

Comparison of three robust impedance estimators for the MT method

Hao Chen (✉ chenhaomagnetic@gmail.com)

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

Hideki Mizunaga

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

Toshiaki Tanaka

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

Gang Wang

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

Research Article

Keywords: magnetotelluric, response function estimator

Posted Date: January 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-147081/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Comparison of three robust impedance estimators for the MT method

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

Abstract: The initial magnetotelluric (MT) response function estimator is based on the least-square theory; it can be severely disturbed by the cultural noise. The different robust procedures have been developed and improve the performance of response function estimation dramatically. It is hard to say which method is better or not. In a specific situation, a different approach has different performance. Therefore, it is important to know the different property of them. Between the robust procedures, the robust M-estimator gives a small weight to reject the outlier based on the residual between the output (electric field) of the least-squares estimator and the observed data. M-estimator can reduce the influence of unusual data in the electric field (outliers) but are not sensitive to exceptional input (magnetic field) data, which are termed leverage points. The bounded influence (BI) estimator combines the standard robust M-estimator with leverage weighting based on the hat matrix diagonal element's statistics, a standard statistical measure of leverage point. Chave (2004) also creates an open-source code, Bounded Influence Remote Reference Processing (BIRRP), and it is widely used in the MT community. But the leverage point corresponds to the large variation of the magnetic field. It may be an energetic signal or active noise. The performance of the M-estimator and the BI-estimator was dramatically different in the two situations. On the other hand, the repeat median algorithm can protect against unusual data (outlier and leverage point) maximum. We researched the difference property of bound influence (BI-estimator), maximum likelihood (M-estimator), and repeat median (RM-estimator) signal site MT response function estimator. Three independent code (BIRRP code, robust M-estimator code, and RM-estimator code) are used to compare them. At last, two typical field data are used, making the difference between the bound influence estimator and robust M-estimator transparent. We found that when the leverage point is the energetic signal, the M-estimator will perform better than the BI-estimator. When the leverage point is the active noise, the BI-estimator will work much better than the M-estimator. Finally, we also investigated the ability of the three estimators at a single site.

1 INTRODUCTION

MT method is an electromagnetic geophysical method for inferring the earth's subsurface electrical conductivity from measurements of natural geomagnetic and geoelectric field variations at the earth's surface. There is a linear relationship between the geoelectric field and the geomagnetic field in the frequency domain. The first step in MT data processing is to estimate the frequency-dependent response function from the measured time-series data. All the MT data interpretation is based on the result of the response function. The initial technique for MT response function estimation is the least-squares (LS) estimator (Sims et al., 1971); it can be severely disturbed by the cultural noise. Through the decades, there are two main dramatic improvements by two development. The first is the remote reference technique (Gamble et al., 1979). The second is a robust procedure (Egbert and Booker, 1986; Chave and Thomson, 1989; Smimov, 2003; Chave and Thomson, 2004). The combination of them can obtain a more reliable result.

The measure of an estimator's robustness is its breakdown point. An estimator's breakdown point is the proportion of incorrect observations that an estimator can handle before

giving an incorrect result. In other words, the breakdown point may be roughly defined as the smallest percentage of the unusual data that may cause an estimator to collapse. The LS estimator's breakdown point is zero. That means that even a small amount of noise might affect the final result strongly.

In those robust estimators, the standard one is the maximum likelihood estimator (M-estimator). The robust M-estimator gives a small weight to reject the outlier based on the residual between the LS estimator's outputs (electric field) and the observed data. M-estimator can reduce the influence of unusual data in the electric field (outliers) but are not sensitive to exceptional input (magnetic field) data, which are termed leverage points. The bounded influence (BI) estimator combines the standard robust M-estimator with leverage weighting based on the hat matrix diagonal element's statistics, a standard statistical measure of leverage point. Chave (2004) showed that the BI estimator would perform better than the M-estimator.

But not all the case will be that the BI estimator work better than the M-estimator. The leverage point corresponds to the large variation of the magnetic field. It may be an energetic signal or active noise. The general speaking that the low-

frequencies signal (< 1 Hz) is originated from the interaction of the solar wind with the earth's magnetosphere and ionosphere. When there is a geomagnetic storm, the MT signal's strength will increase dramatically. Sometimes the signal strength can be two orders stronger than it in the non-storm period data. At this condition, the leverage point corresponds to the energetic signal. We need a high signal-to-noise ratio (SNR) data to obtain a reliable result at the noisy site. Suppose the BI-estimator rejects this kind of data. It will fail to get a reliable result. In this situation, the M-estimator will work better than BI-estimator. In general, the geomagnetic storm does not happen frequently. The active noise has a large variation and can be detected by the statistical of the hat matrix diagonal element. At this condition, the BI-estimator work better than M-estimator.

On the other hand, the repeat median algorithm can protect against unusual data (outlier and leverage point) maximum. In this paper, BI-estimator, M-estimator, and RM-estimator are compared to investigate the performances of them. Both M-estimator and RM-estimator codes are developed to compare with the BIRRP code. The detailed methods are described in Section 2. The new M-estimator and RM-estimator programs are examined using the synthetic datasets in section 3. The three estimators' performances are demonstrated using two field datasets in different situations that the datasets contain the energetic signal and active noise in section 4. Finally, we conclude that the M-estimator works better than the BI-estimator when the datasets include the energetic natural signal when there is a geomagnetic storm. When the data have active noise, the BI-estimator will work better. Finally, we also investigated the ability of the three estimators. The M-estimator and BI-estimator's breakdown point depends on the data's signal-to-noise ratio, and it is different in a different situation. The breakdown point of the RM-estimator is close to 50%.

2 THE DATA PROCESSING ALGORITHM

This section introduces three MT impedance estimators, the Bounded Influence Remote Reference Processing (BIRRP; Chave and Thomson, 2004), a standard M-estimator and the RM-estimator.

MT method is an electromagnetic geophysical method for inferring the earth's subsurface electrical conductivity from natural geomagnetic and geoelectric field variation measurements at the earth's surface. There is a linear relationship between the geoelectric field and the geomagnetic field in the frequency domain. The impedance tensor at a specific frequency can be calculated in the frequency domain as follow:

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

where E and H are the horizontal electric and magnetic field components at a specific frequency, and Z means the MT impedance. The x denotes the south-north direction; y denotes the east-west direction.

2.1 BIRRP

BIRRP is one kind of the typical robust estimator based on windowed FFT. It divides the original time series into the short segments at first. Each segment is whitened using an autoregressive filter, and the Fourier transforms of the sections are computed after tapering with a prolate window. The segment length is variable, and the selected frequencies within each segment are grouped to calculate the response function later. We will introduce this procedure detailly in 2.2.

The initial response function estimation technique is the least square estimator (Sims et al., 1971). The objective is to minimize the sum squares of the residual 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 \mathbf{E} and \mathbf{H} are the horizontal electric and magnetic field components at a specific frequency, and \mathbf{Z} is the MT impedance tensor. The superscript \dagger denotes the complex conjugate transpose. The M-estimator gives a weight w to reject the outlier based 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:

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

Here the weight is based on the Huber weight function; it is the minimum of $\{1, k/|r_i|\}$, and k is a positive constant, r_i is the i^{th} residual between the observed value and the predicted value. If the residual is bigger than k , the corresponding data is multiplied by the weight. The bigger the residuals are, the smaller the weight is. In this way, we can reduce the influence of the data with the big residual.

For the tuning constant k , the smaller value of k produces more resistance to outliers, but the efficiency is low if the errors are normally distributed. In general, $k = 1.5\sigma$ (σ is the standard deviation of the errors) produces 95% efficiency when the errors are normal. A common approach is to take $\sigma = \text{MAR}/0.44845$, where MAR is the median absolute residual calculated as follows: $\text{MAR} = \text{median}(|r_i - \text{median}(r_i)|)$.

The procedure of the M-estimator can be summarized as follows:

- (1) For the given N estimates of the horizontal electric and magnetic fields at a frequency, perform first regression by the least-squares estimator.
- (2) Compute the residuals r , and the tuning constant $k=1.5 \cdot (\text{MAD}/0.44845)$, and the initial residual squares sum of $\mathbf{r} \mathbf{r}^\dagger$, where the symbol \dagger denotes the complex conjugate transpose.
- (3) Compute the weights \mathbf{w} by Huber weight function and perform the regression by $\mathbf{Z} = (\mathbf{H}^\dagger \mathbf{w} \mathbf{H})^{-1} (\mathbf{H}^\dagger \mathbf{w} \mathbf{E})$. Then compute the squared sum of residuals of $\mathbf{r} \mathbf{w} \mathbf{r}^\dagger$.
- (4) Repeat steps 2 and 3 until the squared sum of residuals changes becomes less than 1%.

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 hat matrix is N by N matrix and

defined as follows:

$$\mathbf{H}_{hat} = \mathbf{H}(\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger. \quad (4)$$

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. The bounded influence estimator (BI-estimator) combines the robust M-estimator with leverage weighting (\mathbf{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:

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

where \mathbf{w} is the weight matrix based on a standard M-estimator, \mathbf{v} is the weight matrix based on the hat matrix diagonal elements' statistics. The BI-estimator performs the weighting procedure similar to the M-estimator iteratively. The detail of the BI-estimator was described by Chave (2012).

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

$$\mathbf{Z} = (\mathbf{H}_r^\dagger \mathbf{w} \mathbf{v} \mathbf{H})^{-1} (\mathbf{H}_r^\dagger \mathbf{w} \mathbf{v} \mathbf{E}), \quad (6)$$

where \mathbf{H}_r denotes the remote magnetic field data. The BIRRP is the abbreviation of the Bounded Influence Remote Reference Processing method.

The jackknife algorithm is used to calculate the error bar at a specified frequency. The jackknife algorithm 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 in the dataset.

2.2 Robust M-estimator

2.2.1 Spectrum estimation

The cascade decimation proposed by Wight and Bostick (1980) was widely used to calculate the spectrum for different frequencies. This scheme has a significant advantage: it can compute the spectrum fast, produces a spectrum representing constant percentage bandwidth, and is evenly spaced on a log-frequency scale. It divides the original time-series data into segments with a specific window length (32, 128, 256 or 512). The selected frequencies within each segment are grouped to calculate the impedance tensor in the frequency domain. The original time-series data are then passed through a cascaded sequence of low pass filters and downsampled by two. The sampling frequency is reduced by two whenever the level raises one stage. It divides the downsampled time-series data into the same window length segments. The same parameter within each segment is extracted to calculate the impedance tensor at a different frequency.

A method similar to BIRRP was adopted to estimate the spectrum instead of downsampling the original time-series data. It divides the original time-series data into segments with the longest window. The segment length is variable with each level. The 5th and 8th number of each segment's frequency coefficient are grouped to calculate each level's response function. The segment length is then reduced by 2, and the process is repeated to calculate the impedance tensor at different frequencies.

For the real MT time-series, e_x, e_y, h_x, h_y, h_z , the electric field unit is transformed into mv/km; the magnetic field unit is transformed into nT. For each level, the procedure to produce the accurate spectrum for each section as follow:

(1) Using the first difference filter to remove the long-period trends and means, in the case of e_x , yielding the new series:

$$\widetilde{e}_x(i) = e_x(i+1) - e_x(i) \quad (7)$$

(2) To reduce the bias of spectral estimation, a multi-tapering method is adopted. Multitapering process is an extension of single taper approaches. The multi-tapering method can be carried out by the Matlab intrinsic function "dpss.m". In this study, we specify the half bandwidth product to be 2.5 and multi-tapering to be 4. It produces four orthogonal discrete prolate spheroidals (Slepian) sequences windows and a column weight vector (lambda), the lambda equal in the length to the number of Slepian sequences. Tapper the 4 Slepian sequences to each segment and get four time-series sets \overline{e}_x .

(3) Fourier transform is carried out to each \overline{e}_x , and calculate the weighted mean by multiplying the lambda to the four Fourier coefficient sets, yielding the Fourier coefficients $E_x[i]$, where $i=1, 2, \dots, N/2$ is the frequency number. The other components (E_y, H_x, H_y, H_z) perform the same procedure. The 5th and 8th number of the frequency coefficient within each segment is grouped to do further processing at each level.

2.2.2 Robust estimation for the response function

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

$$\mathbf{y} = \mathbf{x} \mathbf{R}_x + \boldsymbol{\sigma}_x, \\ \mathbf{Z} = (\mathbf{R}_{hx}, \mathbf{R}_{hy})^{-1} \cdot (\mathbf{R}_{ex}, \mathbf{R}_{ey}), \quad (8)$$

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 the $[H_{xr}, H_{yr}]$; Here $[H_{xr}, H_{yr}]$ can be local $[H_x, H_y]$ or remote reference $[H_{xr}, H_{yr}]$; \mathbf{R}_x denotes the interstation transfer function.

When \mathbf{x} equals the local N by two matrices $[H_x, H_y]$, the $\mathbf{R}_{hx}, \mathbf{R}_{hy}$ equal 1; and the $\mathbf{Z} = (\mathbf{R}_{ex}, \mathbf{R}_{ey})$; in fact, there is no difference between Eqs. 1. The solution to \mathbf{R}_x is a bivariate complex linear regression problem. When \mathbf{y} equals E_x , the equation is equal to $E_x = H_x Z_{xx} + H_y Z_{xy}$, extend the function to complex form:

$$E_{xr_n} + jE_{xi_n} = (H_{xr_n} + jH_{xi_n})(Z_{xxr} + jZ_{xxi}) \\ + (H_{yr_n} + jH_{yi_n})(Z_{xyr} + jZ_{xyi}), \quad (9)$$

where the n denotes the 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. We can transform it into matrix form as follows:

$$\begin{pmatrix} E_{xr_1} \\ E_{xr_2} \\ \vdots \\ E_{xr_n} \\ E_{xi_1} \\ E_{xi_2} \\ \vdots \\ E_{xi_n} \end{pmatrix} = \begin{pmatrix} H_{xr_1} & -H_{xi_1} & H_{yr_1} & -H_{yi_1} \\ H_{xr_2} & -H_{xi_2} & H_{yr_2} & -H_{yi_2} \\ \vdots & \vdots & \vdots & \vdots \\ H_{xr_n} & -H_{xi_n} & H_{yr_n} & -H_{yi_n} \\ H_{xi_1} & H_{xr_1} & H_{yi_1} & H_{yr_1} \\ H_{xi_2} & H_{xr_2} & H_{yi_2} & H_{yr_2} \\ \vdots & \vdots & \vdots & \vdots \\ H_{xi_n} & H_{xr_n} & H_{yi_n} & H_{yr_n} \end{pmatrix} \begin{pmatrix} Z_{xxr} \\ Z_{xxi} \\ Z_{xyr} \\ Z_{xyi} \end{pmatrix} \quad (10)$$

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 \mathbf{R}_x by Eq. 3, following the M-estimator. We compared the two methods. The result was almost the same. Both of them can give a weight \mathbf{w} to reject the outlier in the electric field.

