 10

The polarization direction for a period of 6-second at site L7158 from 3:00:00 to 22:00:00 in UCT time. The upper figure shows the electric field's polarization directions, and the lower is the polarization direction of the magnetic field.

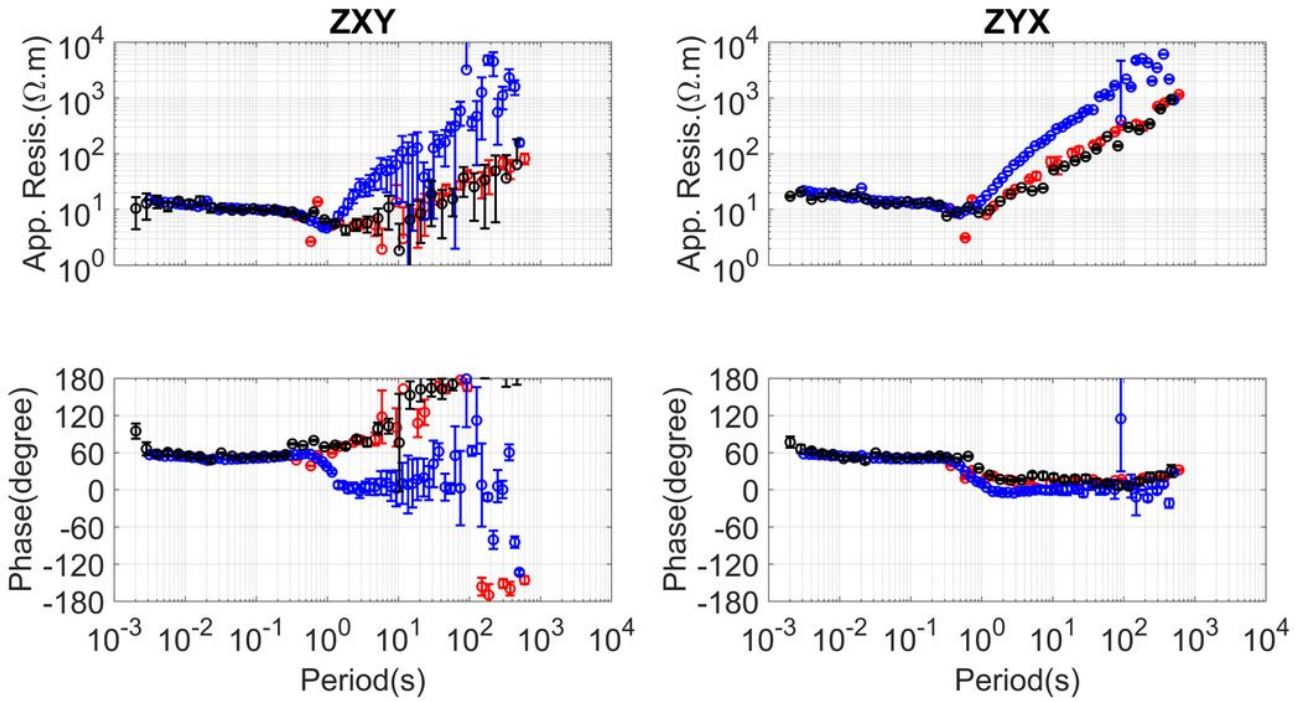

Figure 11

The magnetotelluric sounding curves calculated from the data at site L6274. SSMT- 2000 was used to calculate the blue curves using one-day data; the BIRRP was used to calculate the red curves. And the RMHHT was used to calculate the black curves. Both BIRRP and RMHHT results were computed using nighttime data from 12:00:00 to 16:00:00 in UTC on June 26. The upper figures show the apparent resistivity curves, and the lower figures show the impedance phase curves.

