New robust remote reference estimator using robust multivariate linear regression

Yoshiya Usui [1], Makoto Uyeshima,[1] Shin'ya Sakanaka,[2] Tasuku Hashimoto,[1] Masahiro Ichiki,[3] Toshiki Kaida,[3] Yusuke Yamaya,[4] Yasuo Ogawa,[3,5] Masataka Masuda[1] and Takahiro Akiyama[1]

[1]*Earthquake Research Institute, the University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-0032, Japan. E-mail: yusui@eri.u-tokyo.ac.jp*
[2]*Faculty of Engineering and Resource Science, Akita University, 1-1 Tegatagakuen-machi, Akita, 010-8502, Japan*
[3]*Research Center for Prediction of Earthquakes and Volcanic Eruptions, Graduate School of Science, Tohoku University,6–6 Aoba, Aramaki, Aoba–ku, Sendai,980-8578, Japan*
[4]*Renewable Energy Research Center, Fukushima Renewable Energy Institute, National Institute of Advanced Industrial Science and Technology, 2-2-9 Machiikedai, Koriyama, Fukushima, 963-0298, Japan*
[5]*Institute of Innovative Research, Tokyo Institute of Technology, Tokyo, Japan*

SUMMARY

The solution of the remote reference method, a frequently used technique in magnetotelluric (MT) data processing, can be viewed as a product of the two-input–multiple-output relationship between the local electromagnetic field and the reference field at a remote station. By applying a robust estimator to the two-input–multiple-output system, one can suppress the influence of outliers in the local magnetic field as well as those in the local electric field based on regression residuals. Therefore, this study develops a new robust remote reference estimator with the aid of robust multivariate linear regression. By applying the robust multivariate regression S-estimator to the multiple-output system, the present work derives a set of equations for robust estimates of the transfer function, noise variances, and the scale of the Mahalanobis distance simultaneously. The noise variances are necessary for the multivariate analysis to normalize the residuals of dependent variables. The Mahalanobis distance, a distance measure for multivariate data, is a commonly used indicator of outliers in multivariate statistics. By updating those robust estimates iteratively, the new robust remote reference estimator seeks the transfer function that minimizes the robust scale estimate of the Mahalanobis distance. The developed estimator can avoid bias in the MT transfer function even if there are significant noises in the reference magnetic field and handle outlying data more robustly than previously proposed robust remote reference estimators. The authors applied the developed method to a synthetic data set and real-world data. The test results demonstrate that the developed method downweights outliers in the local electric and magnetic fields and gives an unbiased transfer function.

Key words: Magnetotellurics; Statistical methods; Time-series analysis.

1 INTRODUCTION

The magnetotelluric (MT) method is a passive electromagnetic (EM) exploration method that images subsurface electrical resistivity structures from orthogonal components of natural EM fields at the earth's surface. In the absence of noise, a set of linear relations exists among the orthogonal components of natural EM fields. In the frequency domain, the linear relationship can be written as

$$\boldsymbol{e} = Z\boldsymbol{b}, \quad \boldsymbol{e} = \begin{pmatrix} E_x \\ E_y \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} B_x \\ B_y \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix}, \quad (1)$$

where $\boldsymbol{e} \in \mathbb{C}^2$ is the vector of the electric field in mV km^{-1}; $\boldsymbol{b} \in \mathbb{C}^2$ is the vector of the magnetic field in nT and $\boldsymbol{Z} \in \mathbb{C}^{2 \times 2}$ is the transfer function (TF), called the impedance tensor, in mV km^{-1} nT^{-1}. In this study, vectors and matrices are written in bold lowercase and uppercase, respectively. Eq. (1) can be broken down into a pair of two-input–one-output linear systems, bivariate linear regressions in other words:

$$E_x = Z_{xx}B_x + Z_{xy}B_y, \quad (2)$$

$$E_y = Z_{yx}B_x + Z_{yy}B_y. \quad (3)$$

However, in the real world, eqs (1), (2) and (3) do not hold, owing to the presence of noise. Cultural noises from direct current electric railways, power lines, electrical fences, motor vehicles running near MT stations and other sources cause large errors in estimating the MT TF (Szarka 1988; Junge 1996; Ferguson 2012). The least-square estimate of the TF can be obtained by minimizing the autospectral density of the noise in the dependent variable of each bivariate linear regression.

However, if data are heavily distorted by noises, using the least-squares method can lead to a biased estimate of the MT TF (Chave 2012). For example, even one single outlier can significantly affect the least-square estimate because the ordinary least-square method is extremely sensitive to outliers. This problem has been overcome by mainly using two principal procedures (Chave 2012): (1) the remote reference method and (2) robust weighting schemes. In the remote reference method (Goubau *et al.* 1978; Gamble *et al.* 1979), the horizontal magnetic field is recorded simultaneously at a reference station, where noises are uncorrelated with those at a local MT station. In the TF estimation, the cross-power spectral density of the local and reference magnetic field is used instead of the autopower spectral density of the local magnetic field. Because there is no correlated noise between the local and reference magnetic fields, the bias due to the autopower spectral density of noise, which is inherent in the ordinary least-square estimate (Sims *et al.* 1971; Simpson & Bahr 2005), can be avoided. Robust weighting schemes are often used in combination with the remote reference method to suppress the influence of heavily distorted data that can act as outliers (e.g. Chave & Thomson 1989; Larsen 1989; Larsen *et al.* 1996). Their effectiveness has been recognized, and various data-adaptive robust weighting schemes have been proposed (Chave & Thomson 1989; Larsen 1989; Larsen *et al.* 1996; Egbert 1997; Ritter *et al.* 1998; Chave & Thomson 2004; Smirnov & Egbert 2012).

In robust weighting schemes, the M-estimator, which detects and downweights outliers based on regression residuals, has been most frequently used (e.g. Egbert & Booker 1986; Chave & Thomson 1989). However, the M-estimator cannot effectively handle the outliers in predictor variables (horizontal magnetic field data in the MT TF estimation). Such outliers, called leverage points, can significantly distort TF estimates (Chave & Thomson 2004; Chave 2012; Maronna *et al.* 2019). To overcome the limitation of the M-estimator, Chave & Thomson (2004) developed the two-stage bounded influence (BI) method. Their method estimates the TF at a local MT station in two stages. In the first stage, the interstation horizontal magnetic TF (HMTF) between local and reference stations is estimated. In the second stage, the local MT TF is estimated using the estimate of the local magnetic field as the predictor variable. The two-stage BI method can handle leverage points because outliers in local magnetic field data can be detected in the first stage based on regression residual in addition to the leverage weighting based on the hat matrix. Chave & Thomson (2004) and Chave (2012) have shown the superiority of the two-stage BI method to the M-estimator.

If the weight matrix for the second stage is a constant multiplication of that for the first stage, the two-stage method yields a result equivalent to that of the original remote reference method (Goubau *et al.* 1978; Gamble *et al.* 1979) because the matrix containing the autospectral density of the estimated local magnetic field is equal to the matrix consisting of the cross-spectral density of the estimated magnetic field and local magnetic field (Chave 2012) in that case:

$$\hat{\boldsymbol{B}}^H \boldsymbol{W} \hat{\boldsymbol{B}} = \hat{\boldsymbol{B}}^H \boldsymbol{W} \boldsymbol{B}. \tag{4}$$

where $\hat{\boldsymbol{B}} \in \mathbb{C}^{n_d \times 2}$ and $\boldsymbol{B} \in \mathbb{C}^{n_d \times 2}$ (n_d is the number of data segments) consist of the Fourier transforms of the estimated and observed magnetic field, respectively, and $\boldsymbol{W} \in \mathbb{R}^{n_d \times n_d}$ is a diagonal weight matrix in the second stage. However, if such a condition is not satisfied, which generally occurs, the solution of the two-stage approach is different from that of the original remote reference method because

$$\hat{\boldsymbol{B}}^H \boldsymbol{W} \hat{\boldsymbol{B}} = \boldsymbol{B}^H \boldsymbol{W}_1 \boldsymbol{B}_r \left(\boldsymbol{B}_r{}^H \boldsymbol{W}_1 \boldsymbol{B}_r\right)^{-1} \boldsymbol{B}_r{}^H \boldsymbol{W} \boldsymbol{B}_r \left(\boldsymbol{B}_r{}^H \boldsymbol{W}_1 \boldsymbol{B}_r\right)^{-1}$$
$$\boldsymbol{B}_r{}^H \boldsymbol{W}_1 \boldsymbol{B} \neq \hat{\boldsymbol{B}}^H \boldsymbol{W} \boldsymbol{B}. \tag{5}$$

where $\boldsymbol{B}_r \in \mathbb{C}^{n_d \times 2}$ consists of the Fourier transform of the reference magnetic field, and $\boldsymbol{W}_1 \in \mathbb{R}^{n_d \times n_d}$ is a diagonal weight matrix in the first stage. Even in such cases, the two-stage BI method still gives an unbiased estimate if the magnetic field at a reference station does not contain noise. However, if the reference magnetic field contains noises, the two-stage approach can lead to a biased TF estimate. This is because the HMTF can be underestimated in the first stage due to the autopower spectral density of noise, resulting in the underestimation of the local magnetic field (the predictor variable in the second stage).

When one returns to the original remote reference method (Goubau *et al.* 1978; Gamble *et al.* 1979), it can be viewed as a product of the two-input–multiple-output relationship between the local EM field and the reference magnetic field (Schmucker 1984; Neska 2006; Campanyà *et al.* 2014). Under the uniform source assumption, the local EM field at an MT station can be linearly related to the magnetic field at a reference station (e.g. Schmucker 1984):

$$\begin{pmatrix} \boldsymbol{e} \\ \boldsymbol{b} \end{pmatrix} = \boldsymbol{T} \boldsymbol{b}_r, \ \boldsymbol{T} = \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix}, \ \boldsymbol{b}_r = \begin{pmatrix} B_{rx} \\ B_{ry} \end{pmatrix}, \tag{6}$$

where $\boldsymbol{b}_r \in \mathbb{C}^2$ indicates a vector of the horizontal magnetic field at a reference station, and $\boldsymbol{T} \in \mathbb{C}^{4 \times 2}$ represents the inter-station TF (ISTF) between the local EM field and the reference magnetic field. The ISTF $\boldsymbol{T}$ consists of $\boldsymbol{U} \in \mathbb{C}^{2 \times 2}$ and $\boldsymbol{V} \in \mathbb{C}^{2 \times 2}$:

$$\boldsymbol{U} = \begin{pmatrix} U_{xx} & U_{xy} \\ U_{yx} & U_{yy} \end{pmatrix}, \boldsymbol{V} = \begin{pmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{pmatrix}. \tag{7}$$

where $\boldsymbol{U}$ denotes the ISTF between the local electric field and the reference magnetic field, called the pseudo-remote impedance tensor (Muñoz & Ritter 2013) or the quasi-MT TF (Campanyà *et al.* 2014), and $\boldsymbol{V}$ indicates the HMTF. One can obtain exactly the same equation as by the original remote reference method (Goubau *et al.* 1978; Gamble *et al.* 1979) for the impedance tensor $\boldsymbol{Z}$ by combining $\boldsymbol{U}$ and $\boldsymbol{V}$:

$$\boldsymbol{Z} = \boldsymbol{U} \boldsymbol{V}^{-1}, \tag{8}$$

as stated by Schmucker (1984), Neska (2006), Campanyà *et al.* (2014). One can robustify the resultant MT TF by using data-adaptive weights for solving eq. (6). To illustrate the characteristics of the aforementioned methods more clearly, the authors compare them in the supplemental file, assuming a one-input–one-output system.