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

$$\begin{aligned} \mathbf{R}_e &= (\mathbf{H}_r^\dagger \mathbf{H}_r)^{-1} (\mathbf{H}_r^\dagger \mathbf{E}), \\ \mathbf{R}_h &= (\mathbf{H}_r^\dagger \mathbf{H}_r)^{-1} (\mathbf{H}_r^\dagger \mathbf{H}). \end{aligned} \quad (11)$$

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

$$\mathbf{Z} = \mathbf{R}_h^{-1} \mathbf{R}_e = (\mathbf{H}_r^\dagger \mathbf{H}_r)^{-1} (\mathbf{H}_r^\dagger \mathbf{E}). \quad (12)$$

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 noises with the local site, it can 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 the uncertainty.

2.3 Repeated Median Estimator

The least-squares method can give its optimum solution if the sum of the squared residuals is minimum. The residual is defined by the difference between the observed electric field and the electric field E_p predicted from $E_p = \mathbf{Z}\mathbf{H}$, where \mathbf{Z} is obtained by LS estimator. The least-squares estimator (LS-estimator) is very sensitive to gross errors in the data and has a breakdown point equal to zero.

M-estimator is much less sensitive to outliers than the LS-estimator. Siegel's repeated median estimator indicates the highest possible breakdown point equal to 50 %. If there are no more than half of the data contaminated, the repeated median estimator will produce a reliable result.

The repeated median estimator (RM-estimator) was proposed by Siegel (1982). It is helpful to consider the simple linear regression model $y_i = b + a x_i + e_i$, to explain the estimator in detail. The RM-estimator for the slope is then given by:

$$a = \underset{i \neq j}{\text{med med}} \left(\frac{y_j - y_i}{x_j - x_i} \right) \quad (13)$$

where $i, j = 1, \dots, N$; when $j = i$, the inner median is skipped. N

denotes the number of the point. For the solution of the slop, two points can determine a result. The inner median result for one point is the median of all the combinations with other points. The final result of the slop is the median of the first median.

By the situation of MT, if there are two datasets, we can get a result. The real and imaginary part of each impedance tensor is calculated separately. The repeat median operator is given by:

$$\text{Re}\{Z\} = \underset{i \neq j}{\text{med med}} \text{Re} (H_{ij}^\dagger * H_{ij})^{-1} (H_{ij}^\dagger * E_{ij}),$$

$$\text{Im}\{Z\} = \underset{i \neq j}{\text{med med}} \text{Im} (H_{ij}^\dagger * H_{ij})^{-1} (H_{ij}^\dagger * E_{ij}), \quad (14)$$

where the indices $i, j = 1, \dots, N$. The inner median skips when $j = i$.

The Siegel repeat median is calculated as follow:

(1) The first median operator of eq. 14 is calculated as:

$$\begin{aligned} S1_{\text{Re}\{Z\}_i} &= \underset{j \neq i}{\text{med}} \text{Re} (H_{ij}^\dagger H_{ij})^{-1} (H_{ij}^\dagger E_{ij}); \\ S1_{\text{Im}\{Z\}_i} &= \underset{j \neq i}{\text{med}} \text{Im} (H_{ij}^\dagger H_{ij})^{-1} (H_{ij}^\dagger E_{ij}); \end{aligned} \quad (15)$$

The $Z_i = S1_{\text{Re}\{Z\}_i} + i^* S1_{\text{Im}\{Z\}_i}$, where Z_i denotes the i^{th} result of the inner median.

(2) The second median operator is calculated as:

$$\begin{aligned} S2_{\text{Re}\{Z\}} &= \underset{i}{\text{med}} S1_{\text{Re}\{Z\}_i}; \\ S2_{\text{Im}\{Z\}} &= \underset{i}{\text{med}} S1_{\text{Im}\{Z\}_i}; \end{aligned} \quad (16)$$

The $Z_s = S2_{\text{Re}\{Z\}} + i \cdot S2_{\text{Im}\{Z\}}$, where Z_s denotes the result of impedance calculate by Siegel RM-estimator,

(3) Confidence limits are estimated from the median of absolute deviations (MAD):

$$Z_{\text{MAD}} = 1.483 \underset{i}{\text{med}} \{Z_i - Z_s\}. \quad (17)$$

This estimate of the scale parameter is straightforward to calculate and is insensitive to outliers. For Gaussian errors, a 95 % confidence limit can be defined as (error bar):

$$\Delta Z = 1.96 \cdot Z_{\text{MAD}} / \sqrt[3]{N}, \quad (18)$$

where N is the number of data.

3 COMPARISON USING SYNTHETIC DATA

To assess the performance of the Robust M-estimator and RM-estimator. We use the synthetic data with a known impedance Z to test the new code. The way to create synthetic data is similar to the method proposed by Chen (2012). We used the 1-day magnetic data from Memambetsu (MMB) station as the synthetic magnetic time series data. MMB station is one of Magnetic Observatory performs geomagnetic and geoelectric observations in Japan. The sampling rate of synthetic data is 1 second, and its unit is nT. The impedances $Z(\omega)$ are calculated from our simple 1-D model, and then transform the magnetic field time-series ($h(t)$) into the frequency domain by using the Fourier transform. Then the impedance $Z(\omega)$ 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 $e(t)$. Fig.

1 compares the real model's impedance curves and the result calculated from the time-series data using the BIRRP code, the M-estimator and the RM-estimator. All of them coincide with each other.

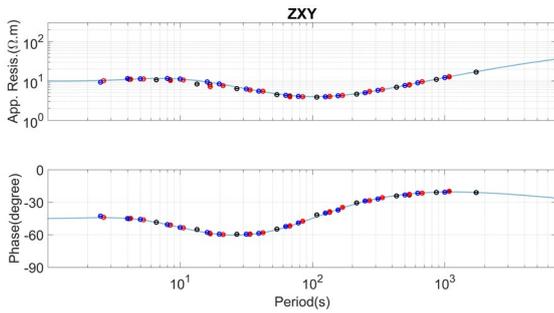


Fig. 1 The comparison of the impedance curves between the true model and the result calculated from the time series. The blue line is the true curve of a simple 1-D model. Blue circles were calculated using the BI-estimator, and red circles were calculated using the M-estimator. Black circles were calculated using the RM estimator.

4 APPLICATION TO THE FIELD DATA

The most significant difference between the M-estimator and the BI-estimator is the leverage weighting. The leverage point is the energetic magnetic field data. It can be an energetic signal or active noise. We used two datasets to show the two different situations. In the first dataset, the signal has a large variation when there is a geomagnetic storm. In the second dataset, the noise has a large variation.

4.1 Application for KAP03

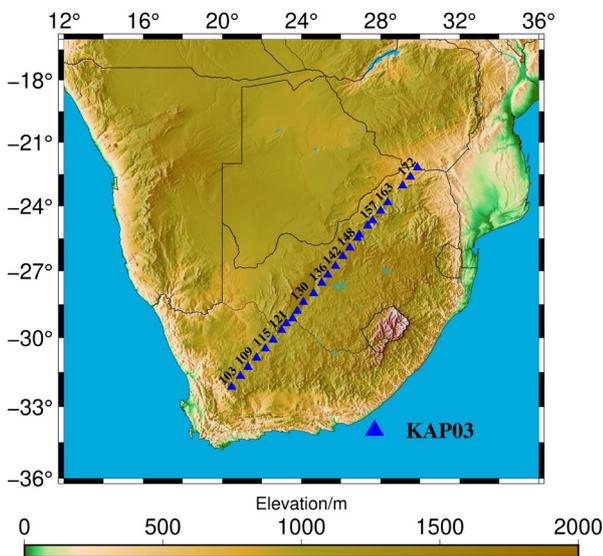


Fig. 2 Location map of the survey line KAP03 in South Africa.

In the first dataset, the Kaapvaal 2003 (KAP03) long-period 5-component magnetotelluric time series data is used. Data were recorded with a 5-second sampling rate for around a month at each site using GSC LIMS systems in 2003 as a part of the SAMTEX project. The 26 long-period sites distributed in a NE-SW profile are shown in Fig.2. In the middle of the survey line (KAP127-KAP145), datasets were heavily contaminated by DC noises from the power line of trains

running between Kimberley and Johannesburg. There are two geomagnetic storm events in the observation periods. Both of the storms lasted about two days. The first storm occurred around Oct. 29, 2003, to Oct. 31, 2003, the second storm occurred around Nov. 20, 2003, to Nov. 22, 2003. We used the different period time-series dataset at noisy sites 130 and 133 to analyze the three estimators' performance. We show the result derived from site 130 first.

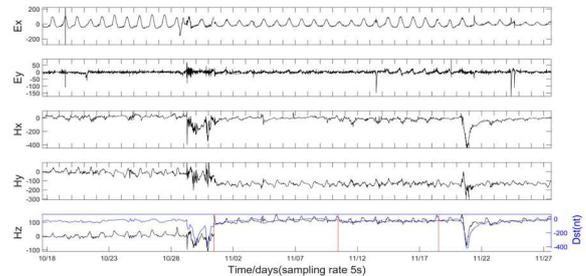


Fig.3 Time series of magnetotelluric data whose sampling rate is 5 seconds at site 130. The red vertical lines show the missing data section, and the black lines show the 5-component MT time series data. The blue line shows the Dst index of magnetic data. The unit of the electric field is mV/km, and the magnetic field in nT.

Fig.3 shows the MT time-series data at site 130. The red vertical lines indicate the missing data section, and the black lines indicate the 5-component MT time-series data. The sampling rate is 5 seconds. The unit of the electric field is mV/km, and the magnetic field in nT. The blue line shows the Dst (disturbance storm time) index.

Fig.4 and Fig.5 show the XY and YX component of magnetotelluric sounding curves calculated during both storm day and the non-storm day at site 130. The first column is calculated in the non-storm day data from 00:00:00 on Oct. 22 to 00:00:00 on Oct. 25. The second column is computed in the first storm from 00:00:00 on Oct. 29 to 00:00:00 on Oct. 31. The third column is calculated in the second storm data 00:00:00 on Nov. 20 to 00:00:00 on Nov. 22. M-estimator calculates the red curve, and BI-estimator calculates the black curve. The upper figures show the apparent resistivity. The lower figures show the impedance phase. Using the non-storm data, both of the methods failed to recover the result. The phase of the YX and XY component was close to 180° or 0° . That is the phenomenon of artificial noise (Zonge and Hughes, 1987); 180° or 0° would correspond to a dipole electric source, which could be the train line. And the long period of the XY component is very scattered.

In the XY component result in Fig.4, we think the period below 200-second is improved and gets a reliable result using the storm data. Coming back to Fig.3, the time-series data at site 130, we find the sine signal dominates the Ex. In general, the natural MT signal usually shows as a random waveform. The continuous noise was too strong to get a reliable result in the long period data, even if we used the geomagnetic storm data.

Comparing all the results, we find that the YX component can be improved using storm period data. In Fig.5, the first

storm period data calculates the second column. Both curves are similar to each other. After comparing the result with the remote reference result, we think that the first geomagnetic storm's result is the true curve. The third column is calculated by the second storm period from 00:00:00 on Nov. 20 to 00:00:00 on Nov. 22. M-estimator result was considered close to the true curve comparing to the result derived from the first geomagnetic storm. But the result calculated by the BI-estimator is different from the M-estimator result. The phase is also close to 180° . In general, the BI-estimator performs better than the M-estimator. But in this case, it is not.

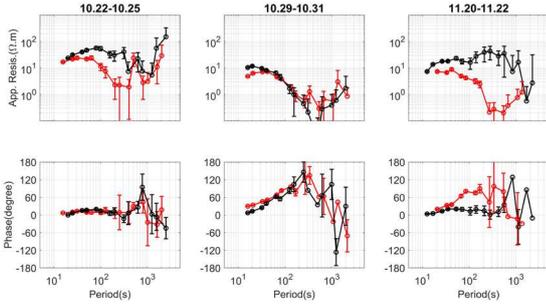


Fig.4 The XY component of magnetotelluric sounding curves calculated during both storm day and the non-storm day at site 130. The M-estimator calculated the red curves, and the BI-estimator calculated the black curves. The left figures were calculated using the non-storm data, and the central figures were calculated using the second storm data. The right figures were calculated using the third storm data. The upper figures show the apparent resistivity curves, and the lower figures show the impedance phase curves.

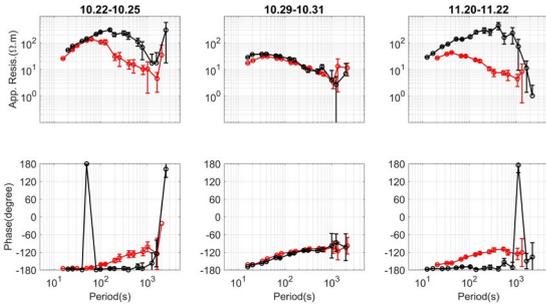


Fig.5 The YX component of magnetotelluric sounding curves calculated during both storm day and the non-storm day at site 130. The colors are the same as Fig.4.

Next, we tried to introduce some non-storm noisy data to the first geomagnetic storm dataset to compare BI-estimator, M-estimator, and RM-estimator's performances. Fig. 6 shows the YX component of magnetotelluric sounding curves. The left figures were calculated using the BIRRP. The center figures were calculated using the M-estimator. The right figured were calculated using the RM-estimator. The blue curves were calculated using the dataset from Oct. 27 to Oct. 31. The black curves were calculated using the dataset from Oct. 28 to Oct. 31. The red curves were calculated using the dataset from Oct. 29 to Oct. 31. In the three datasets, all of them contain the geomagnetic storm data from Oct. 29 to Oct.

31. The BI results get biased with the longer non-storm data introduced. But M-estimator results almost don't change and are reliable. We think two factors lead to the results.