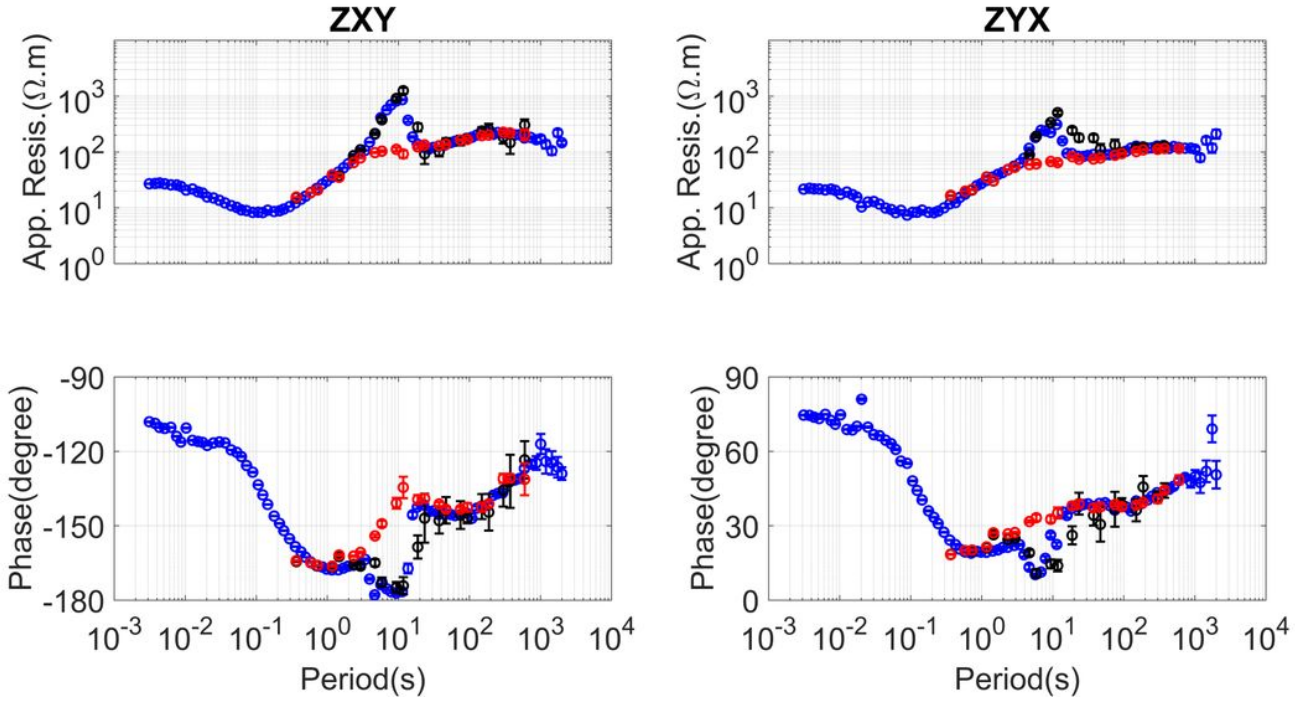

Figure 12

The magnetotelluric sounding curves calculated using the data at site Y0626. SSMT-2000 was used to calculate the blue curves from one-day data, BIRRP was used to calculate the black curves using daytime data from 12:00:00 to 16:00:00 in UTC on June 26. BIRRP was used to calculate the red curves from nighttime data from 16:00:00 to 20:00:00 in UTC on June 26.