One of the most straightforward ways to obtain a robust MT TF estimate from eqs (6) and (8) is the use of robust multivariate linear regression. By applying robust multivariate linear regression, one can solve eq. (6), the two-input–multiple-output system, simultaneously with the same weight for all dependent variables. It is important to use the same weight for all dependent variables because, otherwise, differences in the downward bias among ISTFs (eq. 7) can lead to a biased estimate of the impedance tensor if the magnetic field at a reference station ($\boldsymbol{b}_r$) contains noises.
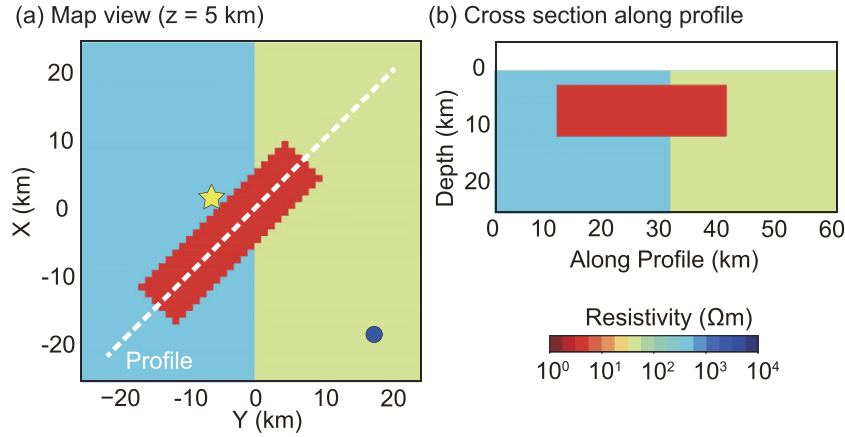
Figure 1. Oblique Conductor model (Tietze *et al.* 2015) used for making a synthetic time-series data set. (a) Map view at a depth of 5 km. The yellow star denotes the observation point of the synthetic EM data. The dark-blue circle at the bottom right denotes the observation point of the magnetic field for the remote reference method. (b) The cross-section along the profile which is shown as the white dashed line in (a).
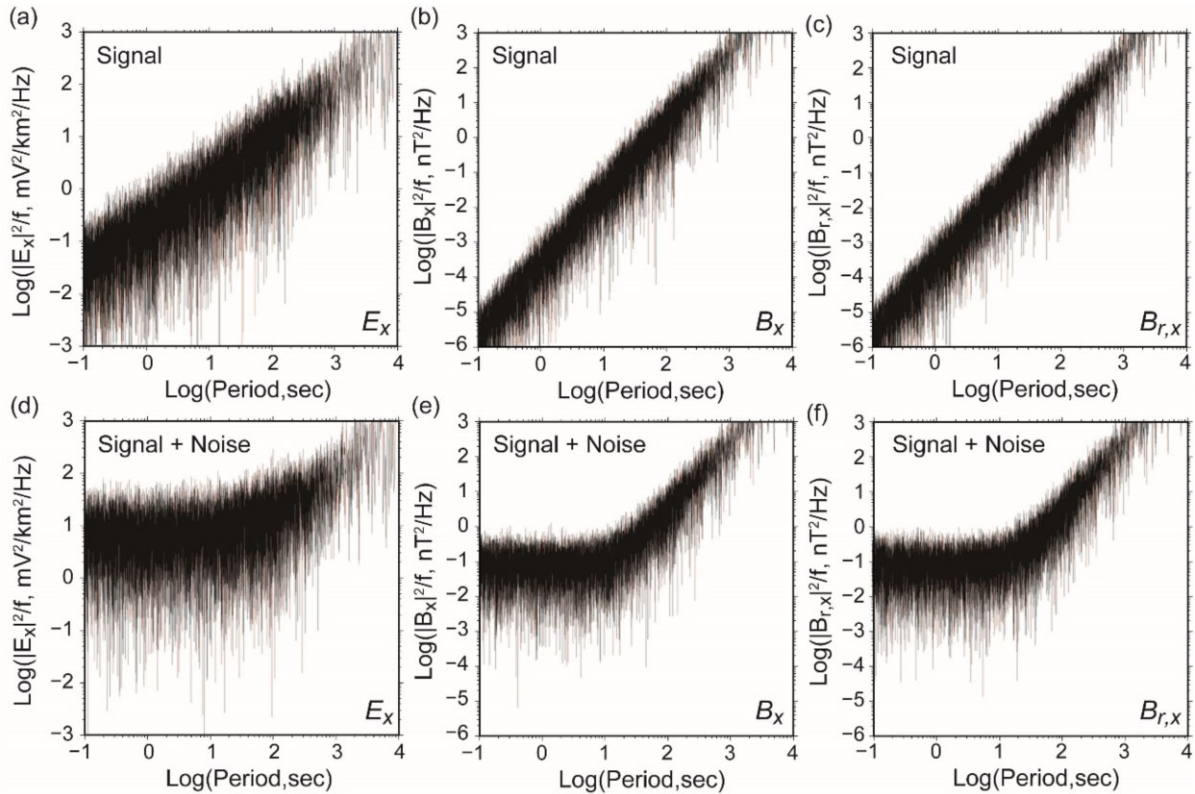

Figure 2. Power spectra of *x*-components of the synthetic EM field at the MT station (left and middle panels) and the reference station (right-hand panel). (a–c) Power spectra of the EM field containing only signals. (d–f) Power spectra of the EM field containing signals and noises.

Therefore, this study developed an estimator to solve eq. (6) simultaneously using robust multivariate linear regression. In particular, the authors used the multivariate regression S-estimator (Van Aelst & Willems 2005). The S-estimator, which was originally proposed by Rousseeuw & Yohai (1984) to solve univariate regression, seeks the estimate that minimizes a robust scale estimate among a number of M-estimates (Salibián-Barrera & Yohai 2006; Maronna *et al.* 2019). One needs to use a redescending influence curve, in which data with large residuals have no influence on the estimate, to handle heavily outlying data. However, there can be numerous local minima if the influence curve is redescending (Maronna *et al.* 2019). Hence, the M-estimate using a redescending influence curve inherently depends on the initial estimate. To avoid the limitation of the M-estimator, the S-estimator performs iteratively reweighted least squares starting from a number of different initial estimates. The S-estimator can handle leverage points because they may be detected as outlying data with large residuals depending on initial estimates. Appendix A demonstrates the differences between the M- and S-estimators using a simple linear regression problem.
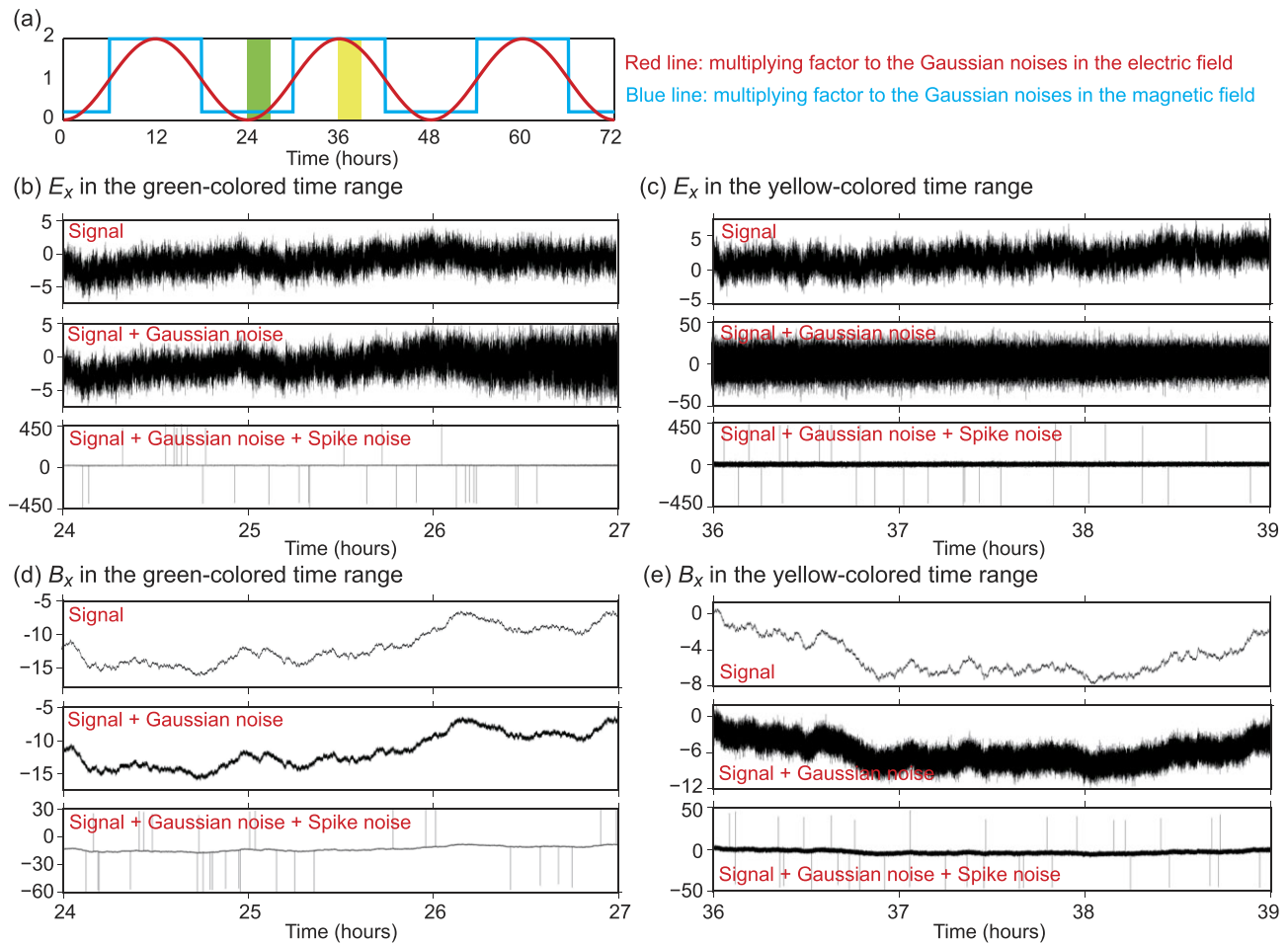
Figure 3. (a) Multiplying factors to the raw Gaussian noises in the electric field data (red line) and the magnetic field data (blue line). Due to the multiplication of those factors, the amplitudes of the Gaussian noises in the synthetic time-series vary with time. The green portion corresponds to the time ranges of (b) and (d), while the yellow portion corresponds to the time ranges of (c) and (e). (b, c) Synthetic time-series of the *x*-component of the local electric field (mV km^{-1}). (d, e) Synthetic time-series of the *x*-component of the local magnetic field (nT). In (b), (c), (d) and (e), time-series for signal, signal + Gaussian noise and signal + Gaussian noise + spike noise are depicted from top to bottom.

It should be noted that this work is not the first one that utilizes the multiple-output system (eq. 6) for estimating MT TF. Egbert (1997) and Smirnov & Egbert (2012) solved the multiple-output system in their robust estimators, although they derived clean reference channels from all available channels. However, no previous studies have solved the multiple-output system (eq. 6) as it is by robust multivariate linear regression. Our method gives the MT TF estimate with exactly the same form as the traditional robust remote reference method (e.g. Chave & Thomson 1989), which gives an unbiased MT TF, although data-adaptive weights are determined by the multivariate regression S-estimator in our method. Because the authors derived the method by straightforwardly applying robust linear regression, it has the potential to take advantage of some findings of statistics. For example, the fast and robust bootstrap method (Salibián-Barrera *et al.* 2008), an efficient bootstrap method for robust estimators, and Bayesian linear regression (e.g. Marin & Robert 2007) are potentially applicable to our method.

In this paper, first, the authors describe the details of the new robust remote reference method. Next, the authors test the developed method using a synthetic EM time-series data set. Finally, the authors apply the developed method to actual MT data measured in Northeastern Japan.

2 METHOD

This section presents the new robust remote reference method using the multivariate regression S-estimator (Van Aelst & Willems 2005). The overlapped section averaging method (Percival & Walden 1993; Bendat & Piersol 2010; Chave 2012) is used to estimate the TF, as in Chave & Thomson (1989) and Chave & Thomson (2004). After subtracting the mean from the original time-series data, each time-series is divided into 50 per cent overlapping segments. A robust pre-whitening method, originally proposed by Martin & Thomson (1982) and Maronna *et al.* (2019), is then applied. Subsequently, in the frequency domain, adjustments for the pre-whitening and the tapering are performed in addition to instrument calibrations. From the lowest target frequency, the frequency of interest is progressively increased by reducing the segment length.

The developed method is based on the two-input–multiple-output relationship between the local EM field and the reference magnetic field (eq. 6):

$$v = \begin{pmatrix} e \\ b \end{pmatrix} = \hat{T} b_r + \varepsilon, \tag{9}$$

(a) Case 1

Original remote reference method

(b) Case 2

Two-stage BI method

(c) Case 3

ISTF of each bivariate equation in eq. (6) was separately estimated by M-estimator
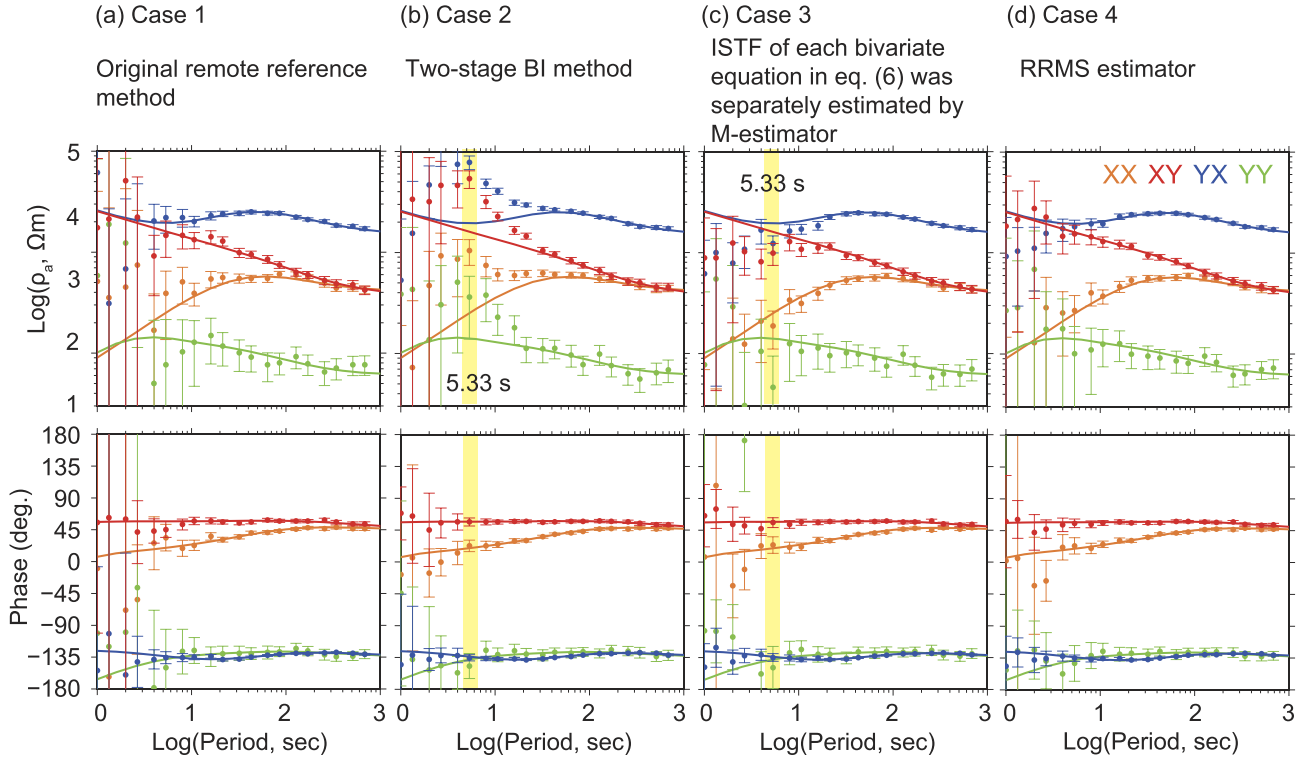
(d) Case 4

