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

Hao Chen¹, Hideki Mizunaga², Toshiaki Tanaka² and Gang Wang³

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: 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 worldwide, but it was hardly applicable in urban areas because of large artificial electromagnetic noise. But in one day, there are several hours at midnight that most of the electric equipment is shutdown. The MT time-series data at midnight is much quiet than in the daytime. Therefore, we focused on calculating the MT impedance using the quiet time-series data. In this research, the data observed by the Phoenix System is used. We introduced a robust impedance estimator based on the Hilbert-Huang transform (RMHHT) and used a new strategy to calculate the broadband MT time-series data. We indicated that this technique needs 4-hour time-series data to get a reliable resistivity structure up to 1,000 seconds in the numerical simulation. This short
measurement time makes it possible to carry out MT surveys in urban areas with strong noises. However, the biggest problem of the impedance estimator based on HHT is time-consuming. The total computation time can be reduced significantly by using this strategy. Finally, a successful case study using the midnight time-series data was demonstrated to get reliable resistivity structures in the areas contaminated by heavy noises.

Keywords: impedance; magnetotelluric; Hilbert-Huang transform

1 INTRODUCTION

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 large artificial electromagnetic noise.

The first step in MT data processing is to estimate the impedance tensor in the frequency domain from the measured time-series data. In standard MT data processing routines, the original time series is divided into the short segments and then transforms the segments into the frequency domain by Fourier Transform. Here we should know one segment is one dataset when doing the impedance estimation in the frequency domain. 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), particularly urban areas. If most of the time series data contain strong noise, any impedance estimator will fail to get reliable results. But in one day, there is always several hours' time window at midnight that most of the electric equipment is shutdown. The time-series data in the period is much quiet. In this
paper, we consider using the quiet period time-series data only to calculate the impedance tensor.

Huang et al. (1998) introduced empirical mode decomposition (EMD) in the framework of the Hilbert-Huang Transform (HHT), a novel time series analysis tool, which 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. Huang et al. (1998) argued that the IMF allows 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, the method based on the HHT performs better than the traditional method based on windowed FFT.

Phoenix System is widely used for observing the broadband frequency MT time series data. In this paper, the data sets observed by the MTU-A Geophysical Instruments is used. The time-series data are stored in three files. Two of them store the high and middle-frequency band (2,400 and 150 Hz) in a few seconds at intervals of several minutes from the beginning of the minute. The other file stores the low-frequency data (15 Hz) continuously. When the noise is strong, the SSMT-2000 program, the standard set of Phoenix software, will fail to get a reliable result occasional. Therefore, demanding for developing new data processing technology for the noisy data.

For the conventional method, windowed FFT is widely used to calculate the spectrum for a different frequency. In this paper, as an alternative way, we proposed a new strategy using a robust M-estimator based on HHT (RMHHT) to calculate the full-scale broadband MT time-series data. In the synthetic data test, we have shown that there is continuous 4-hour noise-free time series data. We could get a reliable result until 1,000 seconds. At midnight, most of the electrical equipment is shutdown. Usually, the data at midnight is much quiet than in the daytime. This strategy is suitable to get a reliable result by selecting the data in the time domain. That makes it possible to use the midnight time-series data only to calculate the impedance tensor. The
biggest problem of the estimator based on HHT is time-consuming. By this strategy, the
computation time can be shortened much.

At last, we compared the performance of RMHHT with the conventional method Bounded
Influence Remote Reference Processing, BIRRP (Chave and Thomson, 2004), and the SSMT
2000 program and showed the successful case study using the midnight time-series 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 section 4.3.

2.1 BIRRP

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

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

\[
\begin{pmatrix}
E_x(\omega) \\
E_y(\omega)
\end{pmatrix} = \begin{pmatrix}
Z_{xx}(\omega) & Z_{xy}(\omega) \\
Z_{yx}(\omega) & Z_{yy}(\omega)
\end{pmatrix} \begin{pmatrix}
H_x(\omega) \\
H_y(\omega)
\end{pmatrix}.
\]

(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:

\[
Z = (H^\dagger H)^{-1}(H^\dagger E),
\]

(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 \(\dagger\) denotes the complex conjugate transpose.
The M-estimator 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.