The most significant difference between the BI-estimator and M-estimator is the leverage weighting. We investigated the hat matrix diagonal element value situation, as shown in Fig.7. The leverage points are corresponding to the geomagnetic storm data. The expectation value of the hat matrix diagonal element is $2/N$, N denotes the number of the data; the more the data is, the smaller the expectation value is. That also means the more the leverage point will be detected. In this case, all of the leverage points are correspond to the geomagnetic event. For the noisy site, we need the geomagnetic storm period data to get a reliable result. But BI-estimator will give a much small weight to reject this kind of data. That is the first factor. The other factor is, the more the non-storm data introduce, the more noise data is introduced. The BI-estimator will deviate further from the true curve with the more non-storm data introduced by the two factors.

On the other hand, the M-estimator is not sensitive to the leverage point. As the solution of least square is based on averaging of the auto and cross power density spectra, for the south-north direction, the function can be written as follow:

$$\begin{aligned} \langle E_x H_x^\dagger \rangle &= Z_{xx} \langle H_x H_x^\dagger \rangle + Z_{xy} \langle H_y H_x^\dagger \rangle, \\ \langle E_x H_y^\dagger \rangle &= Z_{xx} \langle H_x H_y^\dagger \rangle + Z_{xy} \langle H_y H_y^\dagger \rangle, \end{aligned} \quad (13)$$

where $\langle \rangle$ denotes the average of the auto-power and cross-power density spectra. In this function, the impedances are dominated by the energetic auto-power density $\langle H_x H_x^\dagger \rangle$ and $\langle H_y H_y^\dagger \rangle$. Therefore, even the non-storm noisy data is introduced, but the energetic storm period data will dominant the final result. And the M-estimator is not biased with introducing the non-storm noisy data.

The RM-estimator can obtain a reliable result by the data from Oct. 29 to Oct. 31 and Oct. 28 to Oct. 31. In contrast, it gets biased after introduced 2-day non-storm noisy data into the geomagnetic data from Oct. 29 to Oct. 31. We thought that the data contain about one and a half-day high signal-to-noise data from Oct. 29 to Oct. 31. In this dataset, the M-estimator performed best. The BI-estimator performed worst.

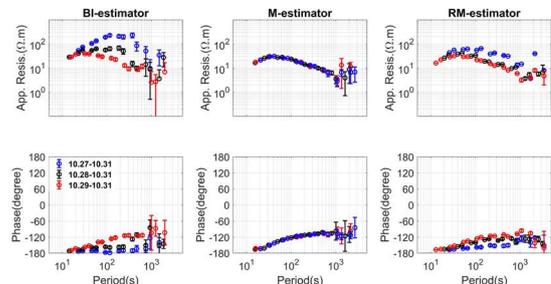


Fig.6 The YX component of magnetotelluric sounding curves calculated using the different periods around the first geomagnetic storm at site 130. The left figures were calculated using the BIRRP. The center figures were calculated using the M-estimator. The right figured were calculated using the RM-estimator. The blue curves were calculated using the dataset from Oct. 27 to Oct. 31. The black curves were calculated using the dataset from Oct. 28 to Oct. 31. The red curves were

calculated using the dataset from Oct. 29 to Oct. 31.

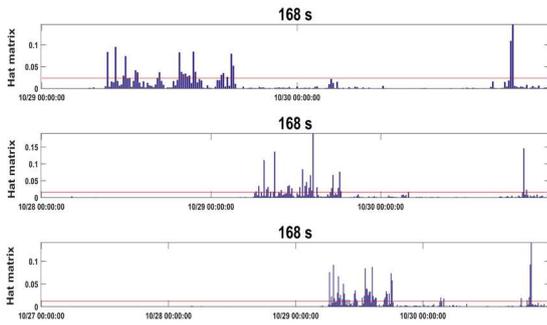


Fig.7 The hat matrix diagonal element distribution at period 168 second. The upper figure shows the data from Oct. 29 to Oct. 31. The central figure shows the data from Oct. 28 to Oct. 31. The lower figure shows the data from Oct. 27 to Oct. 31. The red dotted lines denote the three times of the $2/N$.

Moreover, Fig.8 shows the time-series data of the second geomagnetic storm from 00:00:00 on Nov. 20 to 00:00:00 on Nov. 22. The first and third figures are the electric (e_y) and magnetic field (h_x), respectively. The second and fourth figures are the first directive of the electric and magnetic field, respectively. There is a significant transient change around 27th and 33th hour in the electric field due to the non-magnetotelluric source. And the geomagnetic storm almost focuses on from the 6th to 24th hours without any obvious noise data.

Fig. 9 shows the YX component of magnetotelluric sounding curves calculated using the different periods around the second geomagnetic storm at site 130. The black curves were calculated using the data from 06:00:00 on Nov. 20 to 00:00:00 on Nov. 21. The red curves were calculated using the data from 00:00:00 on Nov. 20 to 00:00:00 on Nov. 22. The RM-estimator could obtain a reliable result using the data from 06:00:00 on Nov. 20 to 00:00:00 on Nov. 21. But a reliable result could not be obtained using the data from 00:00:00 on Nov. 20 to 00:00:00 on Nov. 22. The time-series data around 27th and 33rd hours in the electric field have a great bias due to the non-magnetotelluric source, as shown in Fig. 8. We think the noisy data focus on the data from 00:00:00 on Nov. 21 to 00:00:00 on Nov. 22. The BI-estimator could obtain a reliable result using the geomagnetic storm data from 06:00:00 on Nov. 20 to 00:00:00 on Nov. 21. Still, it failed to obtain a reliable result after introduced the time series data from Nov. 21 to Nov. 22. The same two factors as we analyzed in Fig.6 lead to this result. On the other hand, the M-estimator don't change and get a reliable result. Because of the energetic geomagnetic storm data dominant the final result.

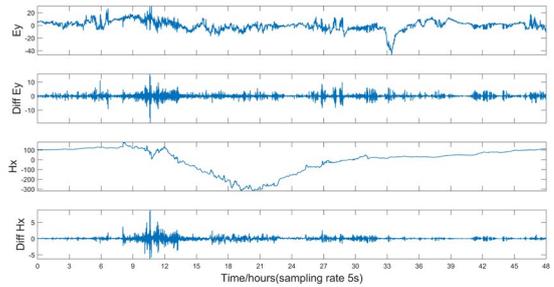


Fig.8 MT time-series data at site 130 from Nov. 20 to Nov. 22. The top figure is the e_y time-series. The second figure is the first derivative of e_y . The third figure is the h_x time-series. The bottom figure is the first derivative of h_x . The unit of the electric field is mV/km, and the unit of the magnetic field is nT.

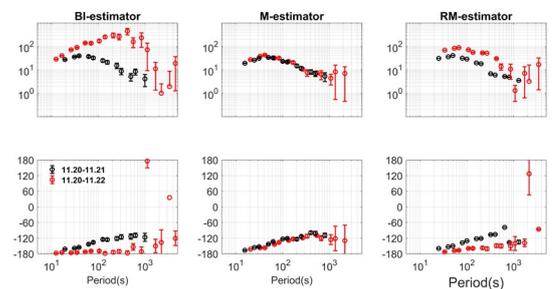


Fig.9 The YX component of magnetotelluric sounding curves calculated using the different periods around the second geomagnetic storm at site 130. The left, middle and right figures were calculated using the BI-estimator, the M-estimator and the RM-estimator. The black curves were calculated using the data from 06:00:00 on Nov. 20 to 00:00:00 on Nov. 21. The red curves were calculated using the data from 00:00:00 on Nov. 20 to 00:00:00 on Nov. 22.

Fig. 10 and Fig.11 show the XY and YX components of MT sounding curves calculated using the different periods at site 133, respectively. The red curves were calculated using the geomagnetic storm data from 00:00:00 on Oct. 29 to 00:00:00 on Oct. 31. The black curves were calculated using the data from 00:00:00 on Oct. 28 to 00:00:00 on Oct. 31. The blue curves were calculated using the data from 00:00:00 on Oct. 27 to 00:00:00 on Oct. 31. The green curves were calculated using the non-storm data from 00:00:00 on Oct. 26 to 00:00:00 on Oct. 29. We interpreted the result calculated using the geomagnetic storm data from Oct. 29 to Oct. 31 as a preferable curve. All BI-estimator, M-estimator, and RM-estimator failed to recover the result using the non-storm data (green curve). The apparent resistivity gets biased, comparing the result computed from the storm period. Compared to the result calculated using different period data around the first geomagnetic storm. In Fig.10, the RM-estimator becomes deviated from the true model after introduced one-day non-storm data from Oct. 28 to Oct. 31. That means the SNR between the non-storm data is low. There are around one day high SNR data from Oct. 29 to Oct. 31 in the XY component.

In Fig.11, the RM-estimator becomes deviated from the true model after introduced two-day non-storm data from Oct.

27 to Oct. 31. There is around one and a half-day high SNR data from Oct. 28 to Oct. 31 in the YX component. The BI-estimator becomes deviated from the true model further with introduced more non-storm noisy data. On the other hand, the M-estimator perform robustly and the result covered with each other. That is caused by two factors, as we analyzed in Fig. 6.

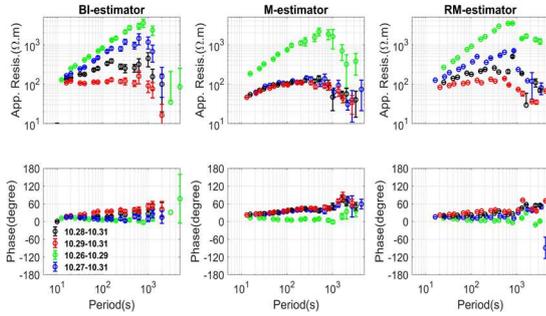


Fig. 10 The XY component of MT sounding curves calculated from the different period data at site 133. The left figures were calculated using the BI-estimator. The central figures were calculated using the M-estimator. The right figures were calculated using the M-estimator. The red curves were calculated using the period from 00:00:00 on Oct. 29 to 00:00:00 on Oct. 31. The black curves were calculated using the period from 00:00:00 on Oct. 28 to 00:00:00 on Oct. 31. The blue curves were calculated using the period from 00:00:00 on Oct. 27 to 00:00:00 on Oct. 31. The green curves were calculated using the period from 00:00:00 on Oct. 26 to 00:00:00 on Oct. 31.

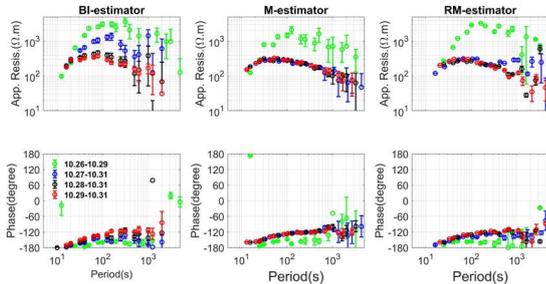


Fig. 11 The YX component of MT sounding curves calculated from different period data at site 133. The colors are the same as Fig. 10.

4.2 Application for L6-7

In the second dataset, 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. Fig.12 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 inset map shows the detail of survey line L7-1; The lower right inset map shows the detail of survey line L6-1; The upper right inset map shows the detail of survey line L6-2.

At first, we introduced polarization directions to estimate the background noise. 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:

$$\alpha_E = \arctan \frac{2 \cdot \text{Re}([E_x E_y^\dagger])}{[E_x E_x^\dagger] - [E_y E_y^\dagger]}, \quad (13)$$

$$\alpha_H = \arctan \frac{2 \cdot \text{Re}([H_x H_y^\dagger])}{[H_x H_x^\dagger] - [H_y H_y^\dagger]}. \quad (14)$$

The polarization directions describe the time-harmonic variation of two orthogonal field components with constant phasing. A variety of sources generates natural magnetic signals. 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. The intensities of natural electromagnetic fields are very weak at the dead band (1-10s); it is easy to influence the local noise. The typical polarization directions for the dead band is shown in Fig.13. Fig.13 shows the result of polarization directions calculated using 6-second data at site L7158. 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.

On the other hand, the polarization directions of the electric field changed around 16:00:00 in UTC. We thought the magnetic field was quiet from 15:00:00 to 16:00:00, but the electric field was noisy. This situation was categorized into the outlier (the unusual in the electric field). In this research, we used the data observed at site L7158 on June 26, 2016, to compare the three estimators in the second situation when the noise is energetic.

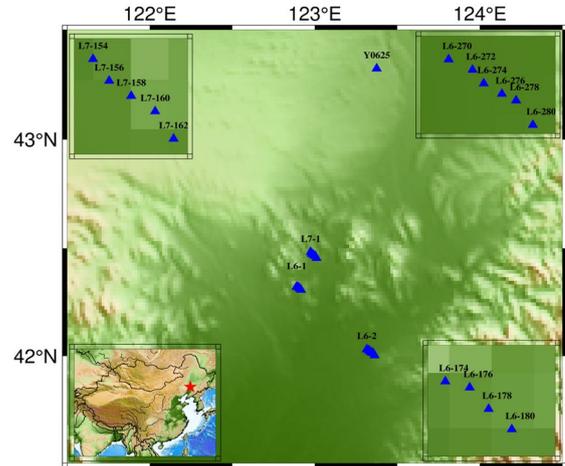


Fig. 12 Location map of the study area. Y0625 is the remote reference site, and L6-1, L6-2, L7-1 in the middle are the local survey lines. The lower left map shows the survey area in China. The red star denotes the research location. The upper left map shows the detail of survey line L7-1; The lower right map shows the detail of survey line L6-1; The upper right map shows the detail of survey line L6-2.

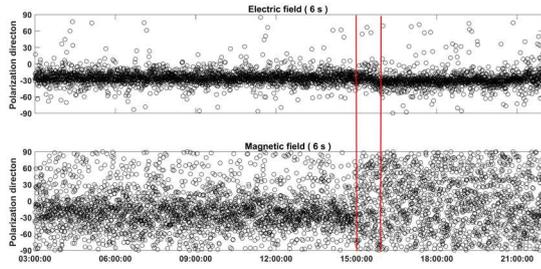


Fig. 13 Polarization direction for period 6 second at site L7158 from 3:00:00 to 22:00:00 in UCT. The upper figure shows the electric field's polarization directions, and the lower figure shows the magnetic field's polarization direction. The red vertical lines denote the boundary of the property of the polarization directions changed.