RRMS estimator

Figure 4. (a–d) Comparison of the sounding curves of the estimated MT TFs for the synthetic time-series data with the true sounding curves. Yellow areas indicate the period of 5.33 s, for which the data-adaptive weights of Case 2 and Case 3 are examined (Figs 5 and 6). In all figures, filled circles indicate the apparent resistivity and phase calculated from the estimated impedance tensor, and error bars (1.96σ) were computed from the standard errors of the impedance tensor components by the error propagation laws using the first-order Taylor expansion. Solid lines indicate the true sounding curves.
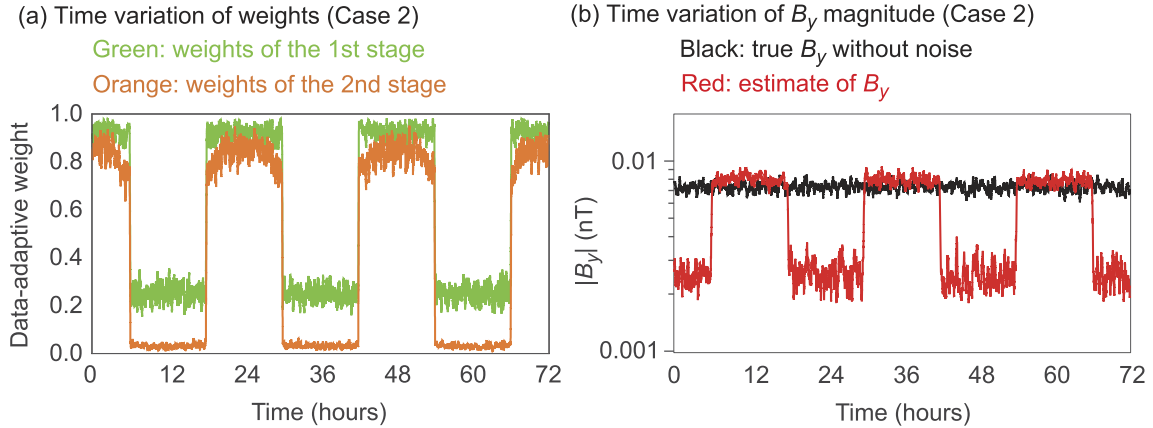
(a) Time variation of weights (Case 2)

Green: weights of the 1st stage

Orange: weights of the 2nd stage

(b) Time variation of B_y magnitude (Case 2)

Black: true B_y without noise

Red: estimate of B_y

Figure 5. (a) The data-adaptive weights used for the first and second stages of the two-stage BI method (Case 2) (Fig. 4b). The weights of the first stage are those for the bivariate equations whose dependent variable is B_y, and the weights of the second stage are those for the bivariate equations whose dependent variable is E_x. The time variations of the weights at a period of 5.33 s are shown. (b) Comparison of the magnitude of B_y estimated at the first stage of Case 2 (red line) with the true magnitude of B_y without noise (black line). The time variations of the magnitudes at a period of 5.33 s are shown. The moving averages using 101 points are used for all the figures to facilitate visualization because the short-term fluctuation makes the trend of the time variations unclear when the weight at each segment is directly used.

$$\hat{\boldsymbol{T}} = \begin{pmatrix} \hat{\boldsymbol{U}} \\ \hat{\boldsymbol{V}} \end{pmatrix} = \begin{pmatrix} t_{1,1} & t_{1,2} \\ \vdots & \vdots \\ t_{q,1} & t_{q,2} \end{pmatrix}, \tag{10}$$

$$\hat{\boldsymbol{Z}} = \hat{\boldsymbol{U}}\hat{\boldsymbol{V}}^{-1}, \tag{11}$$

where $\boldsymbol{v} \in \mathbb{C}^q$ is a complex vector comprising dependent variables, $\hat{\boldsymbol{T}} \in \mathbb{C}^{q \times 2}$ and $\hat{\boldsymbol{Z}} \in \mathbb{C}^{(q-2) \times 2}$ are estimates of $\boldsymbol{T}$ and $\boldsymbol{Z}$, respectively,

and $\boldsymbol{\varepsilon} \in \mathbb{C}^q$ denotes regression residuals. The integer q denotes the number of complex-dependent variables. In the case of estimating the impedance tensor, what is described in the following, q is four. Different from eq. (6), eq. (9) contains residual vector $\boldsymbol{\varepsilon}$ because noises in the data are considered in this equation. Van Aelst & Willems (2005) developed the robust multivariate regression S-estimator for multivariate regression using real variables. Therefore, the multivariate regression model used in this study (eq. 9),
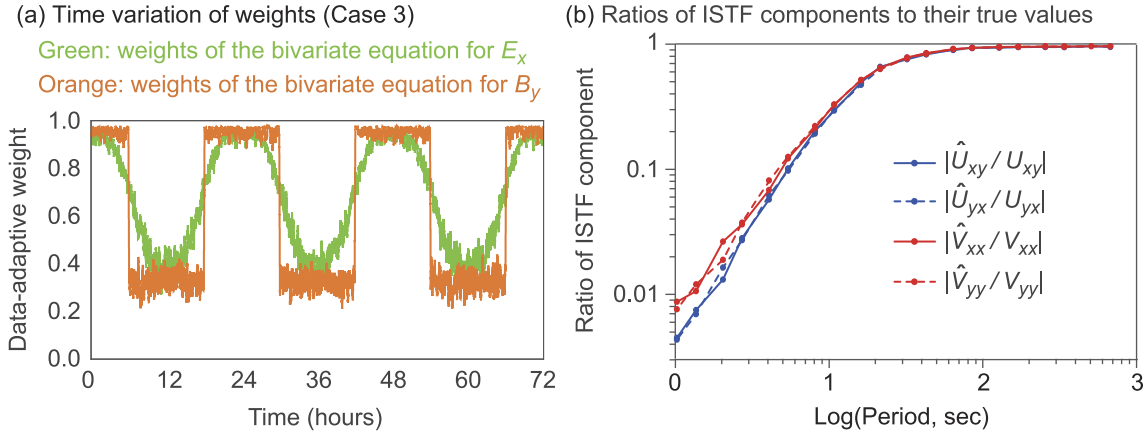
Figure 6. (a) The data-adaptive weights used for the bivariate equations for E_x and B_y in Case 3 (Fig. 4c). The time variations of the weights at a period of 5.33 s are shown. The moving averages using 101 points are used as in Fig. 6. (b) Magnitude ratios of the ISTF components (eq. 7) estimated in Case 3 to their true values. The magnitude ratios of the off-diagonal components of the pseudo-remote impedance tensor $\hat{U}$ and the diagonal components of the HMTF $\hat{V}$ are depicted.
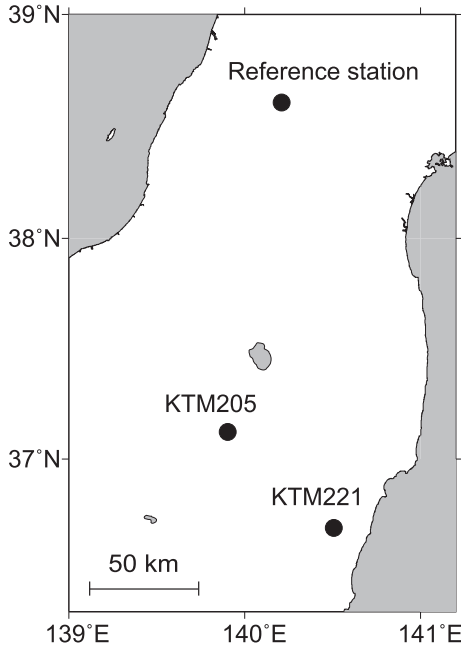

Figure 7. Locations of the MT stations (KTM221 and KTM205) and the reference station of the real-world data set used in this study. Those stations are located in Northeastern Japan.

which uses complex variables, is transformed into a corresponding real-valued model:

$$y = A^T x, \tag{12}$$

$$x = \begin{pmatrix} Re\,(B_{rx}) & Im\,(B_{rx}) & Re\,(B_{ry}) & Im\,(B_{ry}) \end{pmatrix}^T \in \mathbb{R}^4, \tag{13}$$

$$y = \begin{pmatrix} Re\,(v_1) & Im\,(v_1) & \cdots & Re\,(v_q) & Im\,(v_q) \end{pmatrix}^T \in \mathbb{R}^{2q}, \tag{14}$$

$$A = \begin{pmatrix} Re(t_{1,1}) & Im(t_{1,1}) & Re(t_{q,1}) & Im(t_{q,1}) \\ -Im\,(t_{1,1}) & Re\,(t_{1,1}) & -Im\,(t_{q,1}) & Re\,(t_{q,1}) \\ Re\,(t_{1,2}) & Im\,(t_{1,2}) \,... & Re\,(t_{q,2}) & Im\,(t_{q,2}) \\ -Im\,(t_{1,2}) & Re\,(t_{1,2}) & -Im\,(t_{q,2}) & Re\,(t_{q,2}) \end{pmatrix} \in \mathbb{R}^{4 \times 2q}, \tag{15}$$

where the elements of v are denoted by $v_1, \ldots, v_q$, and the real matrix A and vectors x, y are composed of the real and imaginary parts of the variables in eq. (9). Notably, each real and imaginary component appears twice in the coefficient matrix A: the elements of odd rows are the same as those of even rows, except for the sign of the imaginary components. The covariance matrix for the regression residuals is denoted by $C \in \mathbb{R}^{2q \times 2q}$. The S-estimator for multivariate regression is the solution to the problem of minimizing the determinant of C, subject to

$$\frac{1}{n_d} \sum_{i=1}^{n_d} \rho\,(d_i) = b_0, \tag{16}$$

$$d_i = \left(r_i^T C^{-1} r_i \right)^{1/2}, \tag{17}$$

$$r_i = y_i - A^T x_i, \tag{18}$$

(Van Aelst & Willems 2005). The subscript i in these equations indicates that the variables are calculated from the ith data segment. The real scalar variable d_i is the Mahalanobis distance (Maronna *et al.* 2019). The Mahalanobis distance is a commonly used indicator of outliers in multivariate statistics, and its usefulness for MT data processing was confirmed by Platz & Weckmann (2019). The integer n_d is the number of data segments used for the TF estimation. The loss function ρ for Tukey's bi-weight

$$\rho\,(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4}, & for \;\; |x| \leq c, \\ \frac{c^2}{6}, & for \;\; |x| > c, \end{cases} \tag{19}$$

is used in the same manner as in Lopuhaä (1989) and Van Aelst & Willems (2005). The parameter b_0 is selected to be the expectation of $\rho(d_i)$, as in Lopuhaä (1989), assuming that the underlying distribution of the Mahalanobis distance is a chi distribution with $2q$ degrees of freedom. The parameter c in eq. (19) is determined using the following equation:

$$\frac{b_0}{\sup \rho} = \frac{b_0}{c^2/6} = \frac{n_d - 2q}{2n_d}, \tag{20}$$

which leads to the maximum breakdown point, approximately 0.5 (Lopuhaä 1989).