And the solution of M-estimator at a specific frequency is given by:

\[ Z = (H^\dagger wH)^{-1}(H^\dagger wE). \] (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 weighting \((v)\) 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:

\[ Z = (H^\dagger wvH)^{-1}(H^\dagger wvE). \] (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:

\[ Z = (H_r^\dagger wvH)^{-1}(H_r^\dagger wvE), \] (5)

where \(H_r\) 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, which are then used to calculate the impedance tensor by 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 small 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. It is similar to the method that Berdichevsky (1973) proposed the impedance tensor estimation involved narrow bandpass filtering in the time-frequency domain as follows:

\[
\begin{bmatrix}
\sum_{n=0}^{N} E_{xh}(t_n) H_{xh}^{\dagger}(t_n) \\
\sum_{n=0}^{N} E_{yh}(t_n) H_{xh}^{\dagger}(t_n)
\end{bmatrix} = \begin{bmatrix}
Z_{xx}(\omega) & Z_{xy}(\omega) \\
Z_{yx}(\omega) & Z_{yy}(\omega)
\end{bmatrix} \begin{bmatrix}
\sum_{n=0}^{N} H_{xh}(t_n) H_{xh}^{\dagger}(t_n) \\
\sum_{n=0}^{N} H_{yh}(t_n) H_{xh}^{\dagger}(t_n)
\end{bmatrix},
\]

(6)

and

\[
\begin{bmatrix}
\sum_{n=0}^{N} E_{xh}(t_n) H_{yh}^{\dagger}(t_n) \\
\sum_{n=0}^{N} E_{yh}(t_n) H_{yh}^{\dagger}(t_n)
\end{bmatrix} = \begin{bmatrix}
Z_{xx}(\omega) & Z_{xy}(\omega) \\
Z_{yx}(\omega) & Z_{yy}(\omega)
\end{bmatrix} \begin{bmatrix}
\sum_{n=0}^{N} H_{xh}(t_n) H_{yh}^{\dagger}(t_n) \\
\sum_{n=0}^{N} H_{yh}(t_n) H_{yh}^{\dagger}(t_n)
\end{bmatrix},
\]

(7)

where \(E_{xh}(t_n), E_{yh}(t_n), H_{xh}(t_n), H_{yh}(t_n)\) denote the instantaneous values of analytic signal corresponding components of the electromagnetic field, which has the same instantaneous frequency \(\omega\) at the same time point. Neukirch (2013) showed that the instantaneous parameters (IPs) obtained by the HHT could be used directly because they truly are analytic signals. \(t_n\) denotes the \(n^{th}\) time point. The symbol \(\dagger\) 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:

**Step 1**: Time-frequency transformation

Huang et al. (1998) proposed the Hilbert-Huang transform (HHT). It is based on the combination of the empirical mode decomposition (EMD) and the Hilbert spectral analysis. Huang et al. (1998) presented the application of their technique to univariate data. However, MT data contains at least four data channels, which depend on each other. Rehman and Mandic (2009) developed a new scheme, multivariate empirical mode decomposition (MEMD), to analyze multivariate signals and compute IMFs of each signal's components such that they maintain correlated 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. The instantaneous frequencies (IFs) obtained from IMFs with a similar scale is the same. We will extract the instantaneous parameter in the same frequency band based on the IFs later.

There are different ways to calculate IFs and the complex instantaneous coefficients 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. In this method, the direct quadrature but not the Hilbert transform is considered the imaginary part of the complex time-series. A detailed description of the method was given by Huang et al. (2009). In this way, we can get a set of the complex instantaneous coefficients using time-series data.

**Step 2**: Extract the IPs
Before the regression, the instantaneous parameters calculated by HHT are extracted out in
the same frequency band simultaneously. They were selected to calculate the impedance tensor,
according to Eqs.6 and 7. Neukirch et al. (2014) and Chen et al. (2012) describe how to extract
the IPs for the different frequency bands in detail. At first, we divide the frequency into different
frequency bands. As the impedance tensor changes smoothly with frequency, we can group the
similar frequency around a certain center frequency by the IFs. Extract the IPs 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:
\[
\mathbf{H}_{\text{hat}} = \mathbf{H}(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}.
\]
(8)