At first, we get the MT sounding curve from the Phoenix software SSMT-2000 from the one-day data. We find that the result of the dead band is biased. We then calculate the time series data by BI-estimator, M-estimator, and RM-estimator using the midnight data from 16:00:00 to 20:00:00. The black curves were calculated using the M-estimator. The red curves were calculated using the BI-estimator. The green curves were calculated using the RM-estimator. These results are shown in Fig. 14. The sounding curves become reliable after calculating the midnight data.

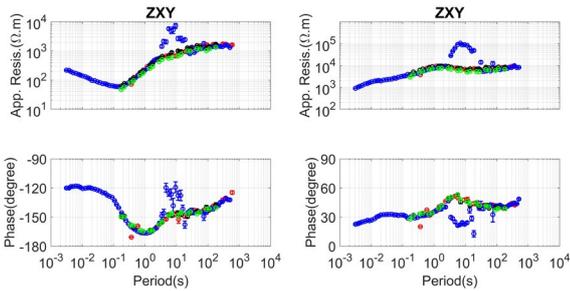


Fig. 14 The magnetotelluric sounding curves calculated using the data at site L7158. SSMT-2000 calculated the blue curves using one-day data. M-estimator calculated the black curves using the midnight data from 16:00:00 to 20:00:00 in UTC. BI-estimator calculated the red curves using the midnight data from 16:00:00 to 20:00:00 in UTC. RM-estimator calculated the green curves using the midnight data from 16:00:00 to 20:00:00 in UTC.

At the next step, we investigated the hat matrix diagonal element. Fig.15 shows the hat matrix diagonal element distribution from 14:00:00 to 20:00:00 in UTC. The magnetic noise was strong from 14:00:00 to 15:00:00, and it was relatively quiet at midnight. We tested the BI-estimator, M-estimator and RM-estimator by introducing different lengths of noise data.

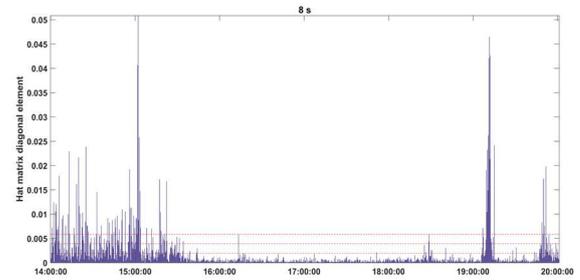


Fig. 15 The hat matrix diagonal element distribution from 14:00:00 to 20:00:00 in UTC at the period of 8 seconds. The red dotted lines denote the one, two, three times of the $2/N$, respectively.

Fig.16 shows the magnetotelluric sounding curves calculated by RM-estimator using different period data at site L7158. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The green curves were calculated using the data from 15:00:00 to 20:00:00 in UTC. The black curves were calculated using the data from 14:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 13:00:00 to 20:00:00 in UTC. The RM-estimator result becomes biased by the data from 14:00:00 to 20:00:00. The breakdown point of RM-estimator is 50%. That means there was about 3-hour quiet data at midnight. And the quiet data focuses on the midnight from 16:00:00 to 20:00:00.

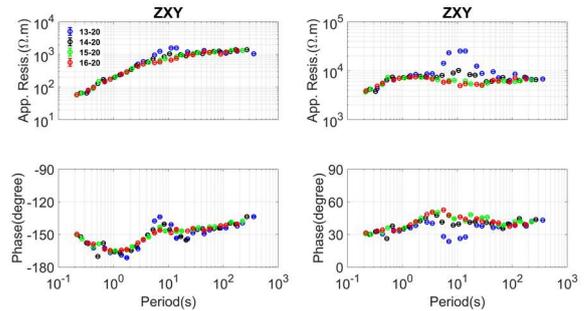


Fig. 16 The magnetotelluric sounding curves calculated by RM-estimator using different period data at site L7158. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The black curves were calculated using the data from 14:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 13:00:00 to 20:00:00 in UTC.

Fig. 17 shows the result calculated by the M-estimator. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The black curves were calculated using the data from 15:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 14:00:00 to 20:00:00 in UTC. After introducing a one-hour outlier in the electric field, the M-estimator also could work effectively. But after introducing one more hour of leverage point data between 14:00:00 to 15:00:00, the M-estimator gave a biased result in the dead band. This study shows that the M-estimator is not sensitive to the leverage point in the magnetic field and the leverage point that dominates the final result.

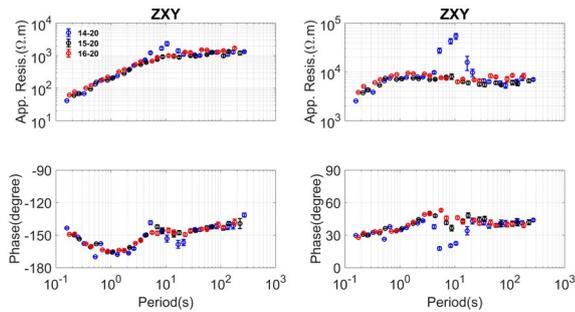


Fig. 17 The magnetotelluric sounding curves calculated by M-estimator using different period the data at site L7158. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The black curves were calculated using the midnight data from 15:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 14:00:00 to 20:00:00 in UTC.

Fig. 18 shows the result calculated by BI-estimator. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The black curves were calculated using the data from 14:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 13:00:00 to 20:00:00 in UTC. Compared the result calculated by the BI-estimator and M-estimator by the data from 14:00:00 to 20:00:00 in UTC, the BI-estimator succeeded to get a reliable result while the M-estimator and RM-estimator failed. It shows that when the leverage point is the active noise, the BI-estimator will work better than M-estimator.

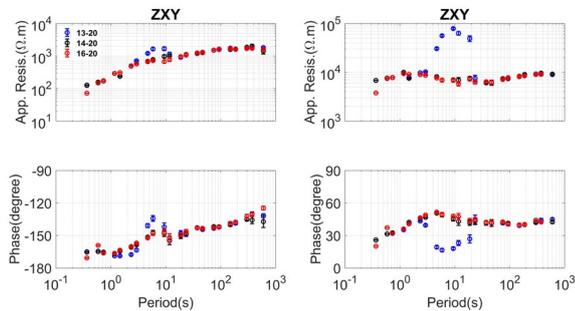


Fig. 18 The magnetotelluric sounding curves calculated by the BI- estimator using different period data at site L7158. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The black curves were calculated using the data from 14:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 13:00:00 to 20:00:00 in UTC.

5 CONCLUSION AND DISCUSSION

In this paper, we compared three impedance estimators; the M-estimator, BI-estimator and RM-estimator. At first, we introduced BI-estimator, M-estimator and RM-estimator created by ourselves without any additional processing. And then, we test the performance of the M-estimator and RM-estimator using the synthetic noise-free data. At last, we used two typical datasets to compare M-estimator, BI-estimator, and RM-estimator performances. As well known that the breakdown point of the least-squares estimator is 0; the breakdown point of the RM-estimator is close to 50%; if there

is no more than half of the data is contaminated, it can produce a good result. The breakdown point of the M-estimator and BI-estimator is different in a different situation. It depends on the strength of the signal and noise. As the final result of BI and M-estimator depends on the reweighed average method (least square estimator).

When the data site is noisy, the most significant difference between the M-estimator and BI-estimator is the leverage weighting. Suppose there is no leverage point, the breakdown points of the M-estimator and BI-estimator become the same. The leverage point is the large variation of magnetic field data. By the least square theory, the energetic magnetic field data will dominant the final result. The M-estimator is not sensitive to the leverage point. BI-estimator will reject the leverage point by using a strict iterative weighting. But the leverage point can be an energetic signal or active noise.

At the first field data test, the leverage point corresponds to the energetic single when there is a geomagnetic storm event. The leverage weighting will bring a negative influence to the reweight least square estimator. The M-estimator will perform better than the BI-estimator and RM-estimator.

When the leverage point is the active noise at the second field data test, the BI-estimator will work much better than the M-estimator and RM-estimator. The breakdown point of RM-estimator is 50%. In this test, the RM-estimator get biased by the data from 14:00:00 to 20:00:00. We think there are three-hour high SNR time-series data from 16:00:00 to 20:00:00. The M-estimator get biased by the data from 15:00:00 to 20:00:00. The breakdown point of the M-estimator is 40%; The RM-estimator was better than M-estimator. The BI-estimator gave biased results using the data from 14:00:00 to 20:00:00. The breakdown point of the BI-estimator is close to 50%. The breakdown point of the M-estimator and BI-estimator is different in a different situation.

ACKNOWLEDGEMENTS

We want to thank the INTERMAGNET (International Real-time Magnetic Observatory Network) for providing the magnetic time-series data to construct the synthetic MT data. The dataset L6-7 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. We want to thank the SAMTEX team for providing the time series data of the KAP03 survey line to do the research.

REFERENCES

- 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.

Egbert, G. D. and Booker, J. R., 1986, Robust estimation of geomagnetic transfer functions, *Geophys. J. R. Astr. Soc.*, **87**, 173–194.

Fowler, R. A., Kotick, B. J. and 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.

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

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.

Smirnov, M.Y., 2003, Magnetotelluric data processing with a robust statistical procedure having a high breakdown point, *Geophys. J. Int.*, **152**, 1–7.

Wight, D., and Bostick, F., 1980, Cascade decimation—A technique for real time estimation of power spectra. In ICASSP'80. IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. **5**, pp. 626–629). IEEE.

Weckmann, U., Magunia, A. and 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, in *Electromagnetic Methods in Applied Geophysics. Applications*, ed. Nabighian, M. N., SEG, 713–809.

Figures

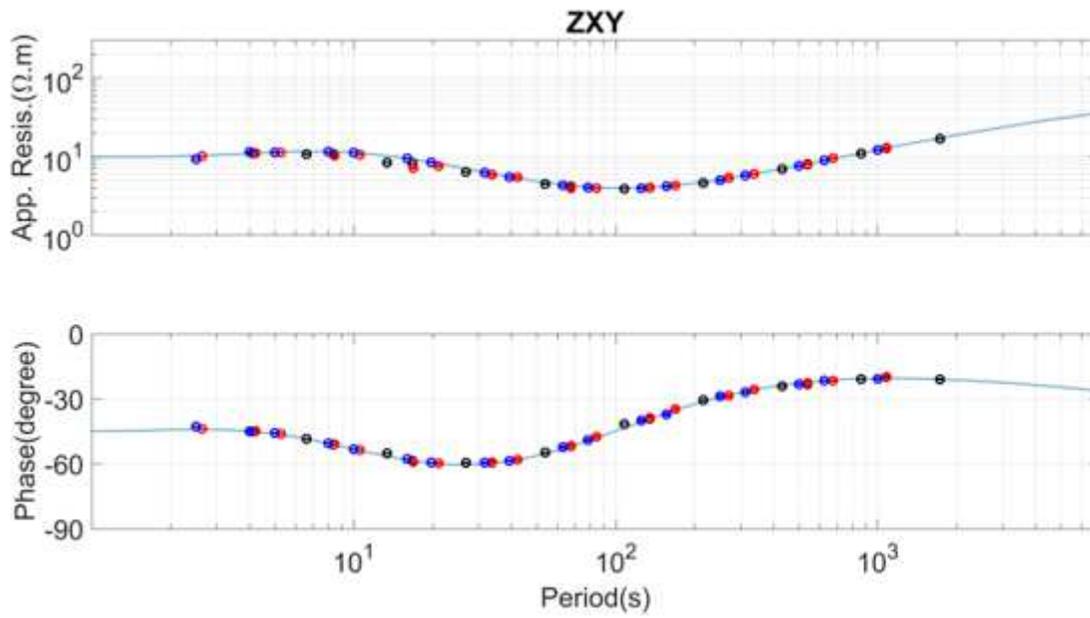


Figure 1

The comparison of the impedance curves between the true model and the result calculated from the time series. The blue line is the true curve of a simple 1-D model. Blue circles were calculated using the BI-estimator, and red circles were calculated using the M-estimator. Black circles were calculated using the RM estimator.

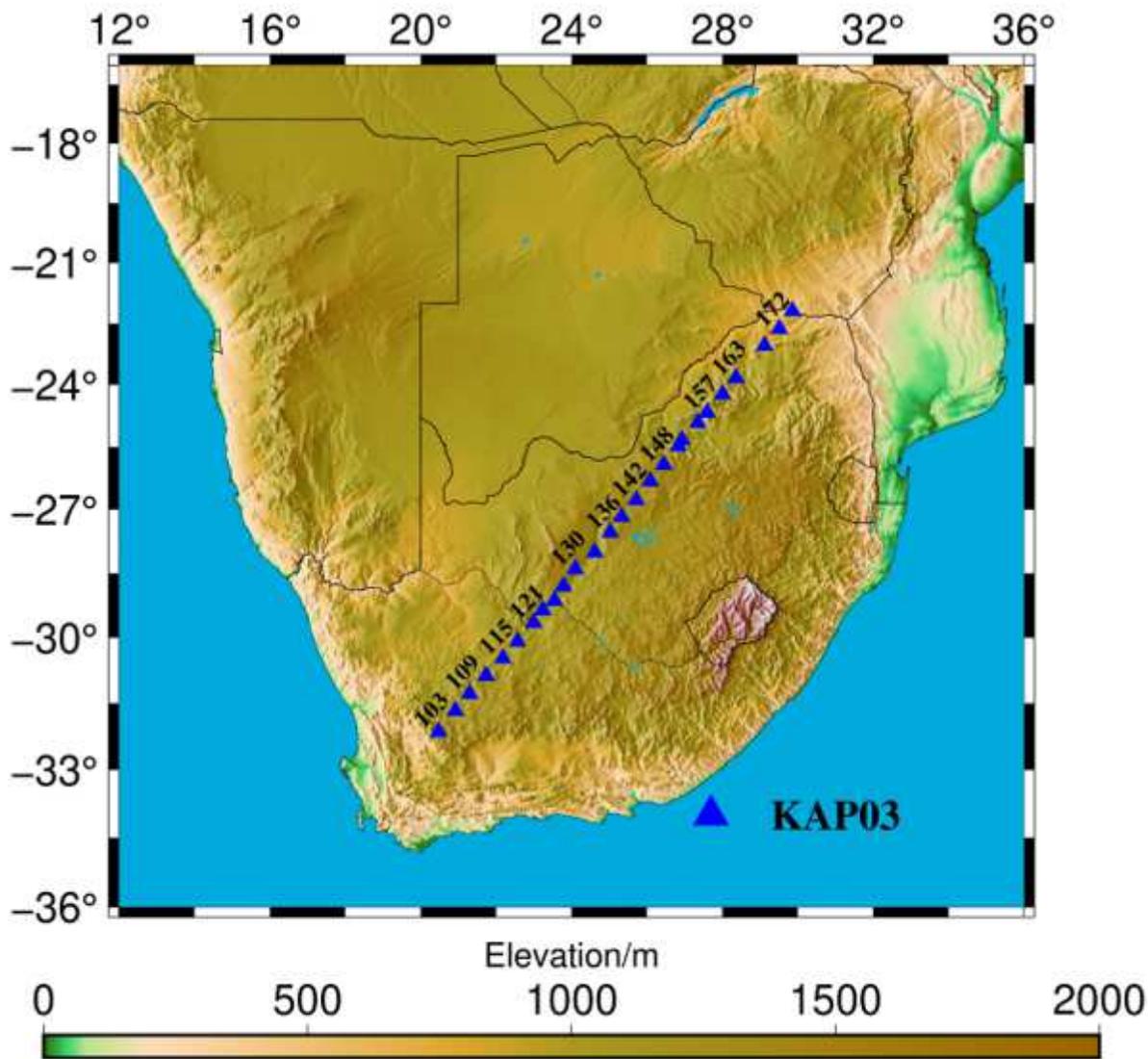


Figure 2

Location map of the survey line KAP03 in South Africa. 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.