Because of the dependence of the odd and even rows of matrix A, only the elements of the odd-numbered rows are treated as unknowns in the following. Specifically, using the matrix that extracts
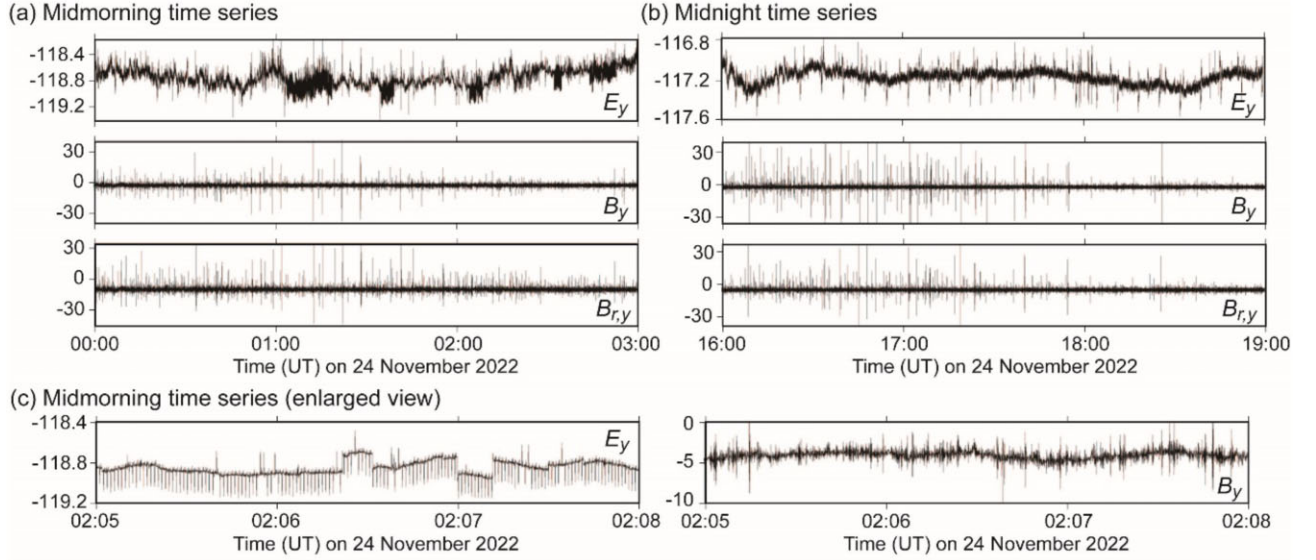
Figure 8. Time-series of E_y and B_y components at KTM221 and $B_{r,y}$ component at the reference station (Fig. 7) on 24 November 2022 (UT). The time range of (a), from 00:00 to 03:00 in UT, corresponds to local midmorning and the time range of (b), from 16:00 to 19:00 in UT, corresponds to local midnight. The unit of the vertical axes is not shown due to the raw time-series not having physical units because the instrument responses are not corrected. (c) Time-series of E_y and B_y components at KTM221 from 02:05 to 02:08 on 24 November 2022 (UT), during which spike and step noises appeared. Because the noises were most prominent in E_y-component, the y-components of the EM field data are shown in all panels.
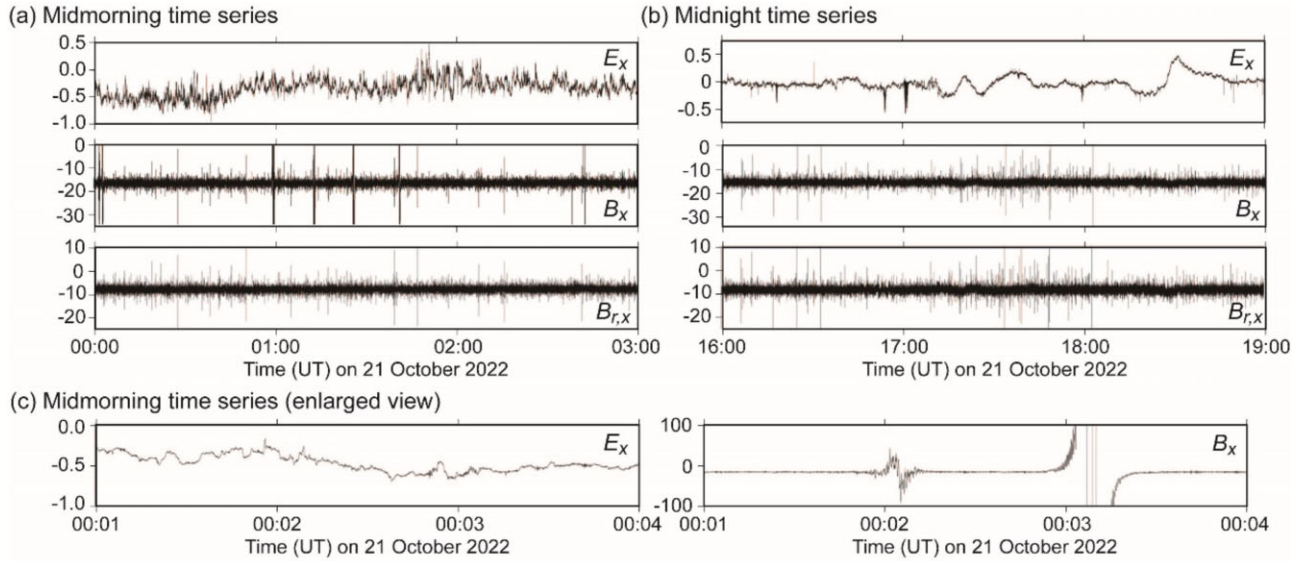

Figure 9. Time-series of E_x and B_x components at KTM205 and $B_{r,x}$ component at the reference station (Fig. 7) on 21 October 2022 (UT). The time range of (a), from 00:00 to 03:00 in UT, corresponds to local midmorning, and the time range of (b), from 16:00 to 19:00 in UT, corresponds to local midnight. The unit of the vertical axes is not shown due to the raw time-series not having physical units because the instrument responses are not corrected. (c) Time-series of E_x and B_x components at KTM205 from 00:01 to 00:04 on 21 October 2022 (UT), during which drastic changes in the B_x component occurred at around 00:02 and 00:03. Because the noises were most prominent in B_x component, the x-components of the EM field data are shown in all panels.

odd rows from A by left-multiplying,

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times4}, \tag{21}$$

the authors define an unknown vector comprising the independent elements of A:

$$t = vec\,(LA) \in \mathbb{R}^{4q}, \tag{22}$$

where the *vec* operator (Harville 1998) is used. The noise covariance matrix C is assumed to be diagonal, similar to that in Egbert (1997):

$$C = diag\left(\;\sigma_1^2\;\sigma_1^2\;\cdots\;\sigma_q^2\;\sigma_q^2\;\right) \in \mathbb{R}^{2q\times2q}, \tag{23}$$

implying that all noise is incoherent between the EM field components and that the real and imaginary parts of the residuals are independent and have the same variance. The latter condition has also been assumed by Chave & Thomson (1989) and Chave & Thomson (2004). Because C contains the same elements twice, the

(a) Case 1

Original remote reference method

(b) Case 2

Two-stage BI method

(c) Case 3

ISTF of each bivariate equation in eq. (6) was separately estimated by M-estimator
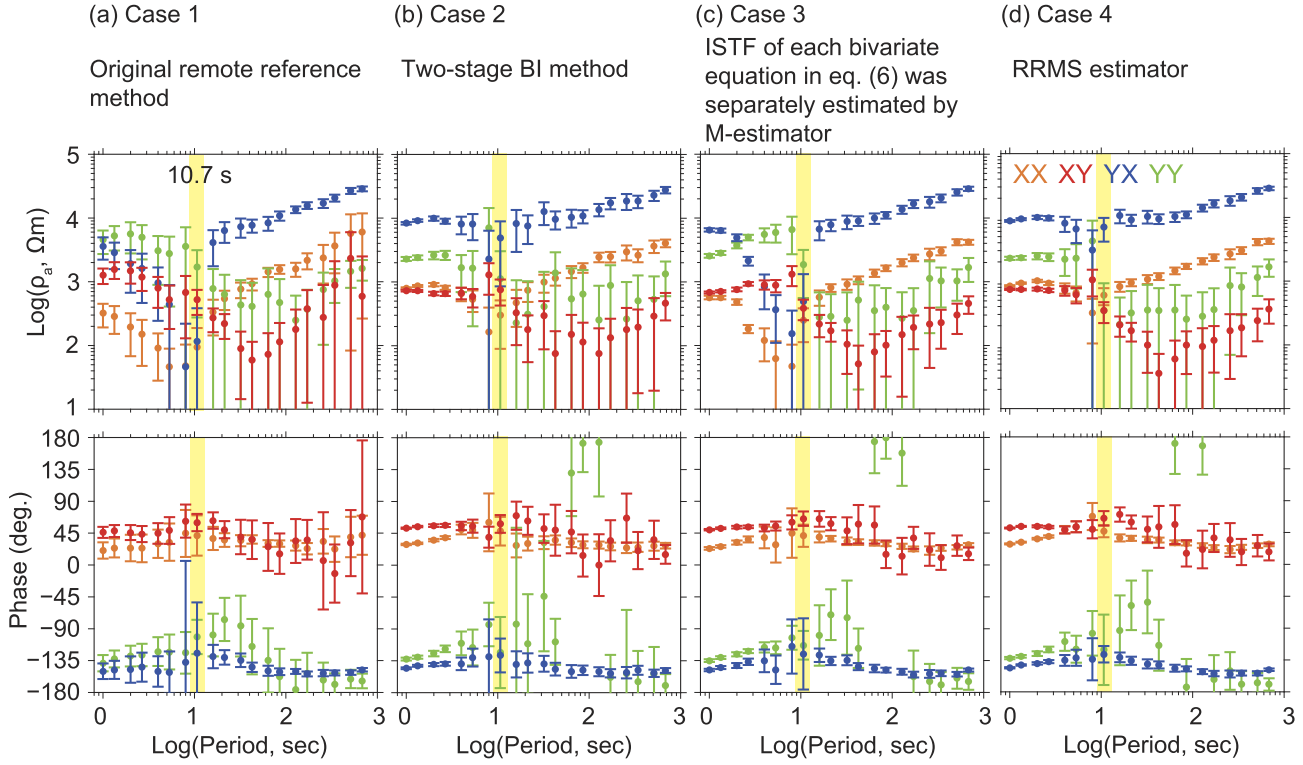
(d) Case 4

RRMS estimator

Figure 10. Sounding curves of the estimated MT TFs of KTM221 station. The filled circles represent the apparent resistivity and phase calculated by the estimated impedance tensor. The error bars (1.96σ) of the apparent resistivity and phase were computed from the standard errors of the impedance tensor components by the error propagation laws using the first-order Taylor expansion. Yellow portions indicate 10.7 s, at which the time variations of the apparent resistivity and the data-adaptive weights are examined in Fig. 11.

authors introduce a degenerated noise covariance matrix:

$$\tilde{C} = diag\left(\sigma_1^2 \cdots \sigma_q^2\right) \in \mathbb{R}^{q \times q}, \tag{24}$$

which has no redundant elements.

The multivariate regression S-estimator satisfies the first-order conditions (Lopuhaä 1989; Van Aelst & Willems 2005):

$$\frac{\partial \psi\left(t, C, \lambda\right)}{\partial t^T} = 0, \tag{25}$$

$$\frac{\partial \psi\left(t, C, \lambda\right)}{\partial \tilde{C}} = 0, \tag{26}$$

$$\psi\left(t, C, \lambda\right) = ln\left(|C|\right) - \lambda \left(\frac{1}{n_d}\sum_{i=1}^{n_d} \rho\left(d_i\right) - b_0\right), \tag{27}$$

where $\lambda \in \mathbb{R}$ denotes the Lagrange multiplier. As shown in Appendix A, eq. (25) leads to the estimate of the ISTF:

$$\hat{T} = \left(\sum_{i=1}^{n_d} w_i \boldsymbol{v}_i \boldsymbol{b}_{r,i}^H\right)\left(\sum_{i=1}^{n_d} w_i \boldsymbol{b}_{r,i} \boldsymbol{b}_{r,i}^H\right)^{-1} \in \mathbb{C}^{q \times 2}, \tag{28}$$

$$w_i = \frac{\rho'\left(d_i\right)}{d_i}. \tag{29}$$

Notably, by applying the multivariate regression S-estimator to the remote reference method, the present work derived a robust estimate of the ISTF, which has a familiar form of the weighted least-square solution. In addition, as shown in Appendix A, eq. (26) enables one to obtain a robust estimate of the noise variance:

$$\tilde{C} = \frac{q\sum_{i=1}^{n_d}\frac{\rho'(d_i)}{d_i}\mathcal{D}\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^H\right]}{\sum_{i=1}^{n_d}\rho'(d_i)\,d_i}, \tag{30}$$

where $\mathcal{D}: \mathbb{C}^{q \times q} \to \mathbb{C}^{q \times q}$ denotes the operator that extracts the diagonal elements from an input matrix. In other words, noise variances, σ_k^2 ($k = 1, \ldots, q$), can be calculated using

$$\sigma_k^2 = \frac{q\sum_{i=1}^{n_d} w_i \left|\varepsilon_{i,(k)}\right|^2}{\sum_{i=1}^{n_d} w_i d_i^2}, \tag{31}$$

where w_i denotes the weight defined in eq. (29).

In this study, instead of the noise covariance being calculated directly from eq. (30), an equivalent formulation is used, as in Bilodeau & Duchesne (2000) and Ruppert (1992), to prevent the smallest element of the noise covariance matrix from converging to zero (Maronna *et al.* 2019). In the alternative formulation, the scale of the Mahalanobis distance, denoted as $s \in \mathbb{R}$, is estimated as an unknown, in addition to the TF and noise variances. Specifically, the alternative formulation seeks the minimum s subject to

$$\frac{1}{n_d}\sum_{i=1}^{n_d} \rho\left(\frac{\delta_i}{s}\right) = b_0, \tag{32}$$

$$\delta_i = \left(\boldsymbol{r}_i^T \Sigma^{-1} \boldsymbol{r}_i\right)^{1/2}, \tag{33}$$

$$\Sigma = \frac{C}{|C|^{1/2q}}. \tag{34}$$

The last equation, eq. (34), imposes the constraint $|\Sigma| = 1$. The matrix Σ does not contain information on the scale of regression residuals, but only has information on the relative differences of the residuals. On behalf of Σ, the parameter s has information on the scale. Scale parameter s can be calculated using an M-scale estimator (Maronna *et al.* 2019):

(a) Case 2

Two-stage BI method

Weights for the bivariate equation whose dependent variable is E_y

(b) Case 3

ISTF of each bivariate equation in eq. (6) was separately estimated by M-estimator

Weights for the bivariate equation whose dependent variable is E_y

(c) Case 3

ISTF of each bivariate equation in eq. (6) was separately estimated by M-estimator