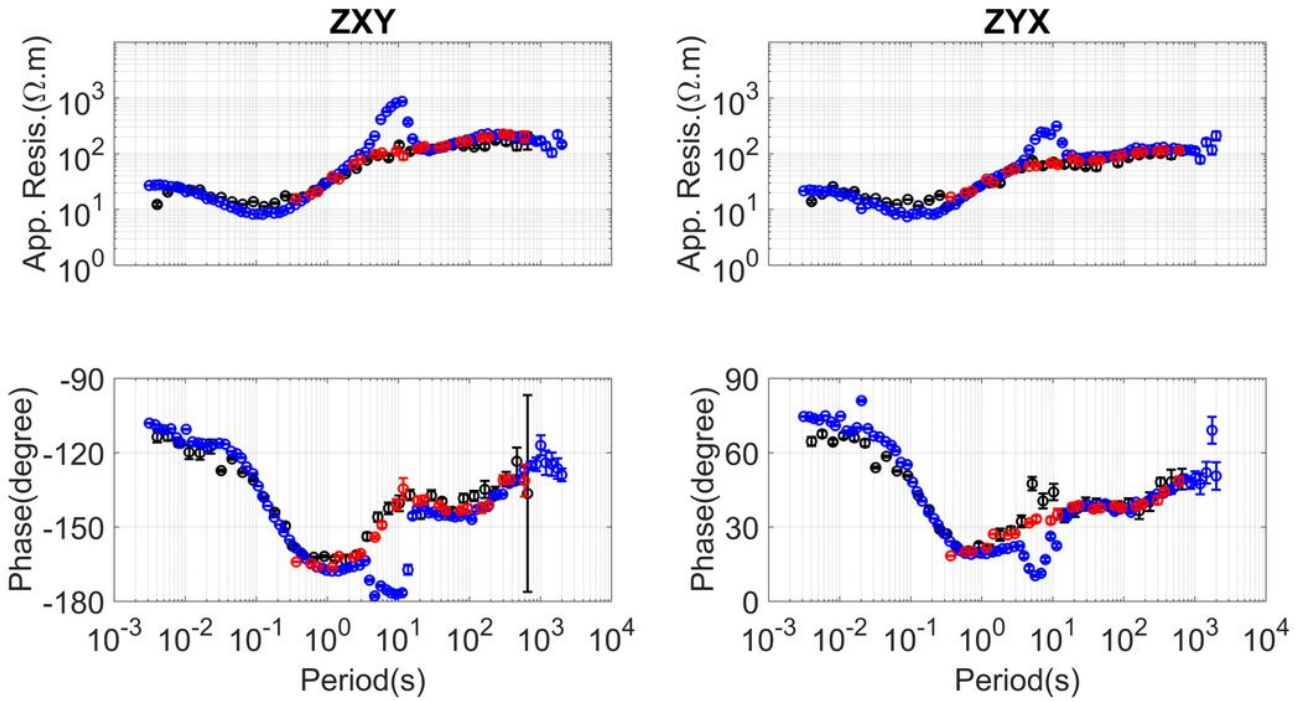

Figure 13

The magnetotelluric sounding curves calculated using the data at site Y0626. SSMT-2000 was used to calculate the blue curves using one-day data, RMHHT was used to calculate the black curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26 at the remote site L7158. BIRRP was used to calculate the red curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26.