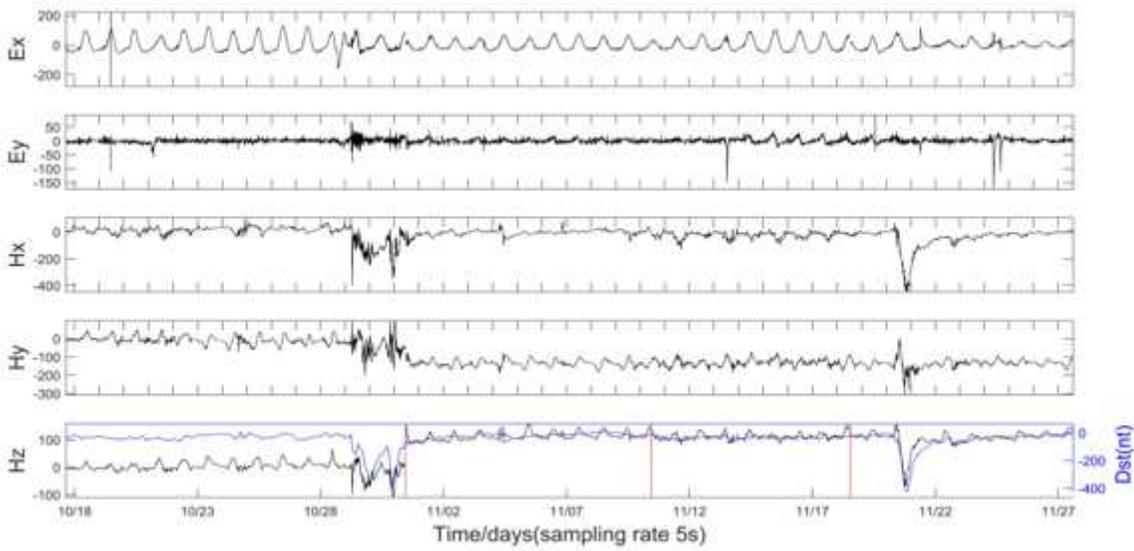


Figure 3

Time series of magnetotelluric data whose sampling rate is 5 seconds at site 130. The red vertical lines show the missing data section, and the black lines show the 5-component MT time series data. The blue line shows the Dst index of magnetic data. The unit of the electric field is mV/km, and the magnetic field in nT.

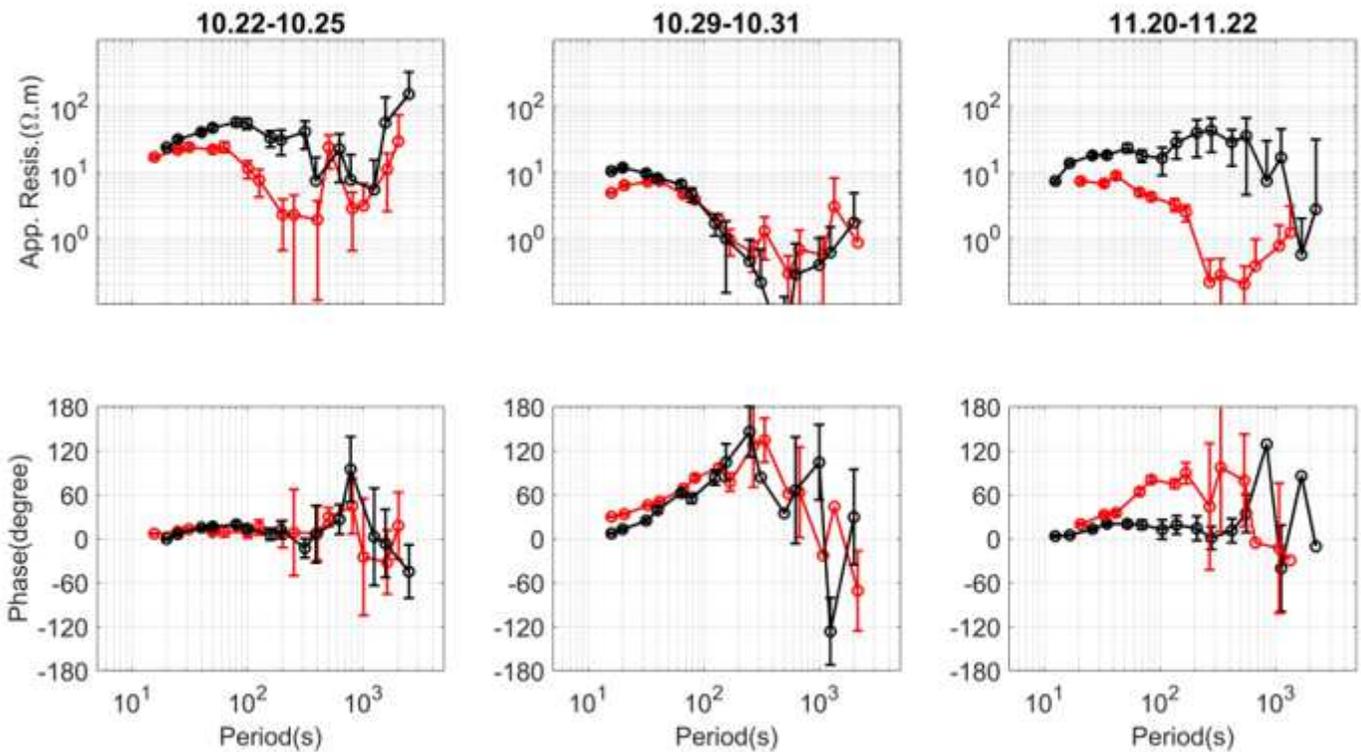


Figure 4

The XY component of magnetotelluric sounding curves calculated during both storm day and the non-storm day at site 130. The M-estimator calculated the red curves, and the BI-estimator calculated the black curves. The left figures were calculated using the non-storm data, and the central figures were calculated using the second storm data. The right figures were calculated using the third storm data. The upper figures show the apparent resistivity curves, and the lower figures show the impedance phase curves.

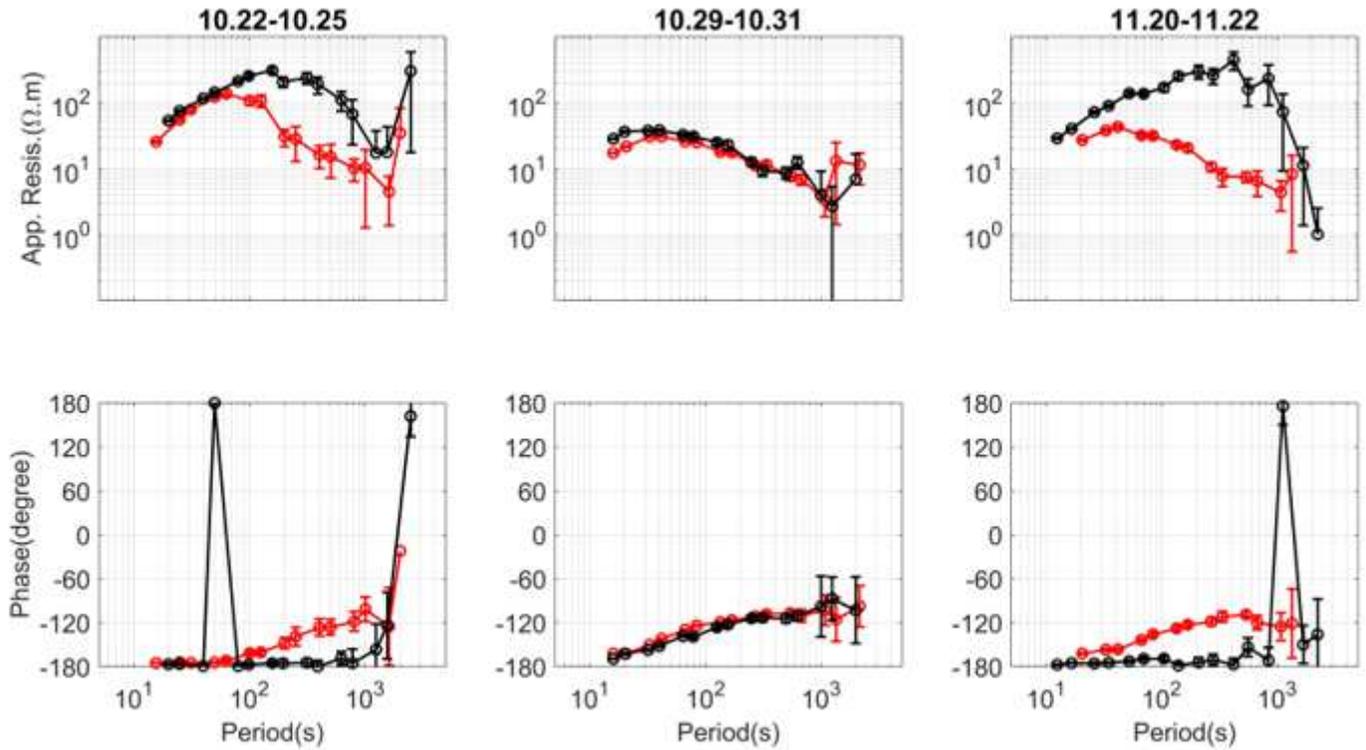


Figure 5

The YX component of magnetotelluric sounding curves calculated during both storm day and the non-storm day at site 130. The colors are the same as Fig.4.

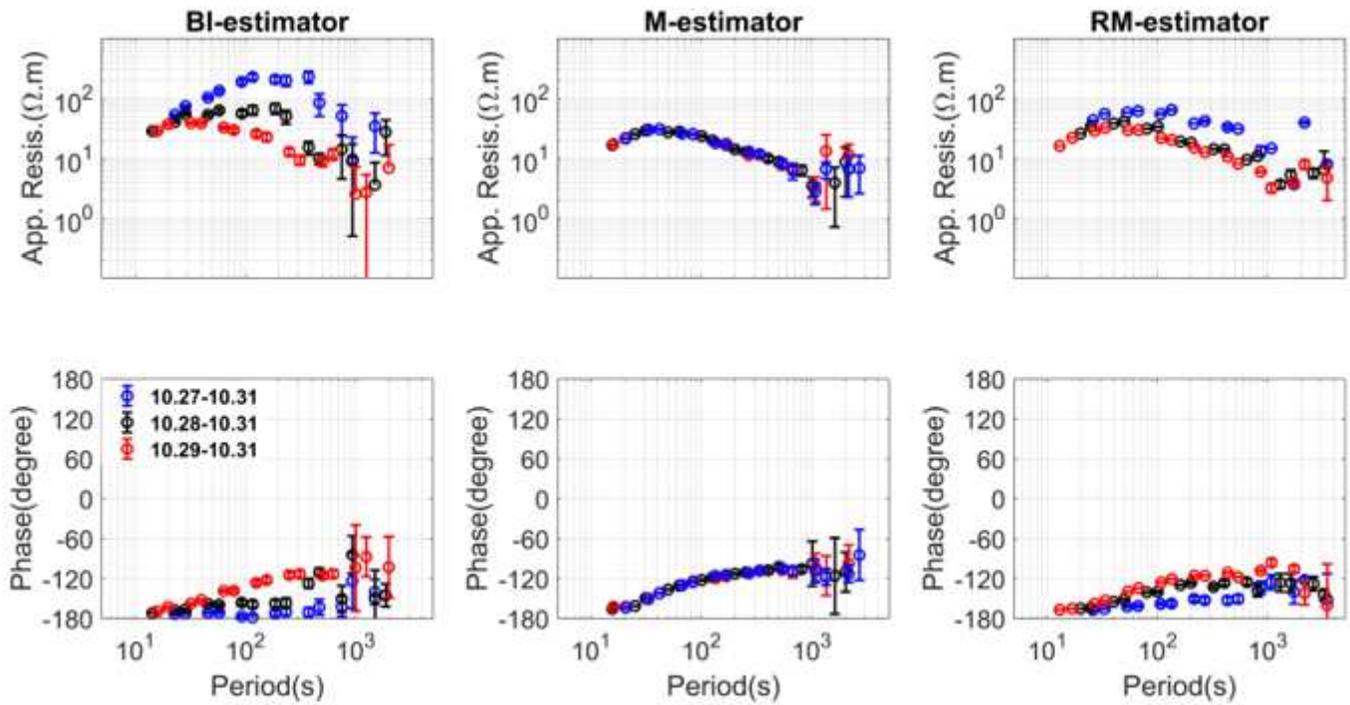


Figure 6

The YX component of magnetotelluric sounding curves calculated using the different periods around the first geomagnetic storm at site 130. The left figures were calculated using the BIRRP. The center figures were calculated using the M-estimator. The right figured were calculated using the RM-estimator. The blue curves were calculated using the dataset from Oct. 27 to Oct. 31. The black curves were calculated using the dataset from Oct. 28 to Oct. 31. The red curves were calculated using the dataset from Oct. 29 to Oct. 31.