Weights for the bivariate equation whose dependent variable is B_x

(d) Case 4
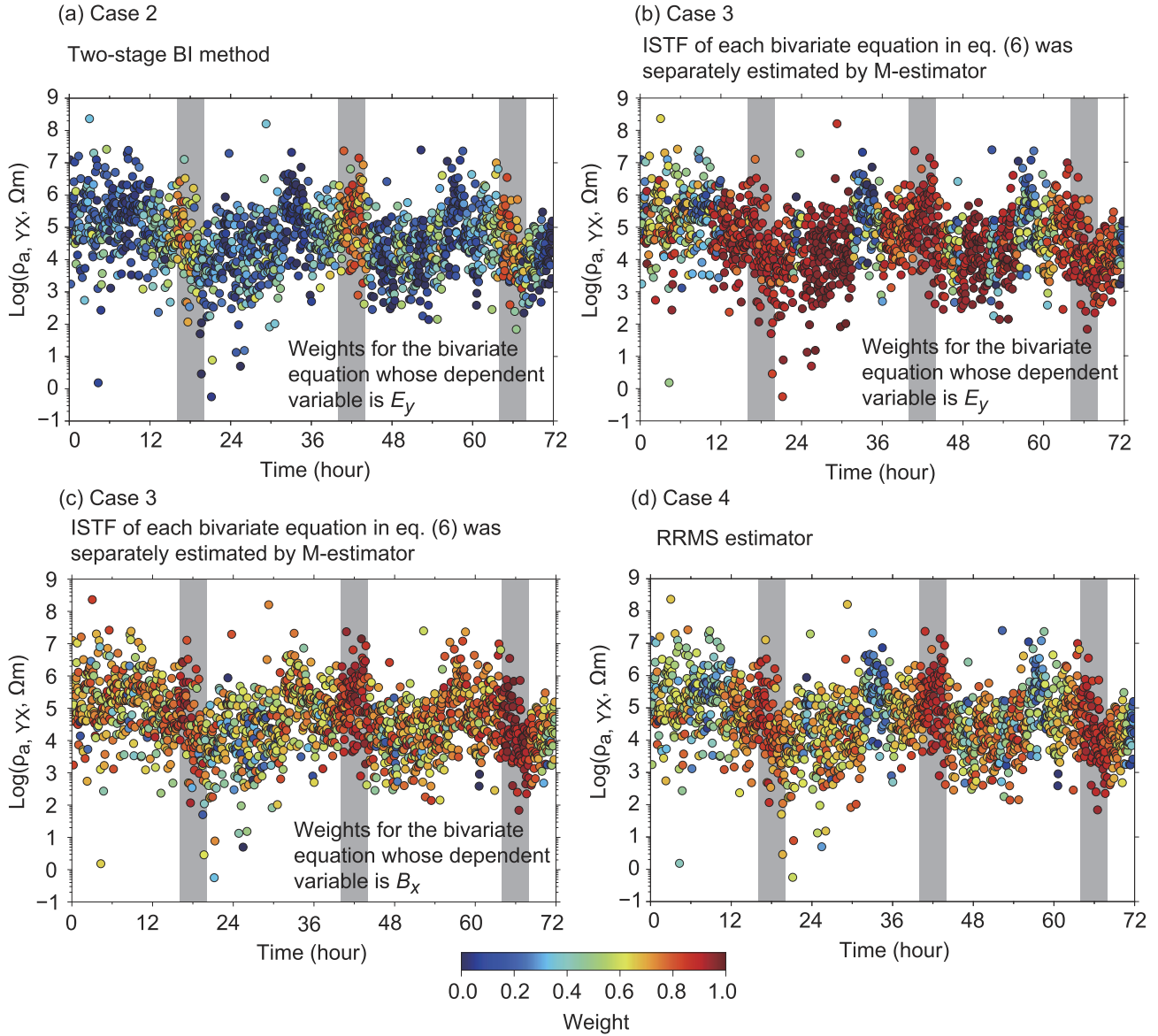
RRMS estimator

Figure 11. Time variation of the apparent resistivity of the Z_{yx} component of KTM221 at a period of 10.7 s. Adjacent ten data segments were used in computing each point of the apparent resistivity. The colours of points depend on the median of the weights of the ten segments used in the calculation of respective points. The horizontal axis is the span of the time-series (3 d). The grey-shaded areas correspond to local midnight (16:00–20:00 UT).

$$s = \sqrt{\frac{1}{n_d b_0} \sum_{i=1}^{n_d} v_i}, \qquad (35)$$

$$v_i = \begin{cases} \frac{\rho''(0)}{2}\delta_i^2 = 0 & if \ \delta_i = 0, \\ \frac{\rho(\delta_i/s)}{(\delta_i/s)^2}\delta_i^2 = \rho(\delta_i/s)s^2 & if \ \delta_i \neq 0. \end{cases} \qquad (36)$$

Eqs (35) and (36) are derived from eq. (32). The M-scale estimator, also called the M-estimator of scale, has been predominantly used in the S-estimator since Rousseeuw & Yohai (1984) proposed the S-estimator. Thus, the equations of the robust multivariate regression S-estimator are

$$\hat{T} = \left(\sum_{i=1}^{n_d} w(\delta_i, s) \, v_i \, b_{r,i}^H\right)\left(\sum_{i=1}^{n_d} w(\delta_i, s) \, b_{r,i} \, b_{r,i}^H\right)^{-1}, \qquad (37)$$

$$\Sigma = \frac{C}{|C|^{1/2q}}, \qquad (38)$$

$$s = \sqrt{\frac{1}{n_d b_0} \sum_{i=1}^{n_d} v(\delta_i, s)}, \qquad (39)$$

$$\delta_i = \left(r_i^T \Sigma^{-1} r_i\right)^{1/2}, \qquad (40)$$

$$\sigma_k^2 = \frac{q \sum_{i=1}^{n_d} w(\delta_i, s) \left|\varepsilon_{i,(k)}\right|^2}{\sum_{i=1}^{n_d} w(\delta_i, s)(\delta_i/s)^2} \quad (k = 1, \ldots, q), \qquad (41)$$

$$w(\delta_i, s) = \begin{cases} 1 & if \ \delta_i = 0, \\ \frac{\rho'(\delta_i/s)}{\delta_i/s} & if \ \delta_i \neq 0, \end{cases} \qquad (42)$$

$$v(\delta_i, s) = \begin{cases} 0 & if \ \delta_i = 0, \\ \rho(\delta_i/s)s^2 & if \ \delta_i \neq 0, \end{cases} \qquad (43)$$

where w and v are the weight functions. Appling the multivariate regression S-estimator to the remote reference method, the present work finally derived a set of robust estimates of the TF (eq. 37), noise variances (eq. 41), and the scale of the Mahalanobis distance
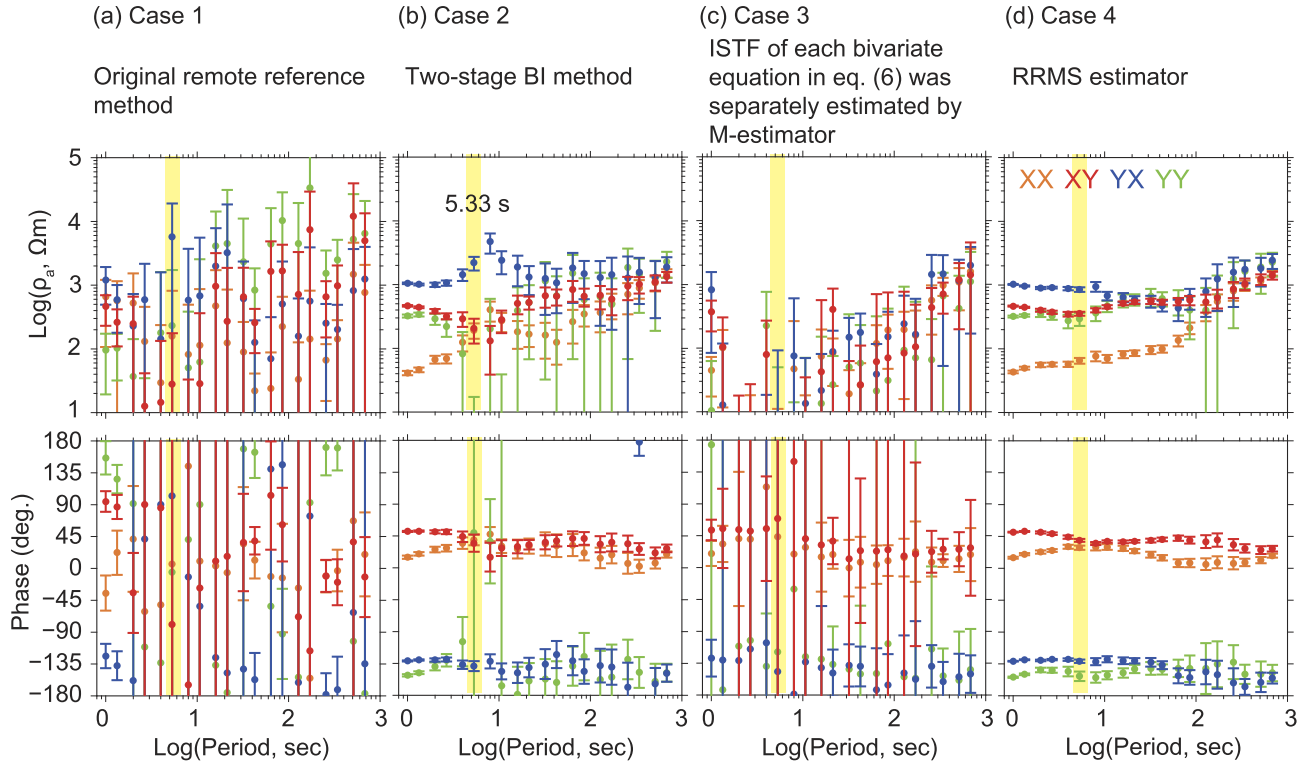
Figure 12. Sounding curves of the estimated MT TFs of KTM205 station. The filled circles represent the apparent resistivity and phase calculated by the estimated impedance tensor. The error bars (1.96σ) of the apparent resistivity and phase were computed from the standard errors of the impedance tensor components by the error propagation laws using the first-order Taylor expansion. Yellow portions indicate 5.33 s, at which the time variations of the apparent resistivity and the data-adaptive weights are examined in Fig. 13.

(eq. 39). These equations constitute a nonlinear system because the equations for updating the unknown parameters (eqs 37, – and 39) depend on the unknowns, which are mutually dependent. Thus, to estimate the TF $\hat{T}$ and the other unknown parameters, updating each parameter alternately and iteratively is necessary.

With the above equations, the developed method seeks the MT TF that minimizes s (eq. 39), a robust scale measure of the Mahalanobis distance. For avoiding the dependency of the final solution of a nonlinear system (eqs 37–43 in our case) on the initial estimate, a number of initial estimates are used in the S-estimator. To perform this procedure in an efficient way, the authors adopted the fast S algorithm of Salibián-Barrera & Yohai (2006). The detailed algorithm of our method is as follows:

Step 1. A number of initial estimates of the ISTF $\hat{T}$ are determined. Each initial estimate is calculated using two randomly selected data segments. The Mersenne Twister pseudorandom number generator (Matsumoto & Nishimura 1998; Nishimura 2000) is used for the random selection of data segments. In this study, 1000 initial estimates were used. In the 1000 candidates, one can select multiple pairs of clean data with a high probability. This is because, even if clean data account for only half of the whole data, the probability we cannot obtain a pair of clean data in 1000 trials is $(1-_{n_d/2}C_2/_{n_d}C_2)^{1000}$, which is approximately zero, even for the maximum n_d used in the present work (i.e. $n_d \cong 100\,000$). The authors used pairs of data segments because an estimate of the TF can be uniquely determined from two segments, and the probability one fails to obtain a set of clean segments increases as the number of segments increases. Although only two segments are used for determining the initial estimates, in subsequent steps, all segments

are used to calculate the estimates of the ISTF, noise variances, and scale parameter.

Step 2. Initial estimate of each noise variance is calculated as the square of the normalized median absolute deviation about the median (MADN) of regression residuals. The MADN is a robust alternative to the standard deviation (Maronna *et al.* 2019). In other words, the initial estimate of σ_k is the MADN of the real and imaginary parts of the kth component of ε. Subsequently, Σ is calculated by eq. (38). The initial estimate of scale s is the median of the Mahalanobis distances δ_i calculated from the initial estimates of $\hat{T}$ and Σ.

Step 3. Each set of initial estimates is iteratively updated. Each ISTF estimate after iterative updates can be regarded as an M-estimate of multivariate linear regression. In each iteration, the calculations of eqs (37), (38) and (39) are performed in that order. The maximum number of iterations is three. Although Salibián-Barrera & Yohai (2006) showed that even one iteration is effective in selecting the candidates to be further updated in Step 5, this work uses three iterations to ensure the improvement of the estimates. If the change in the weighted norm of each residual and the change in scale s are less than 5 per cent within one iteration, the iterative improvement is stopped before the maximum iteration number.

Step 4. Among all candidates, 10 sets of estimates with the smallest s values are selected.

Step 5. The selected candidates are further iteratively updated until the convergence. Specifically, iterative updates are performed until the change in the weighted norm of each residual and the change in scale s are less than 1 per cent within one iteration.