Chave (2003) suggested that the hat matrix's diagonal element, which is more than several times
2/N, is problematic. N denotes the number of data. Before the regression, the data where the
diagonal element of the hat matrix is more than two times 2/N should be removed by the selection
strategy to remove the influence of 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:
\[
\mathbf{y} = \mathbf{xR}_{\mathbf{x}} + \mathbf{H}_{\mathbf{x}}
\]
\[
\mathbf{Z} = \left(\mathbf{R}_{\mathbf{hx},\mathbf{hy}}\right)^{-1}\left(\mathbf{R}_{\mathbf{ex},\mathbf{ey}}\right).
\]
(9)

where \(\mathbf{y}\) denotes the N by one matrix \(E_x, E_y, H_x, H_y, H_z; \mathbf{x}\) denotes the N by two matrices \([\mathbf{H}_{\mathbf{x}}, \mathbf{H}_{\mathbf{yr}}]\)
(Here \([\mathbf{H}_{\mathbf{xr}}, \mathbf{H}_{\mathbf{yr}}]\) can be local \([\mathbf{H}_x, \mathbf{H}_y]\) or remote reference \([\mathbf{H}_{\mathbf{xr}}, \mathbf{H}_{\mathbf{yr}}]\)), \(\mathbf{R}_{\mathbf{x}}\) denotes the interstation
transfer function.
When \( x \) equals the local N by two matrices \([H_x, H_y]\), the \( R_{hx}, R_{hy} \) equal 1; and the 
\[
Z=(R_{ex}, R_{ey});
\]
in fact, there is no difference between Eqs. 6 and 7. The solution to \( R_x \) is a 
bi-variate complex linear regression problem. When \( y \) equals \( E_x \), the equation is equal to 
\[
E_x=H_xZ_{xx}+H_yZ_{xy}.
\]
extend the function to complex form: 
\[
E_{xr} + jE_{xi} = (H_{xr} + jH_{xi})(Z_{xx} + jZ_{xy}) + (H_{yr} + jH_{yi})(Z_{xy} + jZ_{yi}).
\] (10)
where the \( n \) denotes the number of the data, \( H_{xr} \) denote the \( n \)th real part of \( H_x \); \( H_{xi} \) denote the \( n \)th imaginary part of \( H_x \); it is the same to \( E_x, H_y \). \( Z_{xx} \) denotes the real part of \( Z_{xx} \); \( Z_{xy} \) 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:
\[
\begin{pmatrix}
E_{x1} \\
E_{x2} \\
\vdots \\
E_{xn}
\end{pmatrix}
=egin{pmatrix}
H_{x1} & -H_{x1} & H_{y1} & -H_{yi} \\
H_{x2} & -H_{x2} & H_{y2} & -H_{yi} \\
\vdots & \vdots & \vdots & \vdots \\
H_{xn} & -H_{xi} & H_{yr} & -H_{yi}
\end{pmatrix}
\begin{pmatrix}
Z_{xx} \\
Z_{xy}
\end{pmatrix}
\] (11) 

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 \( R_x \) by Eq. 3, following the M-
estimator. We have compared the two methods. The result is almost the same.  

When \( x \) equals the remote reference N by two matrices \([H_{xr}, H_{yr}]\). The interstation transfer 
function is calculated at first as follow:
\[
R_e = (H_r \dagger H_r)^{-1}(H_r \dagger E),
\]
\[
R_h = (H_r \dagger H_r)^{-1}(H_{ri} \dagger H).
\] (12) 

Then impedance tensor is calculated using the relationship between the local transfer function 
and interstation transfer function as follow:
\[
Z = R_h^{-1} R_e = (H_r \dagger H_r)^{-1}(H_r \dagger E).
\] (13)

It 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;

\[ x_1(t) = \sin(2\pi \cdot 50t) + \sin(2\pi \cdot 200t), \]
\[ x_2(t) = \sin(2\pi \cdot 25t) + \sin(2\pi \cdot 100t) + \sin(2\pi \cdot 250t). \]  

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. 1. The
HHT spectrum is shown in Fig. 2. The color denotes the value of 10\cdot\log_{10}(amp/max), where
"amp" denotes the amplitude of the IPs, "max" denotes the maximum amplitude of the IPs. Fig.2
shows the time variation of the spectrum calculated by HHT.
Fig. 1 The time series $x_1(t) = \sin(2\pi \cdot 50t) + \sin(2\pi \cdot 200t)$ and $x_2(t) = \sin(2\pi \cdot 25t) + \sin(2\pi \cdot 100t) + \sin(2\pi \cdot 250t)$. Both of the segments are added with 10% Gaussian noises.