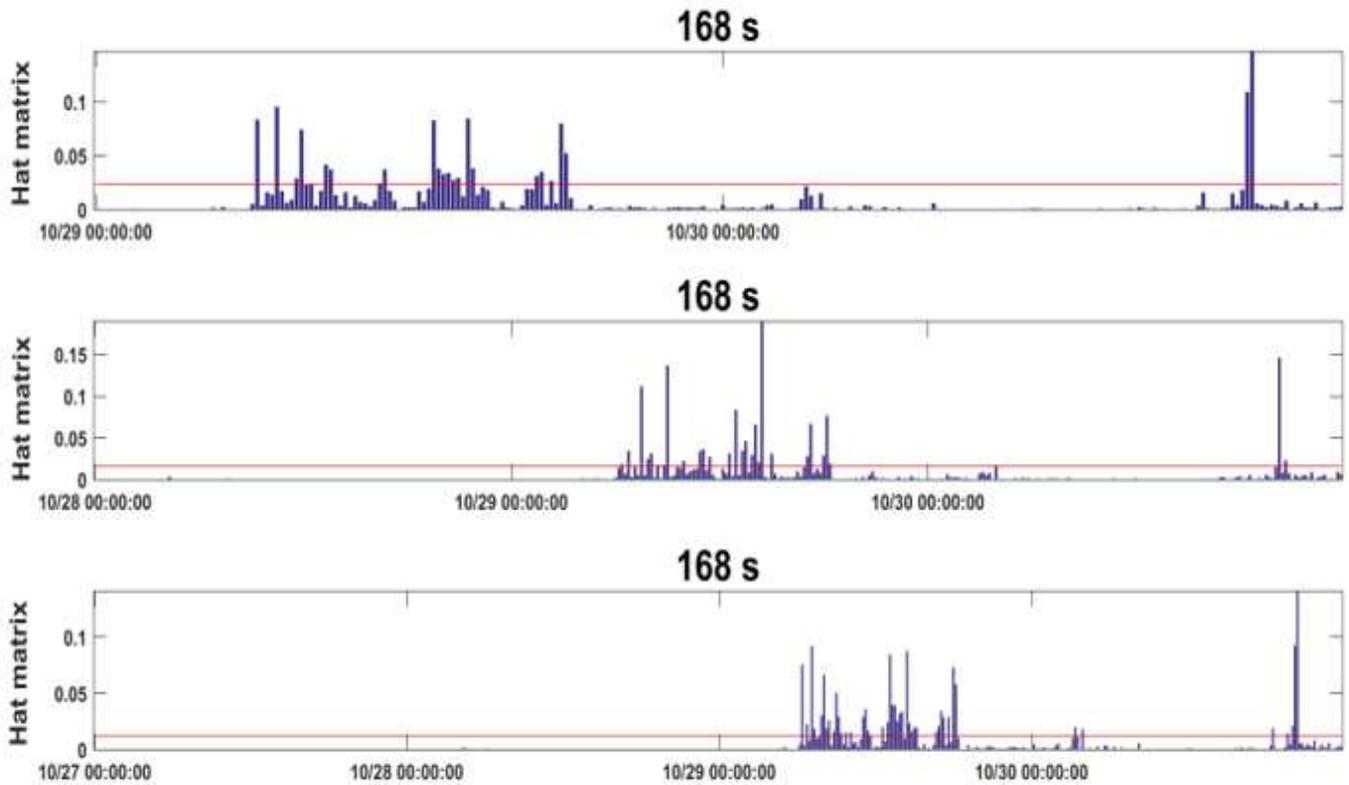


Figure 7

The hat matrix diagonal element distribution at period 168 second. The upper figure shows the data from Oct. 29 to Oct. 31. The central figure shows the data from Oct. 28 to Oct. 31. The lower figure shows the data from Oct. 27 to Oct. 31. The red dotted lines denote the three times of the $2/N$.

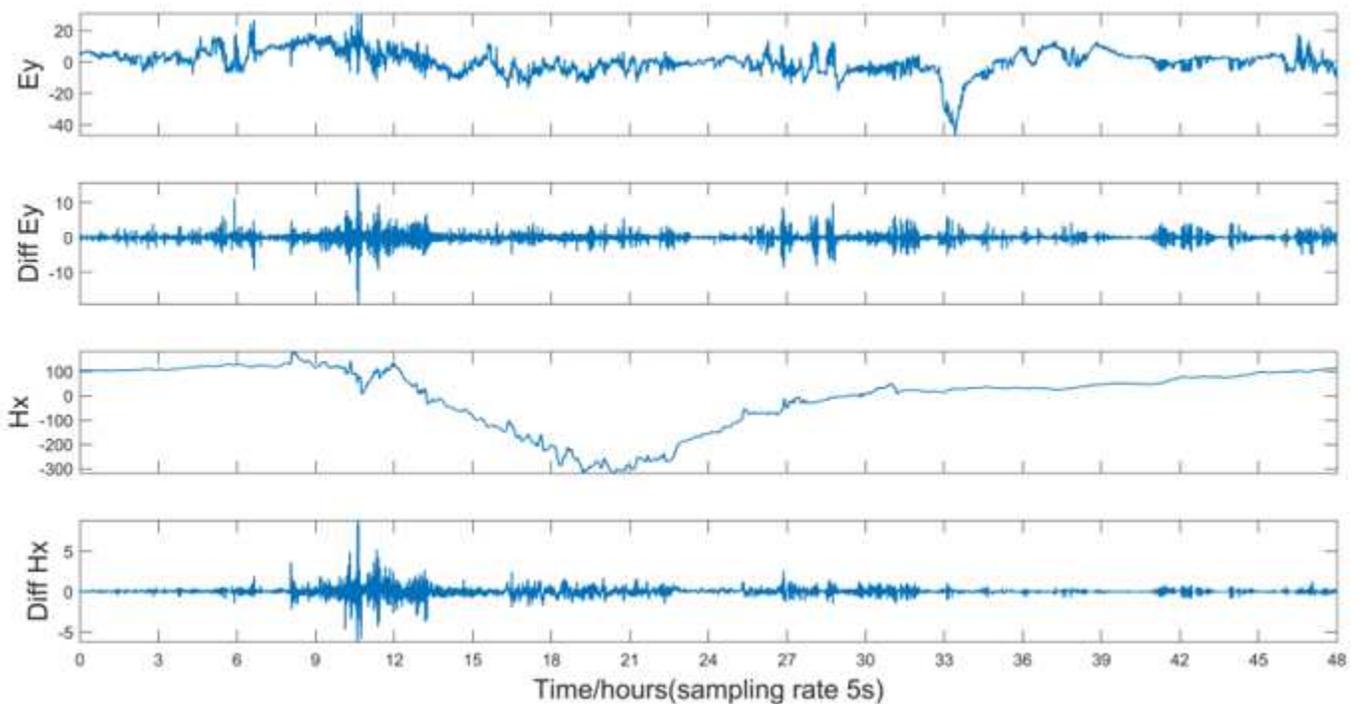


Figure 8

MT time-series data at site 130 from Nov. 20 to Nov. 22. The top figure is the e_y time-series. The second figure is the first derivative of e_y . The third figure is the h_x time-series. The bottom figure is the first derivative of h_x . The unit of the electric field is mV/km, and the unit of the magnetic field is nT.

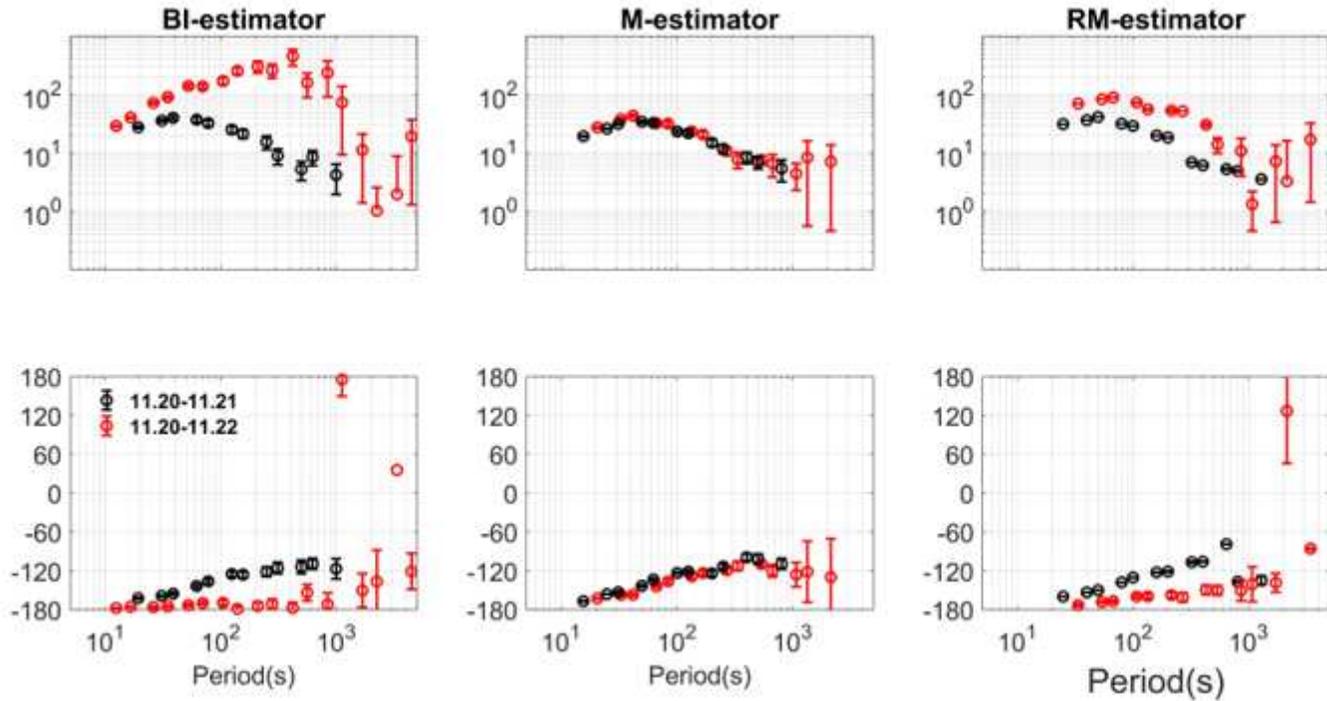


Figure 9

The YX component of magnetotelluric sounding curves calculated using the different periods around the second geomagnetic storm at site 130. The left, middle and right figures were calculated using the BI-estimator, the M-estimator and the RM-estimator. The black curves were calculated using the data from 06:00:00 on Nov. 20 to 00:00:00 on Nov. 21. The red curves were calculated using the data from 00:00:00 on Nov. 20 to 00:00:00 on Nov. 22.

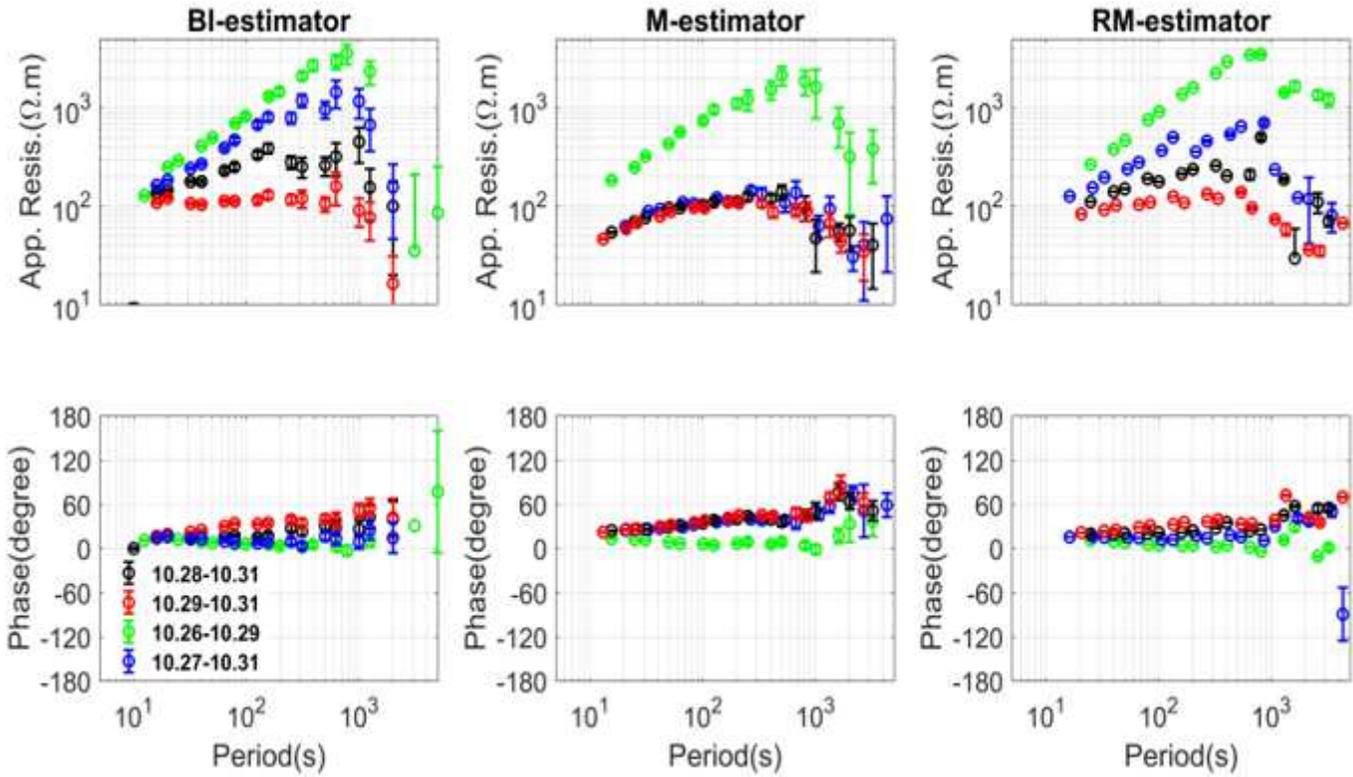


Figure 10

The XY component of MT sounding curves calculated from the different period data at site 133. The left figures were calculated using the BI-estimator. The central figures were calculated using the M-estimator. The right figures were calculated using the M-estimator. The red curves were calculated using the period from 00:00:00 on Oct. 29 to 00:00:00 on Oct. 31. The black curves were calculated using the period from 00:00:00 on Oct. 28 to 00:00:00 on Oct. 31. The blue curves were calculated using the period from 00:00:00 on Oct. 27 to 00:00:00 on Oct. 31. The green curves were calculated using the period from 00:00:00 on Oct. 26 to 00:00:00 on Oct. 31.