Step 6. The estimates with the smallest s among the selected candidates are chosen as the best estimates.
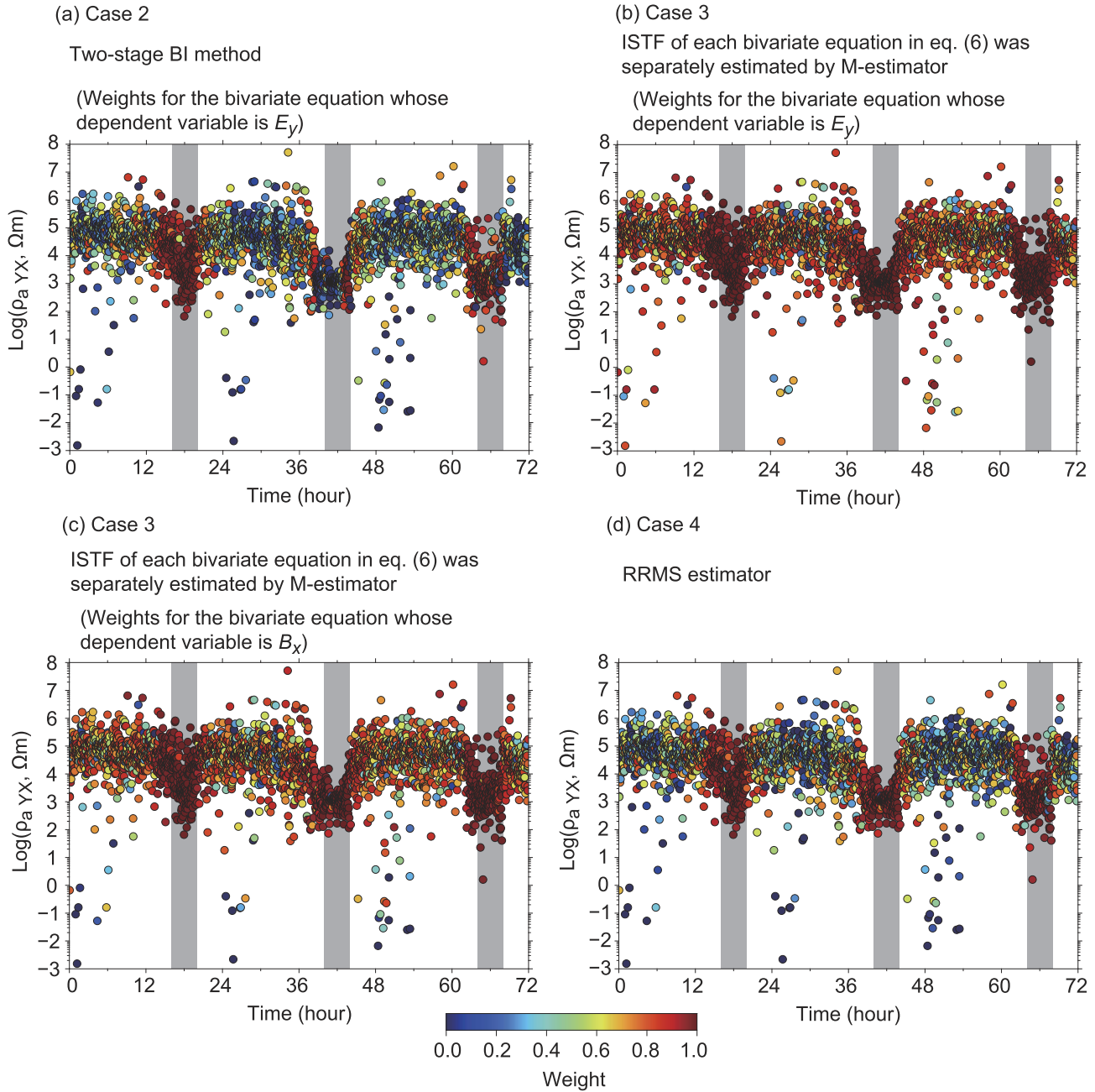
(a) Case 2

Two-stage BI method

(Weights for the bivariate equation whose dependent variable is E_y)

(b) Case 3

ISTF of each bivariate equation in eq. (6) was separately estimated by M-estimator

(Weights for the bivariate equation whose dependent variable is E_y)

(c) Case 3

ISTF of each bivariate equation in eq. (6) was separately estimated by M-estimator

(Weights for the bivariate equation whose dependent variable is B_x)

(d) Case 4

RRMS estimator

Figure 13. Time variation of the apparent resistivity of the Z_{yx} component of KTM205 at a period of 5.33 s. Adjacent ten data segments were used in computing each point of the apparent resistivity. The colours of points depend on the median of the weights of the ten segments used in the calculation of respective points. The horizontal axis is the span of the time-series (3 d). The grey-shaded areas correspond to local midnight (16:00–20:00 UT).

Step 7. From the best estimate of the ISTF, the impedance tensor is calculated using eq. (11).

The authors confirmed that scale *s* decreases monotonically by the algorithm, similar to the fast S algorithm (Salibián-Barrera & Yohai 2006). There are some adjustable parameters in the above algorithm. As for the number of initial estimates, it is desirable to select a large value to sufficiently reduce the probability that one fails to obtain a pair of clean data (e.g. lower than 1 per cent), as noted in the Step 1 description. One could use a smaller maximum iteration number and/or looser convergence criteria in Step 3. However, if one makes looser the convergence criteria in Step 3, it would be better to select

more candidates in Step 4. Even when the authors changed those parameters in several ways when analysing the data sets we show in the subsequent sections, it had little effect on the results, suggesting that the algorithm is not sensitive to those parameters.

Hereafter, the aforementioned TF estimation method is referred to as the robust remote reference multivariate regression S-estimator (RRMS estimator). The RRMS estimator can be applied to the general MT TF estimation, including the estimation of the vertical magnetic TF (in that case, the number of dependent variables *q* is three). Because there is no hyperparameter in the estimator, the RRMS estimator can eliminate outliers in an automated manner without expert knowledge of multivariate statics.

3 APPLICATION TO SYNTHETIC DATA

The authors applied the developed method to a synthetic time-series data set. Synthetic data were calculated using the Oblique Conductor model proposed by Tietze *et al.* (2015) (Fig. 1). In Fig. 1, the yellow star denotes the position of the assumed MT station, and the dark blue circle denotes the assumed reference station for the remote reference method. The impedance tensor of the MT station and the HMTF between the MT station and the reference station, which were used to construct the synthetic data, were computed using the hexahedral-mesh version of the FEMTIC code (Usui 2015; Usui *et al.* 2017). First, the authors calculated the magnetic field at the reference station. The time-series of the reference station was generated based on the algorithm of Timmer & Konig (1995), assuming red noise because the power of the geomagnetic field decays with the increase in frequency (Simpson & Bahr 2005). The assumed sampling frequency was 32 Hz, in other words, the interval of the time-series was 1/32 s. After applying the FFT to the reference geomagnetic field, the authors multiplied the resultant Fourier transforms by the HMTF to obtain the local magnetic field at the MT station. Subsequently, the electric field at the MT station was computed by multiplying the local magnetic field by the impedance tensor. Fig. 2 shows the power spectra of the EM field calculated using the aforementioned processes.

After transforming the local EM field to time-series data by the inverse FFT, the Gaussian and spike noises were added to each time-series. The Gaussian and spike noises of each channel and each station were mutually independent. Because the power of cultural noises generally varies with time, the authors changed the amplitude of the Gaussian noise with time. Specifically, the authors multiplied sinusoidal and rectangular amplification factors to the Gaussian noises in the electric field and magnetic field data, respectively (Fig. 3a). Spike noise was observed with a probability of one ten-thousandth in the synthetic time-series. Fig. 3 shows the resultant synthetic time-series data of the E_x and B_x channels of the MT station. Figs 2(d)–(f) illustrate the powers of the time-series contaminated by noises. The power of the noises is larger than that of the signals at periods of less than approximately 30 s for the magnetic field data. However, the noise in the electric field data influences the data at a longer period range: the noise power is larger than that of signals at periods less than approximately 100 s. The noises of the synthetic data are simple compared to the noises in actual MT data. On the other hand, verifications using actual data have a drawback in that we cannot know how much an estimated TF is different from the true solution. That is because we do not know the true solution for real-world data. The main objective of this synthetic data analysis is to confirm that the RRMS estimator can give unbiased estimates.

The length of the time-series used in the analysis was 72 hr (3 d). The length of the data segment was gradually halved from 65 536 to 128. The authors compared the impedance tensor estimated by four different estimators. The cases using the original remote reference method (Goubau *et al.* 1978; Gamble *et al.* 1979), the two-stage BI method (Chave & Thomson 2004), and the RRMS estimator are denoted as Case 1, Case 2 and Case 4, respectively. In addition, to confirm the importance of using robust multivariate regression, the authors performed the case in which each bivariate equation of the two-input–multiple-output system (eq. 6) was solved separately with the M-estimator (Case 3), that is the bivariate equations for E_x, E_y, B_x and B_y were independently solved. After solving all bivariate equations, the impedance tensor was calculated by eq. (8) as in the RRMS estimator. In Cases 2 and 3, following Chave & Thomson

(1989), the authors used the median absolute derivation from the median to calculate the scale estimate of regression residuals. To avoid the influence of the difference in the weight function, the authors used Tukey's bi-weight function (Maronna *et al.* 2019) for Cases 2, 3 and 4. In Cases 2 and 3, the authors determined the parameter c in eq. (19) by inserting $q = 2$ into eq. (20) because the dependent variable is one complex variable in those cases. Except for the impedance tensor estimation process, the authors used the same procedures for all the estimators. The authors calculated the standard errors of the estimates of the impedance tensor by the bootstrap method, as Neukirch & Garcia (2014) did. In Fig. 4, the estimated apparent resistivity and phase are compared with the true TFs (Fig. 1). The errors of the estimated apparent resistivity and phase were computed from the errors of the impedance tensor by the error propagation laws using the first-order Taylor expansion.

Significant differences were observed among the different TF estimation methods at periods shorter than several tens of seconds (Fig. 4). In the case of the non-robust remote reference method (Case 1), the resultant impedance tensor components were scattered from the true values, especially at periods shorter than 5 *s* (Fig. 4a). Some of the estimated values were out of the vertical-axis range at the shortest two periods. In the other cases, deviations at the shortest two periods were relatively small. However, when we used the two-stage BI method (Case 2), the apparent resistivity was overestimated at periods of around 5 s (Fig. 4b). Conversely, Fig. 4(c) shows that the apparent resistivity was underestimated at periods shorter than several tens of seconds in Case 3. Figs 5a and 6a show the time variations of the data-adaptive weights of Case 2 and Case 3 at a period of 5.33 *s*. These figures illustrate that the data-adaptive weights depended on time, and relatively low weights were assigned to the data at the peaks of the Gaussian noise (Fig. 3a). In Case 2, the weights of the second stage were lower than those of the first stage, especially in the time ranges where the level of noise contamination was high (Fig. 5a), indicating that eq. (4) was not satisfied. Fig. 5b, on the other hand, compares the magnitude of the B_y estimate determined at the first stage with the true magnitude of B_y without noise. B_y was significantly underestimated in the relatively clean time ranges (Fig. 3a), although B_y was subtly overestimated in the time ranges with high noise levels. It is considered that the downward bias of the HMTF due to the noises in the reference magnetic field caused the underestimation of the magnitude of the local magnetic field. The underestimation of the local magnetic field should lead to the overestimation of the apparent resistivity (Fig. 4b).

Fig. 6(b) shows the magnitude ratios of the ISTF components (eq. 7) estimated in Case 3 to their true values. It depicts the magnitude ratios of the off-diagonal components of the pseudo-remote impedance tensor $\hat{U}$ and the diagonal components of the HMTF $\hat{V}$. Both $\hat{U}$ and $\hat{V}$ were underestimated due to the noises in the reference magnetic field. The level of underestimation was higher for $\hat{U}$ compared to $\hat{V}$. It would appear that the difference in the data-adaptive weights (Fig. 6a) caused different biases in $\hat{U}$ and $\hat{V}$, resulting in the underestimation of the apparent resistivity (Fig. 4c) in Case 3. It is trivial that when we used the RRMS estimator (Case 4), the same weights were applied for all dependent variables in eq. (6). There was no significant bias in the TFs estimated by the original remote reference method (Case 1) and the RRMS estimator (Case 4, Figs 4a and d). Those results prove that the RRMS estimator, as well as the original remote reference method, can give unbiased solutions.

4 APPLICATION TO REAL-WORLD DATA

The authors applied the newly developed method to real-world MT data measured at three MT stations in Northeastern Japan (Fig. 7). KTM221 and KTM205 are local MT stations. As a remote reference station, the authors used a station located 150 km or more away from the local MT stations. The ADU (Metronix) measurement systems were used in all stations. The authors used the 32 Hz sampling data for 3 d: the data from 23 to 25 November 2022 (UT) at KTM221 and the data from 21 to 23 October 2022 (UT) at KTM205. Examples of time-series are shown in Figs 8 and 9. Figs 8(a) and 9(a) represent the time-series at midmorning in local time, whereas Figs 8b and 9b represent the time-series at local midnight. KTM221 is a moderately noisy station on a field of grass. The spike and step noises in the electric field (Fig. 8c) might stem from electric fences near the station. KTM205 was located in a noisier place than KTM221. Its location was close to a small hydraulic plant, and motor vehicles occasionally ran a forest road near the station. Remarkable noises in B_x component shown in Fig. 9(c), whose amplitudes were over 100 times larger than those of the magnetic field signals, are thought to stem from running motor vehicles.