To recover the digital data information, we should sample at least two points in one period, followed by the sampling theory. But to get the accurate complex coefficient from the time series, we suggest that it is better to sample 4 points in one period, and one continuous segment should contain at least four times the period. For instance, the 32-second continuous time series with the 1-second sampling rate. We can get a reliable complex coefficient between 4 to 8 second. If each segment satisfies this condition, we can combine the discontinuous segment directly, and the HHT can express the spectrum variation well. We considered that we could combine the discontinuous MT time-series data directly and then transform the time-series into the frequency domain by HHT if each segment satisfies the condition to get a reliable complex coefficient.
259 Fig. 2 The HHT spectrum of time-series data, which is shown in Fig. 1. The color denotes the value of $10 \cdot \log_{10}(\text{amp}/\text{max})$, 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 section, 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 station is one of Magnetic Observatory performs geomagnetic and geoelectric observations 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 $Z(\omega)$, calculated from the simple 1-D model, is multiplied with $H(\omega)$ to determine $E(\omega)$. 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. And the 2-hour time series data can get the maximum reliable period is about 300 seconds. 4-hour time series data is relatively good enough
to get the maximum reliable period is about 1,000 seconds. Fig. 3 shows the results calculated using 2-hour and 4-hour synthetic MT time series data. The light blue lines are the true model, which were 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 result of TE mode calculated using the different lengths of time-series at 874-second is shown in Fig.4. The red dotted lines denote the true model, which was calculated using a 1-D model. The horizontal axis denotes the length of the time-series data. 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 seconds. If there are approximately 4-hour noise-free time-series data, we could get reliable data at least with a period of 1,000 seconds by RMHHT.

Fig. 3 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. 4 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 denotes 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

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.5 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 |
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 follow:

$$E = E^{MT} + E^{CN} + E^{UN},$$
$$H = H^{MT} + H^{CN} + H^{UN}. \tag{15}$$

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, and artificial noise doesn't correlate with each other. 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. 6 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. 7 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. 6 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 midnight at site L6274 and L6276. The variation of the time series at other sites is almost keeping the same scale. Fig. 7 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 midnight. Each segment is ten minutes. The unit of the magnetic field is nT. The variation at site L6274 and L6276 is about 100 times than other sites in the daytime and a similar scale at midnight. That means the noise to signal is 100 in the daytime. It will bias the impedance tensor a lot.

Fig. 8 shows the Hx component's correlation variation between site Y0625(remote reference
site) and L6274 (local site) from 3:00:00 to 22:00:00 in UCT time on June 26. Each segment length is 10 minutes. The blue denotes a negative correlation, and red denotes a positive correlation. With the influence of local noise, the correlation of the magnetic field becomes low or negative in the daytime, and the correlation becomes higher at midnight. It means that the data in the local night is much quiet.

Fig. 8 The Hx component's correlation variation between Y0625 and L6274 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 denotes 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 ($\alpha_E$) and magnetic field ($\alpha_H$) (Fowler et al., 1967) at a specific frequency are defined as:

$$\alpha_E = \arctan \frac{2 \text{Re}[E_x E_y^*]}{[E_x E_x^*]-[E_y E_y^*]}$$  \hspace{1cm} (16)$$

$$\alpha_H = \arctan \frac{2 \text{Re}[H_x H_y^*]}{[H_x H_x^*]-[H_y H_y^*]}$$  \hspace{1cm} (17)$$

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.9. 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 midnight. The magnetic field's polarization direction 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 midnight. The signal to noise ratio is higher at midnight than in the daytime at all frequency bands in this data set. Therefore, we consider using the data at midnight only to calculate the impedance tensor.
Fig. 9 The polarization direction for a period of 6 seconds at site L7158 from 3:00:00 to 22:00:00 in UCT time. The upper figure shows the polarization directions of the electric field, and the lower is the polarization directions of the magnetic field.

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

For the conventional method, the windowed FFT was widely used to calculate the spectrum for a different frequency. The processing of broadband MT time-series data by HHT is demonstrated in this section. The data set observed using the MTU-A geophysical instruments is 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. Suppose each segment satisfies the situation to get an accurate complex coefficient. We can combine the intermittent time-series data directly, as the property of HHT mentioned in section 3.1. 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 contains about 64 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. Using the 150 Hz data calculate the period between 4/150 to 4/15 seconds. One continuous segment contains about 40 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. Using the 15-Hz data, 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. And we downsampled the continuous 15-Hz data to 1-Hz data 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 seconds. Combine all the data, we can calculate the full scall period from high frequency to 1,000 seconds by four hours.