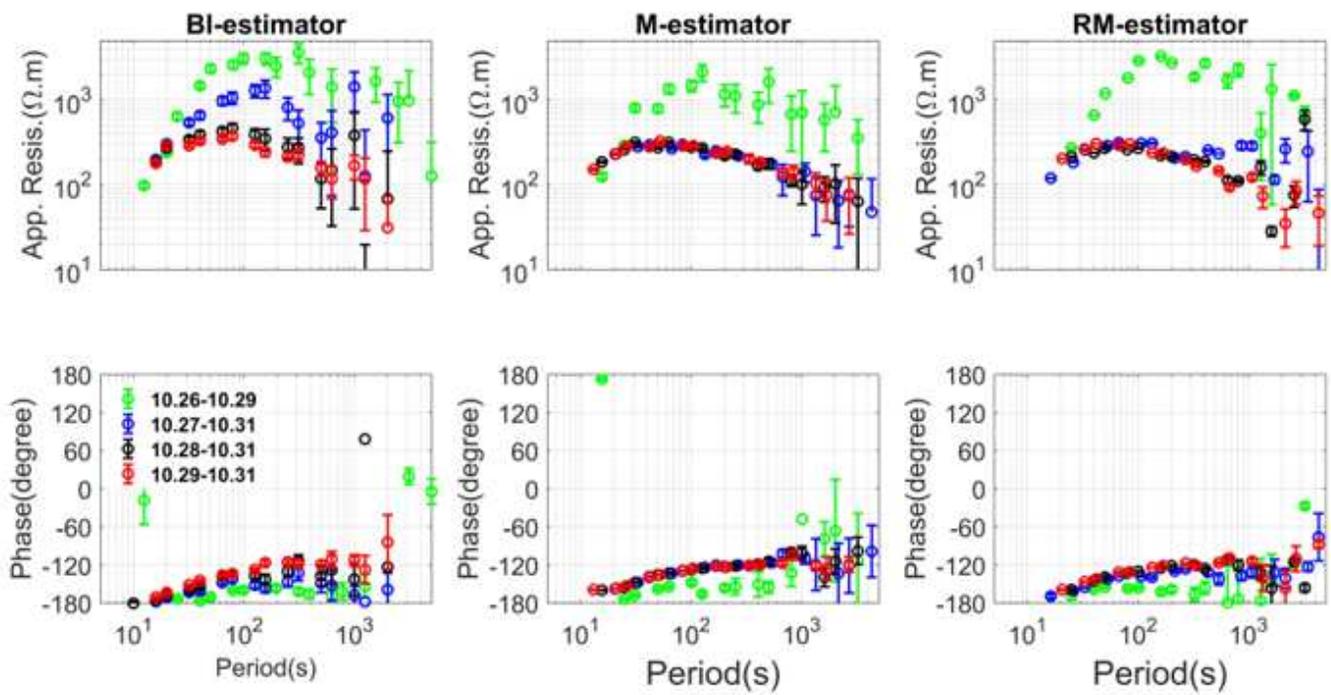


Figure 11

The YX component of MT sounding curves calculated from different period data at site 133. The colors are the same as Fig. 10.

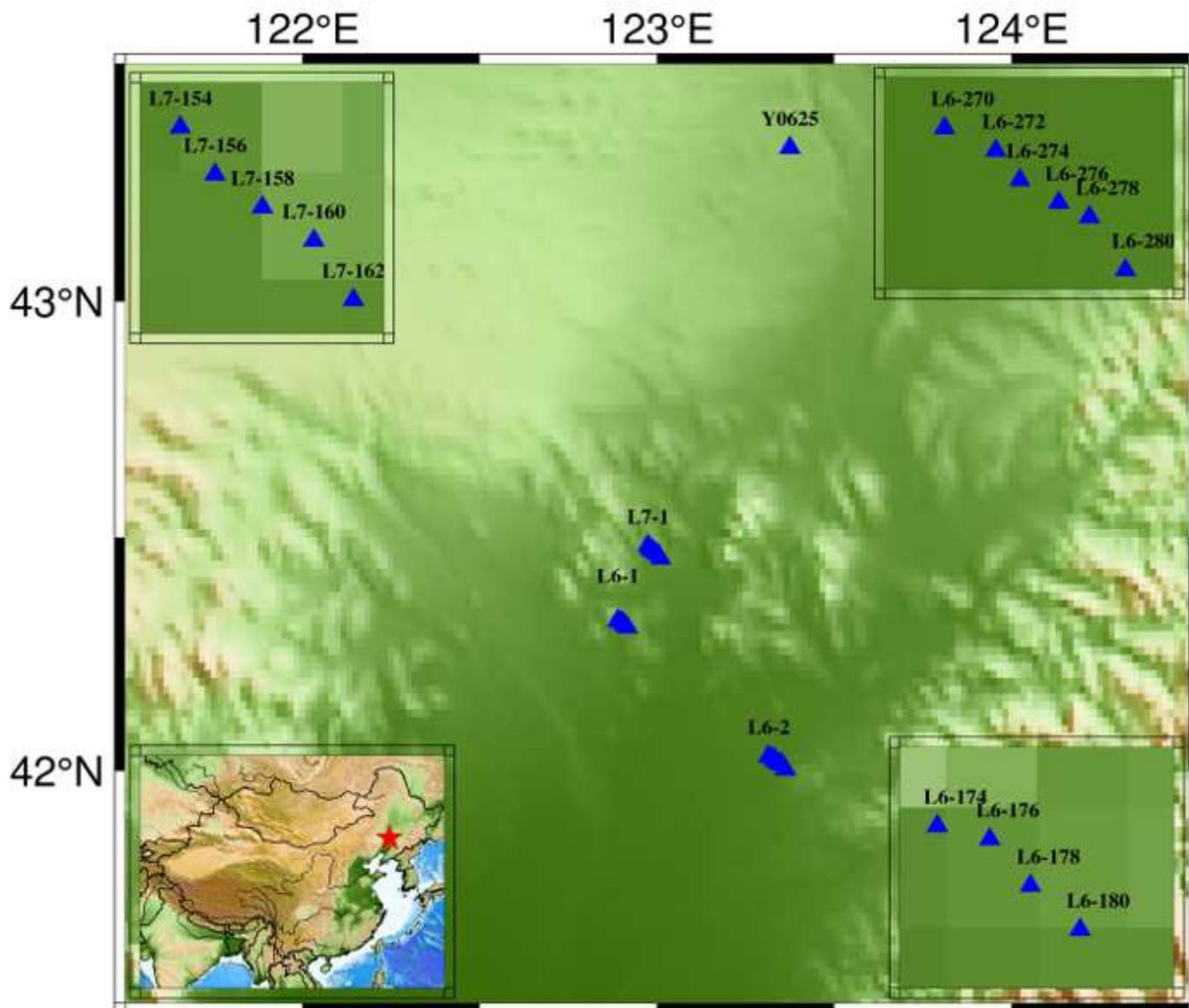


Figure 12

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

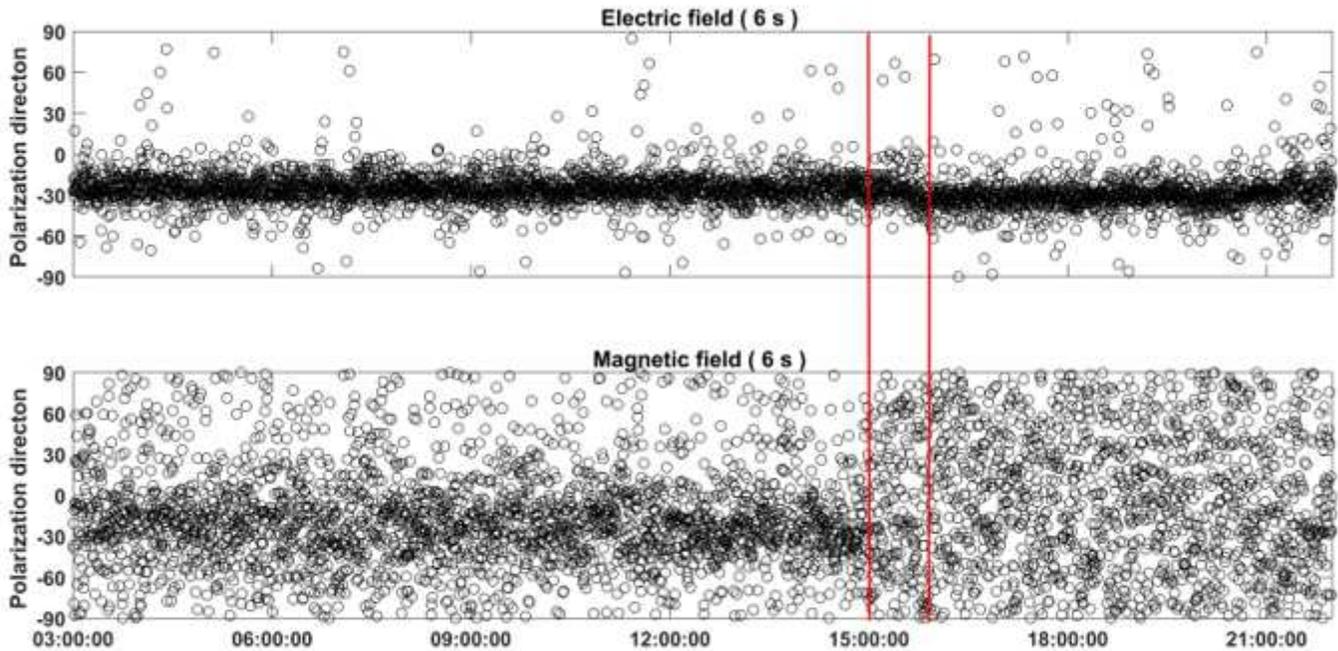


Figure 13

Polarization direction for period 6 second at site L7158 from 3:00:00 to 22:00:00 in UCT. The upper figure shows the electric field's polarization directions, and the lower figure shows the magnetic field's polarization direction. The red vertical lines denote the boundary of the property of the polarization directions changed.

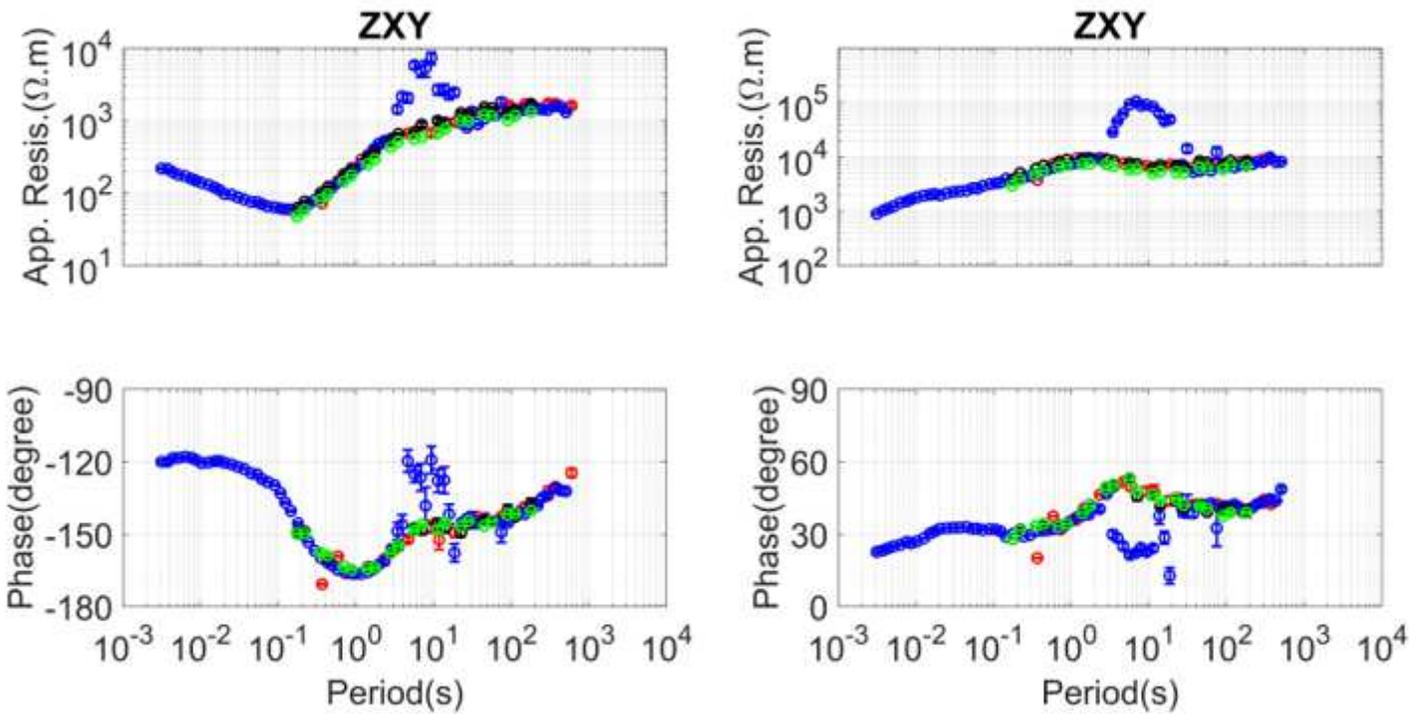


Figure 14

The magnetotelluric sounding curves calculated using the data at site L7158. SSMT-2000 calculated the blue curves using one-day data. M-estimator calculated the black curves using the midnight data from 16:00:00 to 20:00:00 in UTC. BI-estimator calculated the red curves using the midnight data from 16:00:00 to 20:00:00 in UTC. RM-estimator calculated the green curves using the midnight data from 16:00:00 to 20:00:00 in UTC.