The authors estimated the impedance tensor of KTM221 and KTM205 using the same calculation conditions as in the analyses of the synthetic data set (Section 3). Fig. 10 illustrates the estimated sounding curves of KTM221. Although all estimators gave similar results at periods longer than several tens of seconds, there were noticeable differences in the TFs at periods of about 10 s. In Cases 1 and 3, the apparent resistivity at periods from 4 to 10.7 s appeared to change discontinuously from those at longer periods. Fig. 11 compares the time variations of the apparent resistivity of the Z_{yx} component, whose magnitude was the largest among the impedance tensor components, with the data-adaptive weights at a period of 10.7 s. Although all estimators gave relatively high weights to the data during local midnight (remarked by grey-shaded areas), weights varied among different cases (Fig. 11). It should be remembered that, in Case 3, the weights were separately determined for respective bivariate equations in eq. (6). Figs 11(b) and (c) show the weights determined for the bivariate equations whose dependent variables were E_y and B_x, respectively. The differences in biases among the ISTFs might have caused the bias in the apparent at periods from 4 to 10.7 s (Fig. 10c), as in the synthetic data example. The weights of Case 4 (Fig. 11d) had combined features of the weights shown in Figs 11(b) and (c), demonstrating that the RRMS estimator determines the weights by equally valuing outlying data in the electric and magnetic fields. Comparing the TFs of Cases 2 and 4 (Figs 10b and d), we can find that the error bars of Case 2 are larger than those of Case 4. The relatively large error bars of Case 2 were due to the lower weights determined in this case (Fig. 11a) than those in Case 4 (Fig. 11d). It is considered that the multiplication of weight functions in the two-stage BI method determined excessively lower weights.

Fig. 12 illustrates the estimated sounding curves of KTM205. When the non-robust remote reference method was used (Case 1), the resultant sounding curves were discontinuous, and the error bars were large compared to the variations of the TFs (Fig. 12a). This result demonstrates that the second data set was heavily contaminated by noises. The apparent resistivity of Case 3 was notably lower than those of the other cases (Fig. 12). Fig. 13 compares the time variations of the apparent resistivity of the Z_{yx} component, whose magnitude was the largest among the impedance tensor components, with the data-adaptive weighs at a period of 5.33 s. Fig. 13(b)

shows that the data giving underestimated apparent resistivity lower than 10 Ωm were not downweighted when only the electric field was used as the dependent variable, whereas low weights were assigned to those data when we solved the bivariate equation for B_x (Fig. 13c). The differences in biases among the ISTFs due to such weight differences might have contributed to the underdetermined apparent resistivity of Case 3 (Fig. 12c), as in the case with the synthetic data example. The RRMS estimator gave low weights to those data (Fig. 13d), as in the case where we solved the bivariate equation for B_x (Fig. 13c).

The yx-component had a peak at periods from 5 to 10 s in Case 2 (Fig. 12b). In contrast, there was not such a peak in Case 4 (Fig. 12d). Fig. 13(a) shows that, in Case 2, the weights were low at the central midnight (at about 40–44 hr from the beginning of the time-series), whereas Fig. 13(d) illustrates that the weights were high in the same time range in Case 4. The geomagnetic activity was relatively high during the time range. The K indexes in the time range at Kakioka Magnetic Observatory, a member observatory of International Real-time Magnetic Observatory Network (INTERMAGNET), were greater than or equal to 4 (Kakioka Magnetic Observatory 2022). The leverage weight in the two-stage BI method (Chave & Thomson 2004) was considered to downweight the relatively high-power data, which were possibly high signal-to-noise ratio data, allowing the overestimation of the apparent resistivity of the yx-component (Fig. 12b). On the other hand, since those high-power data were not determined as outliers based on residuals, the RRMS estimator gave high weights to the data in all midnight times, preventing the overestimation of the apparent resistivity (Fig. 12d).

5 CONCLUSIONS

The authors developed a new robust remote reference estimator with the aid of robust multivariate linear regression. The authors applied the robust multivariate linear regression S-estimator (Van Aelst & Willems 2005) to the two-input–multiple-output relationship between the local EM field and the reference magnetic field that leads to the same equation as by the original remote reference method (Goubau *et al.* 1978; Gamble *et al.* 1979). As a result, this work obtained a set of robust estimates of the TF, noise variances, and the scale of the Mahalanobis distance (Maronna *et al.* 2019; Platz & Weckmann 2019). The examinations using synthetic and actual MT data sets demonstrate that the RRMS estimator can give an unbiased estimate of MT TF and suppress the influence of outliers in the local electric field and magnetic field.

However, the authors do not pretend that the RRMS estimator can always provide accurate TFs from any MT data set. If influential noises, including non-MT external sources, account for a large portion of the remote reference data, the RRMS can fail to provide accurate estimates. In such a case, for example, the methods of Egbert (1997) and Smirnov & Egbert (2012) are expected to be more effective. However, it seems inappropriate to make a simple comparison between their methods and our estimator. The former estimators utilized the principal component analysis. On the other hand, what the authors pursue in this work is the extension of the robust remote reference method using robust linear regression. Recently, other sophisticated MT TF estimation methods have been developed, for example, the methods using the independent component analysis (Sato *et al.* 2021; Ogawa *et al.* 2023) and the magnetic polarization direction (Weckmann *et al.* 2005; Platz & Weckmann 2019). Some of the methods (e.g. the signal-noise separation method of Sato *et al.* 2021) can be combined with the RRMS

estimator. Such combined use may help us to estimate TFs more accurately.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Yoshiya Usui (Data curation [Equal], Formal analysis [Lead], Funding acquisition [Supporting], Investigation [Lead], Methodology [Lead], Resources [Equal], Software [Lead], Validation [Lead], Visualization [Lead], Writing – original draft [Lead], Writing – review & editing [Lead]), Makoto Uyeshima (Data curation [Lead], Funding acquisition [Lead], Resources [Lead], Writing – review & editing [Equal]), Shin'ya Sakanaka (Data curation [Equal], Funding acquisition [Equal], Resources [Equal], Writing – review & editing [Equal]), Tasuku Hashimoto (Data curation [Equal], Resources [Equal], Writing – review & editing [Equal]), Masahiro Ichiki (Data curation [Equal], Funding acquisition [Equal], Resources [Equal], Writing – review & editing [Equal]), Toshiki Kaida (Data curation [Equal], Resources [Equal], Writing – review & editing [Equal]), Yusuke Yamaya (Data curation [Equal], Funding acquisition [Equal], Resources [Equal], Writing – review & editing [Equal]), Yasuo Ogawa (Data curation [Equal], Funding acquisition [Equal], Resources [Equal], Writing – review & editing [Equal]), Masataka Masuda (Data curation [Equal], Resources [Equal], Writing – review & editing [Equal]), and Takahiro Akiyama (Data curation [Equal], Resources [Equal], Writing – review & editing [Equal])

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Supplementary_Material_R4.docx

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

DATA AVAILABILITY

The code and the synthetic data used in this paper will be shared upon request to the authors.

REFERENCES

Bendat, J.S. & Piersol, A.G., 2010. *Random Data: Analysis and Measurement Procedures,* 4th edn, John Wiley & Sons.

Bilodeau, M. & Duchesne, P., 2000. Robust estimation of the SUR model, *Can. J. Stat.,* **28,** 277–288.

Campanyà, J., Ledo, J., Queralt, P., Marcuello, A. & 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., 2012. Estimation of the magnetotelluric response, in *The Magnetotelluric Method: Theory and Practice,* pp. 165–218, eds Chave, A.D. & Jones, A.G., Cambridge Univ. Press.

Chave, A.D. & Thomson, D.J., 1989. Some comments on magnetotelluric response function estimation, *J. geophys. Res.,* **94**(B10), 14 215–14 225.

Chave, A.D. & Thomson, D.J., 2004. Bounded influence magnetotelluric response function estimation, *Geophys. J. Int.,* **157,** 988–1006.

Egbert, G.D., 1997. Robust multiple-station magnetotelluric data processing, *Geophys. J. Int.,* **130,** 475–496.

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

Ferguson, I.J., 2012. Instrumentation and field, in *The Magnetotelluric Method: Theory and Practice,* pp. 421–479, eds Chave, A.D. & Jones, A.G., Cambridge Univ. Press.

Flores, S.(2015). SOCP relaxation bounds for the optimal subset selection problem applied to robust linear regression, *Eur. J. Oper. Res.,* **246**(1), 44–50.

Gamble, T.D., Goubau, W.M. & Clarke, J., 1979. Magnetotellurics with a remote reference, *Geophysics,* **44,** 53–68.

Goubau, W., Gamble, T. & Clarke, J., 1978. Magnetotelluric data analysis: removal of bias, *Geophysics,* **43,** 1157–1166.

Harville, D.A., 1998. *Matrix Algebra from a Statistician's Perspective,* Springer.

Jalali-Heravi, M. & Konuze, E., 2002. Use of quantitative structure-property relationships in predicting the Krafft point of anionic surfactants, *Internet Electron. J. Mol. Des.,* **1**(8), 410–417.

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

Kakioka Magnetic Observatory, Japan Meteorological Agency, 2022, https://www.kakioka-jma.go.jp/obsdata/K_index/K_index.php

Larsen, J., 1989. Transfer functions: smooth robust estimates by least-squares and remote reference methods, *Geophys. J. Int.,* **99,** 645–663.

Larsen, J., Mackie, R., Manzella, A., Fiordelisi, A. & Rieven, S., 1996. Robust smooth magnetotelluric transfer functions, *Geophys. J. Int.,* **124,** 801–819.

Lopuhaä, H.P., 1989. On the relation between S-estimators and M-estimators of multivariate location and covariance, *Ann. Statist.,* **17,** 1662–1683.

Marin, J.M. & Robert, C.P., 2007. *Bayesian Core: A Practical Approach to Computational Bayesian Statistics,* Springer.

Maronna, R.A., Martin, R.D., Yohai, V.J. & Salibián-Barrera, M., 2019. *Robust Statistics: Theory and Methods (with R),* 2nd edn, John Wiley & Sons.

Martin, R.D. & Thomson, D.J., 1982. Robust-resistant spectrum estimation, *Proc. IEEE,* **70,** 1097–1115.

Matsumoto, M. & Nishimura, T., 1998. Mersenne Twister: a 623-dimensionally equidistributed uniform pseudorandom number generator, *ACM Trans. Model. Comput. Simul.,* **8,** 3–30.

Muñoz, G. & Ritter, O., 2013. Pseudo-remote reference processing of magnetotelluric data: a fast and efficient data acquisition scheme for local arrays, *Geophys. Prospect.,* **61,** 300–316.

Neska, A., 2006. Remote reference versus signal-noise separation: a least-square based comparison between magnetotelluric processing techniques, *PhD thesis,* Institut fur Geologische Wissenschaften, Freie Universitat Berlin.

Neukirch, M. & Garcia, X., 2014. Nonstationary magnetotelluric data processing with instantaneous parameter, *J. geophys. Res.,* **119,** 1634–1654.

Nishimura, T., 2000. Tables of 64-bit Mersenne twisters, *ACM Trans. Model. Comput. Simul.,* **10,** 348–357.

Ogawa, H., Asamori, K., Negi, T. & Ueda, T., 2023. A novel method for processing noisy magnetotelluric data based on independence of signal sources and continuity of response functions, *J. appl. Geophys.,* **213,** doi:10.1016/j.jappgeo.2023.105012.

Percival, D. & Walden, A., 1993. *Spectral Analysis for Physical Applications,* Cambridge Univ. Press.

Platz, A. & Weckmann, U., 2019. An automated new pre-selection tool for noisy magnetotelluric data using the Mahalanobis distance and magnetic field constraints, *Geophys. J. Int.,* **218,** 1853–1872.

Ritter, O., Junge, A. & Dawes, G.J., 1998. New equipment and processing for magnetotelluric remote reference observations, *Geophys. J. Int.,* **132,** 535–548.

Rousseeuw, P.J. & Yohai, V., 1984. *Robust Regression by Means of S-Estimators in Robust and Nonlinear Time-Series Analysis,* pp. 256–272, eds Franke, J., Hardle, W. & Martin, D., Springer-Verlag.

Ruppert, D., 1992. Computing S estimators for regression and multivariate location/dispersion, *J. Comput. Graph. Statist.,* **1,** 253–270.

Salibián-Barrera, M., Van Aelst, S. & Willems, G., 2008. Fast and robust bootstrap, *Stat. Methods Appt.,* **17,** 41–71.

Salibián-Barrera, M. & Yohai, V., 2006. A fast algorithm for S-regression estimates, *J. Comput. Graph. Statist.,* **15,** 414–427.

Sato, S., Goto, T.N., Kasaya, T. & Ichihara, H., 2021. Method for obtaining response functions from noisy magnetotelluric data using frequency-domain independent component analysis, *Geophysics,* **86,** E21–E35.

Schmucker, U., 1984. EM Ubertragungtungen aus Beobachtungen mit mehreren gleichzeitig registrierenden Stationen, in *Kolloquium Elektromagnetische Tiefen-forschung,* pp. 35–36, Grafrath in Oberbayern.

Simpson, F. & Bahr, K., 2005. *Practical Magnetotellurics,* Cambridge Univ. Press.

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

Smirnov, M.Y. & Egbert, G.D., 2012. Robust principal component analysis of electromagnetic arrays with missing data, *Geophys. J. Int.,* **190,** 1423–1438.

Szarka, L., 1988. Geophysical aspects of man-made electromagnetic noise in the Earth—a review, *Surv. Geophys.,* **9,** 287–318.

Tietze, K., Ritter, O. & Egbert, G.D., 2015. 3-D joint inversion of the magnetotelluric phase tensor and vertical magnetic transfer functions. *Geophys. J. Int.,* **203,** 1128–1148.

Timmer, J. & Konig, M., 1995. On generating power law noise, *Astron. Astrophys.,* **300,** 707–710.

Usui, Y., 2015. 3-D inversion of magnetotelluric data using unstructured tetrahedral elements: applicability to data affected by topography, *Geophys. J. Int.,* **202,** 828–849.

Usui, Y., Ogawa, Y., Aizawa, K., Kanda, W., Hashimoto, T., Koyama, T., Yamaya, Y. & Kagiyama, T., 2017. Three-dimensional resistivity structure of Asama Volcano revealed by data-space magnetotelluric inversion using unstructured tetrahedral elements, *Geophys. J. Int.,* **208,** 1359–1372.