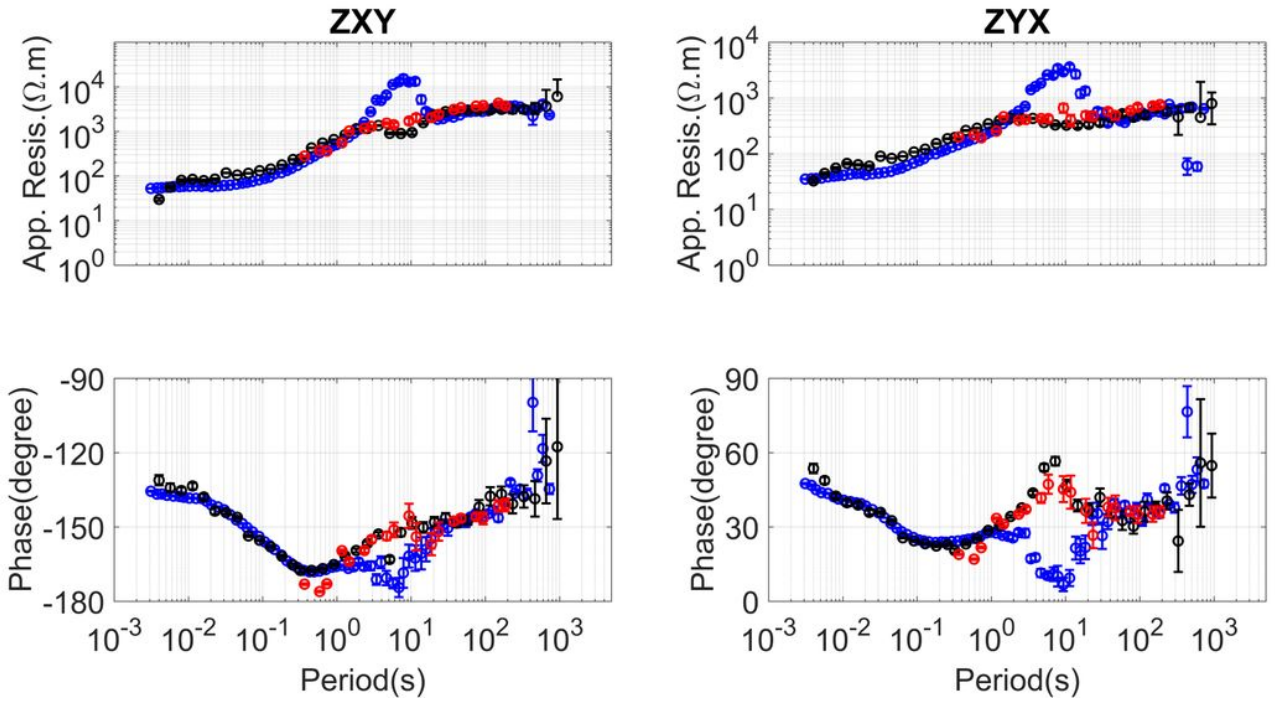

Figure 14

The magnetotelluric sounding curves calculated using the data at site L6178. SSMT-2000 was used to calculate the blue curves using one-day data, RMHHT was used to calculate the black curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26 at the remote site Y0625. BIRRP calculates the red curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26.

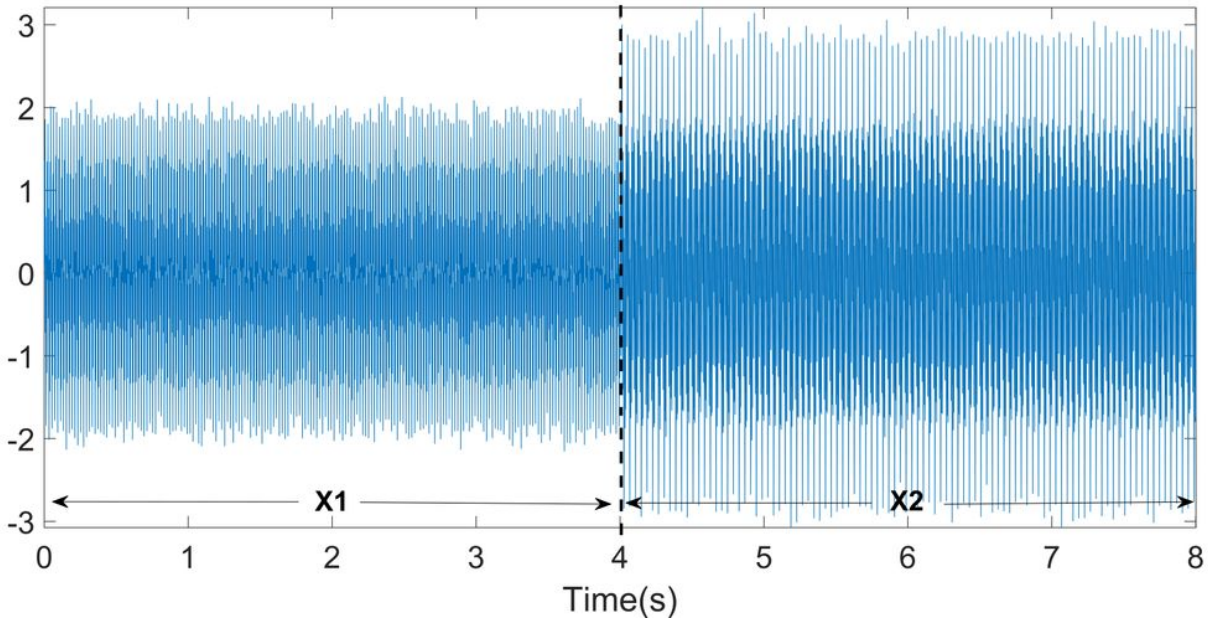

Figure 15

The magnetotelluric sounding curves calculated using the data at site L7158. SSMT-2000 was used to calculate the blue curves using one-day data, RMHHT was used to calculate the black curves using the

nighttime data from 16:00:00 to 20:00:00 in UTC on June 26 at the remote site Y0625. BIRRP was used to calculate the red curves using the nighttime data from 16:00:00 to 20:00:00 in UTC on June 26.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GraphicalAbstract.docx