This strategy allows us to select data in the time domain. There are always several hours at midnight 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 tensor.
Table 3 Strategy for the 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.10 shows the result calculated by site L6274. The RMHHT estimator calculated the black curves. The result was calculated by the midnight 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 midnight 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.
Fig. 10 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 calculated using midnight 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.

In the SSMT-2000 result, the phase of the TM component was close to 0° 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 midnight data is quite different for long periods, so the RMHHT and BIRRP results using midnight 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 midnight data is also noisy. On the other hand, the RMHHT and SSMT-2000 results are consistent and reliable at high-frequency periods.

Fig. 11 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 midnight data from 16:00:00 to 20:00:00
in UTC on June 26. The SSMT-2000 and BIRRP results calculated using the day time data have unnatural humps between 1 and 10 s in the apparent resistivities. On the other hands, BIRRP result calculated by the midnight 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 midnight in the dead band and is consistent with the background noise analysis in section 4.1.

Fig.12 shows the result derived by site Y0625. SSMT-2000 calculates the blue curves from one-day data. BIRRP calculates the red curves by midnight data from 16:00:00 to 20:00:00 in UTC on June 26. It is the same as shown in Fig.11. RMHHT calculates the black curves using the midnight 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 midnight data. But in the TM mode, the phase of RMHHT is changed unnaturally. We think it may be introduced by the noise in the local magnetic field (L7158) as the remote site. In the full-scale, the RMHHT perform well comparing with the BIRRP and SSMT-2000 result.

Fig. 11 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 midnight data from 16:00:00 to 20:00:00 in UTC on June 26.

Fig. 12 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 midnight data from 16:00:00 to 20:00:00 in UTC on June 26 and using L7158 as remote site. BIRRP was used to calculate the red curves using the midnight data from 16:00:00 to 20:00:00 in UTC on June 26.

Similar results were obtained at the L6178 and L7158. Fig. 13 and Fig. 14 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 midnight data, we can combine them directly. In this way, we can get more data at the impedance estimation step in the frequency domain. The result will become more stable.
Fig. 13 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 midnight data from 16:00:00 to 20:00:00 in UTC on June 26 and the Y0625 remote site. BIRRP calculates the red curves using the midnight data from 16:00:00 to 20:00:00 in UTC on June 26.

Fig. 14 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 midnight data from 16:00:00 to 20:00:00 in UTC on June 26 and the Y0625
remote site. BIRRP was used to calculate the red curves using the midnight data from 16:00:00 to 20:00:00 in UTC on June 26.

5 CONCLUSIONS

This paper showed a new strategy using a robust M-estimator based on HHT (RMHHT) to calculate the full-scale broadband MT time-series data containing the discontinuous high-frequency and middle-frequency band data. In the simulation using synthetic data sets, we have shown that we could get a reliable result until 1,000 seconds from continuous 4-hour noise-free time series data. At midnight, most of the electrical equipment is shutdown. Usually, the data at midnight is much quiet than in the daytime. This paper demonstrated that a time window of 4-hour length was almost noise-free during the local night at site L7158, L6178, and Y0625. We used data from this window alone to estimate impedances based on RMHHT. We compared the performance with the conventional method Bounded Influence Remote Reference Processing, BIRRP (Chave and Thomson, 2004), and the SSMT-2000 program and showed the property of this kind of technology. This kind of data processing technology allows us to select data in the time domain and just need about 4-hour quiet data to get a reliable result until 1,000 seconds. This short measurement time makes it possible to do MT exploration in strong electromagnetic noise areas. This strategy also decreased the computation time by down sampling the raw time series to calculate the long period data. Moreover, if there are several midnight time-series data, we can combine them. In this way, we can increase the number of data sets at the impedance estimation step in the frequency domain. The result will become more stable.

Declarations

Availability of data and materials

The magnetic time-series data used to construct the synthetic MT data is downloaded from the INTERMAGNET (International Real-time Magnetic Observatory Network).
Competing interests

We know of 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.

Acknowledgements

We thank Maik Neukirch for his constructive comments and 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
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.