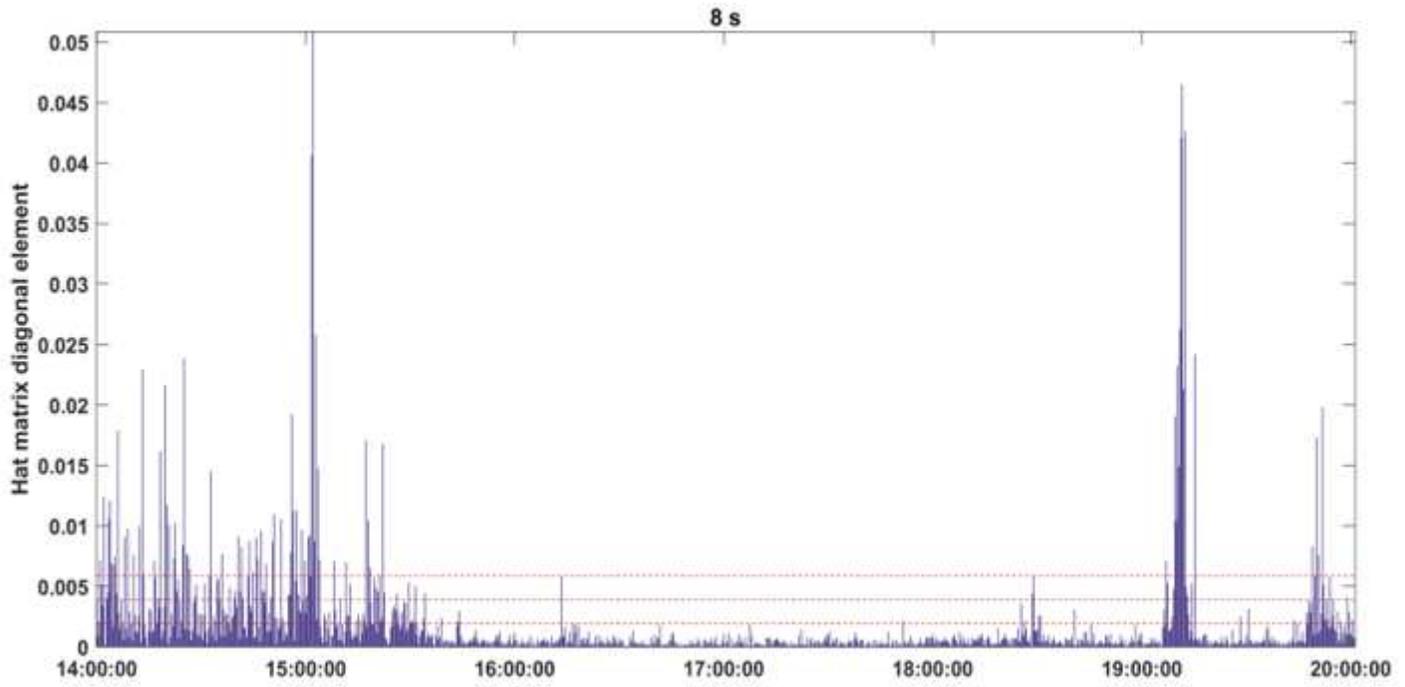


Figure 15

The hat matrix diagonal element distribution from 14:00:00 to 20:00:00 in UTC at the period of 8 seconds. The red dotted lines denote the one, two, three times of the $2/N$, respectively.

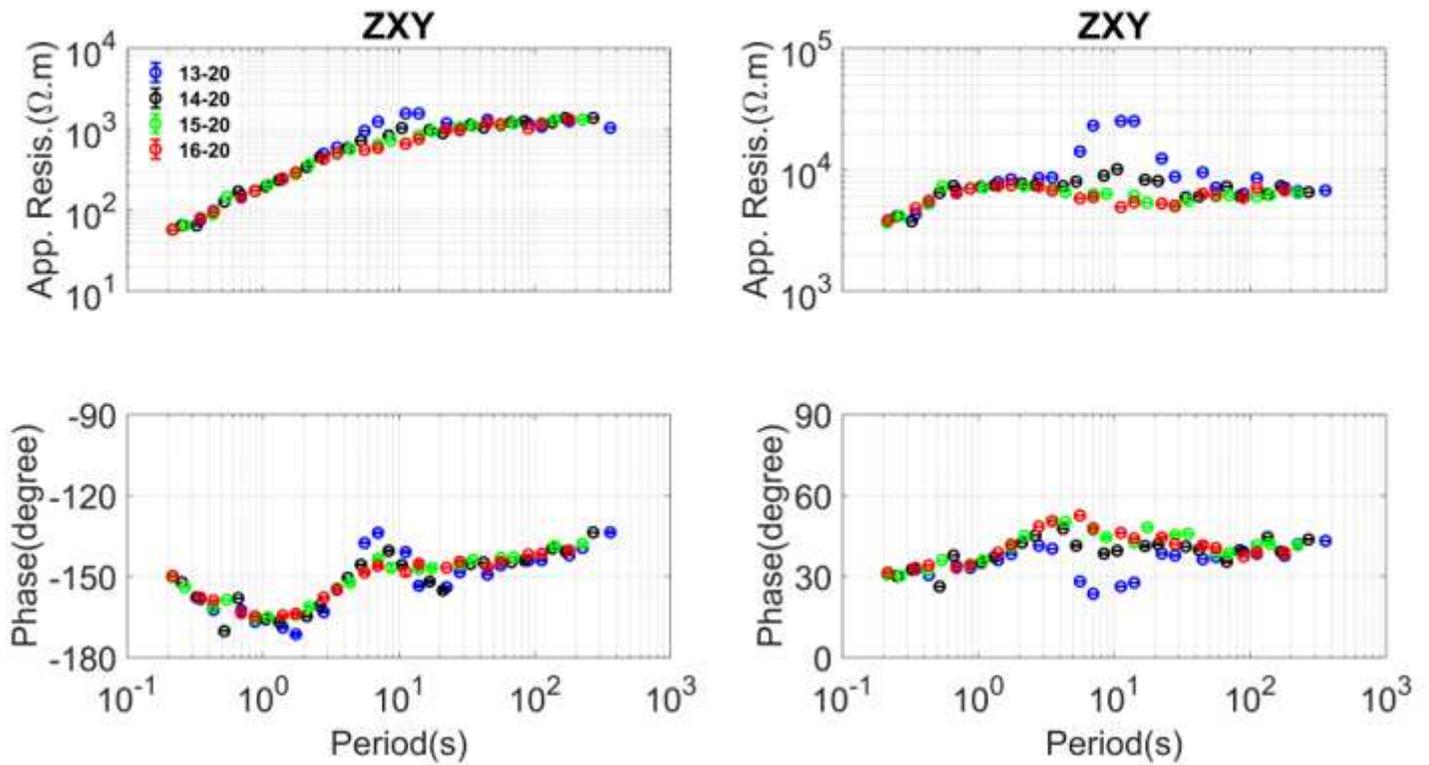


Figure 16

The magnetotelluric sounding curves calculated by RM-estimator using different period data at site L7158. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The black curves were calculated using the data from 14:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 13:00:00 to 20:00:00 in UTC.

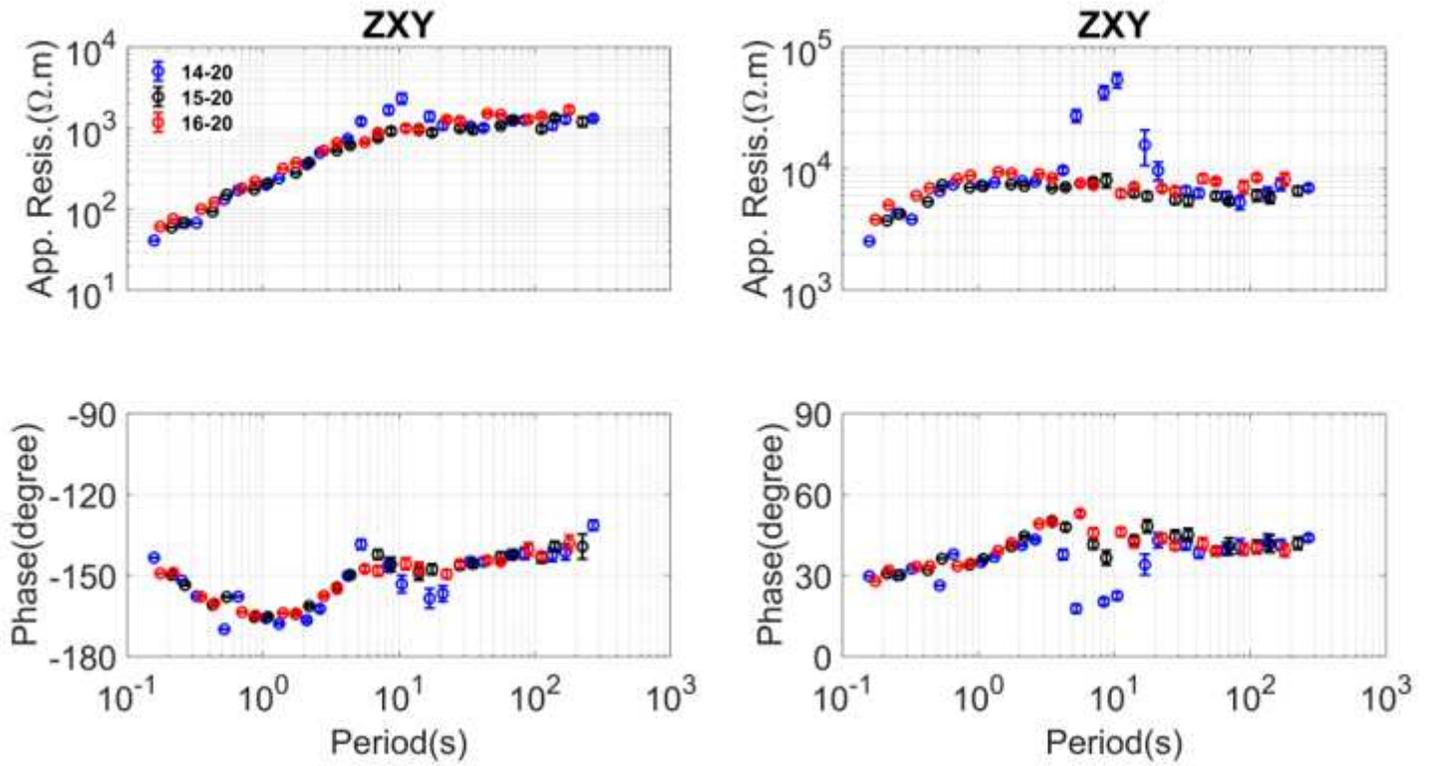


Figure 17

The magnetotelluric sounding curves calculated by M-estimator using different period the data at site L7158. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The black curves were calculated using the midnight data from 15:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 14:00:00 to 20:00:00 in UTC.

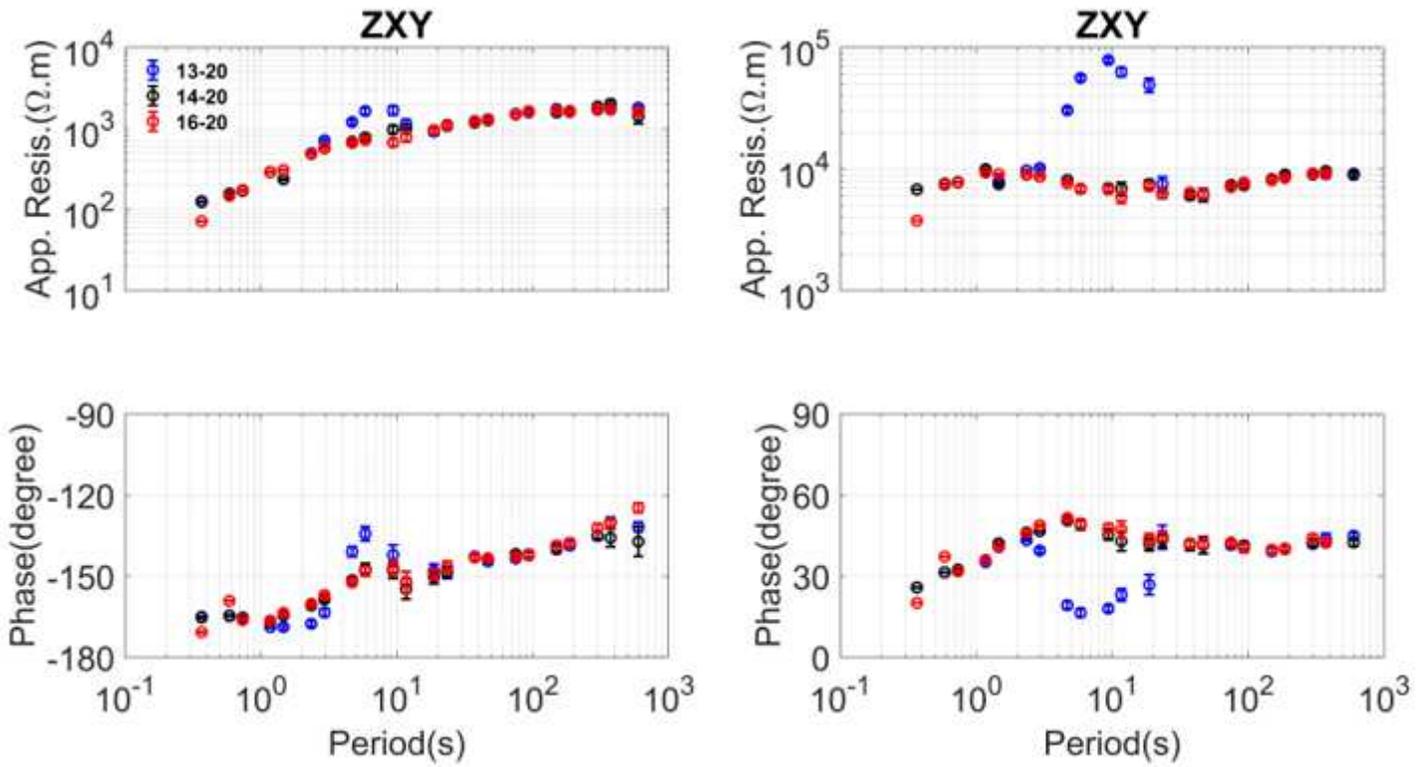


Figure 18

The magnetotelluric sounding curves calculated by the BI- estimator using different period data at site L7158. The red curves were calculated using the midnight data from 16:00:00 to 20:00:00 in UTC. The black curves were calculated using the data from 14:00:00 to 20:00:00 in UTC. The blue curves were calculated using the data from 13:00:00 to 20:00:00 in UTC.