Van Aelst, S. & Willems, G., 2005. Multivariate regression S-estimators for robust estimation and inference, *Stat. Sin.,* **15,** 981–1001.

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.

Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J. & Wobbe, F., 2013. Generic mapping tools: improved version released, *EOS, Trans. Am. geophys. Un.,* **94,** 409–410.

APPENDIX A: COMPARISON OF M- AND S-ESTIMATORS FOR A SIMPLE LINEAR REGRESSION

In this appendix, the authors demonstrate the differences between the M- and S-estimators using a simple real-valued linear regression problem. The authors applied the M- and S-estimators for
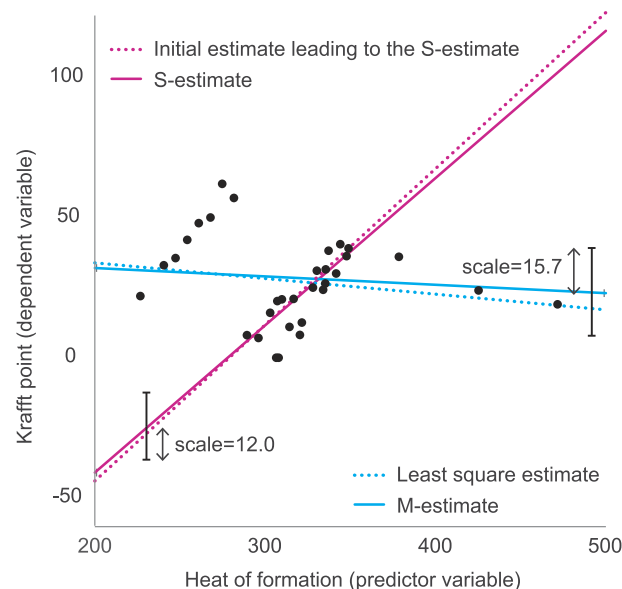

Figure A1. Example of the M-estimate and the S-estimate for simple linear regression. The black dots are the points of the Krafft point data (Jalali-Heravi & Konouz 2002; Flores 2015). The M-estimator gives the M-estimate (solid blue line) by the iteratively reweighted least-squares method from the least-square estimate (dotted blue line). The S-estimator compares estimates starting from a number of different initial estimates and selects the estimate with the smallest robust scale of residuals as the preferred one (S-estimate). The solid red line indicates the S-estimate and the dotted red line is the initial estimate resulting in the S-estimate. Black vertical bars illustrate the robust scale estimates for the M- and S-estimates.

the univariate real-valued linear regression to the Krafft point data (Jalali-Heravi & Konouz 2002; Flores 2015), 32 points data of a quantity called Krafft point versus a molecular descriptor called heat of formation. The black dots in Fig. A1 show the Krafft point data. In addition to the main cluster near the center of the figure, there are outlying data at the right and a smaller cluster at the left.

The ordinary least-square estimate, consisting of the slope and intercept, did not fit the main cluster (the dotted blue line in Fig. A1). When the authors started the iteratively reweighted least-squares method (IRLS) of the M-estimator from the least-square estimate, that is one of the frequently used manners (e.g. Chave & Thomson 1989), the resultant M-estimate (the solid blue line in Fig. A1) also did not conform with the main cluster. That was likely because the rightmost and leftmost points were not detected as outliers (large residual points) since those points were close to the regression line of the least-square estimate. When applying the S-estimator, the authors determined the initial estimates from all possible combinations of two different data points and started the IRLS from each of them. The resultant S-estimate (the solid red line in Fig. A1), which gave the minimum value of the robust scale estimate of regression residuals, conformed with the main cluster. The leftmost and rightmost data points were detected as outliers when the initial estimate leading to the S-estimate (the dotted red line in Fig. A1) was used. It should be noted that the same IRLS algorithm was used for the M- and S-estimators. In the IRLS, the authors used Tukey's biweight function and the M-scale estimator (Maronna *et al.* 2019).

APPENDIX B: DERIVATION OF ESTIMATES OF TRANSFER FUNCTION AND NOISE VARIANCES

In this appendix, the authors derived the estimates of ISTF and noise variances from the first-order conditions (eqs 25, 26 and 27). The left-hand side of eq. (25) can be rearranged as

$$-\frac{\lambda}{n_d}\sum_{i=1}^{n_d}\frac{\partial\rho(d_i)}{\partial\boldsymbol{t}^T}=-\frac{\lambda}{n_d}\sum_{i=1}^{n_d}\rho'(d_i)\frac{\partial d_i}{\partial\boldsymbol{t}^T}. \quad (B1)$$

The partial derivatives of the Mahalanobis distance, with respect to the unknown vector $\boldsymbol{t}$, can be transformed into

$$\frac{\partial d_i}{\partial\boldsymbol{t}^T}=-\frac{1}{d_i}\frac{\partial(\boldsymbol{A}^T\boldsymbol{x}_i)^T}{\partial\boldsymbol{t}^T}\boldsymbol{C}^{-1}\boldsymbol{r}_i. \quad (B2)$$

In the following, $a_{i,j}$ denotes the (i,j) element of $\boldsymbol{A}$ (eq. 15). The elements of $\boldsymbol{A}$ satisfy the relationship

$$a_{2k,2\ell}=a_{2k-1,2\ell-1}, \quad (B3)$$

$$a_{2k,2\ell-1}=-a_{2k-1,2\ell}, \quad (B4)$$

where k $(k=1,\ldots,2)$ and ℓ $(\ell=1,\ldots,q)$ are the integer indices. The partial derivatives of $\boldsymbol{A}^T\boldsymbol{x}_i$, with respect to the elements of the $(2\ell-1)$th or 2ℓth columns of $\boldsymbol{A}$, can have nonzero values only at their $(2\ell-1)$th and 2ℓth elements:

$$\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k,2\ell}}=\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k-1,2\ell-1}}=\begin{pmatrix}0 & \cdots & Re(B_{r,i,(k)}) & Im(B_{r,i,(k)}) & \cdots & 0\end{pmatrix}^T, \quad (B5)$$

$$\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k,2\ell-1}}=-\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k-1,2\ell}}=\begin{pmatrix}0 & \cdots & Im(B_{r,i,(k)}) & -Re(B_{r,i,(k)}) & \cdots & 0\end{pmatrix}^T, \quad (B6)$$

where $B_{r,i,(k)}$ is the kth element of the predictor vector of the ith data segment. Notably, the derivatives with respect to the elements of the first or second columns can have non-zero values only in the first two elements:

$$\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k,2}}=\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k-1,1}}=\begin{pmatrix}Re(B_{r,i,(k)}) & Im(B_{r,i,(k)}) & 0 & \cdots & 0\end{pmatrix}^T, \quad (B7)$$

$$\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k,1}}=\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k-1,2}}=\begin{pmatrix}Im(B_{r,i,(k)}) & -Re(B_{r,i,(k)}) & 0 & \cdots & 0\end{pmatrix}^T, \quad (B8)$$

and those with respect to the elements of the last two columns can have non-zero values only in the last two elements:

$$\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k,2q}}=\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k-1,2q-1}}=\begin{pmatrix}0 & \cdots & 0 & Re(B_{r,i,(k)}) & Im(B_{r,i,(k)})\end{pmatrix}^T, \quad (B9)$$

$$\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k,2q-1}}=\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial a_{2k-1,2q}}=\begin{pmatrix}0 & \cdots & 0 & Im(B_{r,i,(k)}) & -Re(B_{r,i,(k)})\end{pmatrix}^T. \quad (B10)$$

Using these relationships, the partial derivatives of $\boldsymbol{A}^T\boldsymbol{x}_i$, with respect to the unknown vector $\boldsymbol{t}$, can be calculated as

$$\frac{\partial\boldsymbol{A}^T\boldsymbol{x}_i}{\partial\boldsymbol{t}^T}=\boldsymbol{I}_{q\times q}\otimes\begin{pmatrix}\mathcal{R}(\boldsymbol{b}_{r,i}) & \mathcal{I}(\boldsymbol{b}_{r,i})\\-\mathcal{I}(\boldsymbol{b}_{r,i}) & \mathcal{R}(\boldsymbol{b}_{r,i})\end{pmatrix}^T\in\mathbb{R}^{2q\times 4q}, \quad (B11)$$

where $\boldsymbol{I}_{q\times q}$ is the $q\times q$ identity matrix, $\otimes$ denotes the operator for the Kronecker product of two matrices (Harville 1998), and $\mathcal{R}$ and $\mathcal{I}$ denote the operators extracting the real and imaginary components, respectively, from a complex matrix or vector. Therefore, eq. (B2)

can be transformed into

$$\frac{\partial d_i}{\partial\boldsymbol{t}^T}=-\frac{1}{d_i}vec\left[\begin{pmatrix}\mathcal{R}(\boldsymbol{b}_{r,i}) & \mathcal{I}(\boldsymbol{b}_{r,i})\\-\mathcal{I}(\boldsymbol{b}_{r,i}) & \mathcal{R}(\boldsymbol{b}_{r,i})\end{pmatrix}\begin{pmatrix}(\boldsymbol{C}^{-1}\boldsymbol{r}_i)_{odd}^T\\(\boldsymbol{C}^{-1}\boldsymbol{r}_i)_{even}^T\end{pmatrix}\right]^T, \quad (B12)$$

where $(\boldsymbol{C}^{-1}\boldsymbol{r}_i)_{odd}^T$ and $(\boldsymbol{C}^{-1}\boldsymbol{r}_i)_{even}^T$ are the transposes of the vectors consisting of only the odd- and even-numbered elements of $\boldsymbol{C}^{-1}\boldsymbol{r}_i$, respectively. By transforming eq. (B12) using the assumption of the form of the noise covariance (eq. 23), the partial derivatives of the Mahalanobis distance can be obtained, with respect to the unknown vector $\boldsymbol{t}$:

$$\frac{\partial d_i}{\partial\boldsymbol{t}^T}=-\frac{1}{d_i}vec\left(\begin{matrix}\mathcal{R}\left[\boldsymbol{b}_{r,i}^*\left(\tilde{\boldsymbol{C}}^{-1}\boldsymbol{\varepsilon}_i\right)^T\right]\\\mathcal{I}\left[\boldsymbol{b}_{r,i}^*\left(\tilde{\boldsymbol{C}}^{-1}\boldsymbol{\varepsilon}_i\right)^T\right]\end{matrix}\right)^T. \quad (B13)$$

From eqs (25), (B1) and (B13), the identity can be obtained.

$$\sum_{i=1}^{n_d}\frac{\rho'(d_i)}{d_i}\left(\begin{matrix}\mathcal{R}\left[\boldsymbol{b}_{r,i}^*\left(\tilde{\boldsymbol{C}}^{-1}\boldsymbol{\varepsilon}_i\right)^T\right]\\\mathcal{I}\left[\boldsymbol{b}_{r,i}^*\left(\tilde{\boldsymbol{C}}^{-1}\boldsymbol{\varepsilon}_i\right)^T\right]\end{matrix}\right)=0\in\mathbb{R}^{4\times q}, \quad (B14)$$

which can be transformed into

$$\sum_{i=1}^{n_d}\frac{\rho'(d_i)}{d_i}\tilde{\boldsymbol{C}}^{-1}\left(\boldsymbol{v}_i-\hat{\boldsymbol{T}}\boldsymbol{b}_{r,i}\right)\boldsymbol{b}_{r,i}^H=0\in\mathbb{R}^{q\times 2}. \quad (B15)$$

By rearranging the equation after the left-multiplying of $\tilde{\boldsymbol{C}}$, an equation can be obtained to compute the TF:

$$\hat{\boldsymbol{T}}=\left(\sum_{i=1}^{n_d}w_i\boldsymbol{v}_i\boldsymbol{b}_{r,i}^H\right)\left(\sum_{i=1}^{n_d}w_i\boldsymbol{b}_{r,i}\boldsymbol{b}_{r,i}^H\right)^{-1}\in\mathbb{C}^{q\times 2}, \quad (B16)$$

$$w_i=\frac{\rho'(d_i)}{d_i}. \quad (B17)$$

Next, the authors transformed eq. (26). The diagonal elements on the left side of eq. (26) can be rearranged:

$$\frac{\partial ln(|\boldsymbol{C}|)}{\partial\sigma_k^2}-\frac{\lambda}{n_d}\sum_{i=1}^{n_d}\rho'(d_i)\frac{\partial d_i}{\partial\sigma_k^2}\quad(k=1,\ldots,q). \quad (B18)$$

The first term of eq. (B18) can be transformed into

$$tr\left(\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{C}}{\partial\sigma_k^2}\right)=\frac{2}{\sigma_k^2}. \quad (B19)$$

The partial derivatives of the Mahalanobis distance, with respect to the diagonal elements of $\tilde{\boldsymbol{C}}$, can be rearranged to

$$\frac{\partial d_i}{\partial\sigma_k^2}=-\frac{1}{2d_i}\frac{|\varepsilon_{i,(k)}|^2}{\sigma_k^4}\quad(k=1,\ldots,q). \quad (B20)$$

where $\varepsilon_{i,(k)}\in\mathbb{C}$ is the kth element of the residual vector for the ith data segment ($\boldsymbol{\varepsilon}_i$). By substituting eq. (B20), the second term in eq. (B18) can be transformed into

$$\frac{\lambda}{n_d}\sum_{i=1}^{n_d}\frac{\rho'(d_i)}{2d_i}\frac{|\varepsilon_{i,(k)}|^2}{\sigma_k^4}. \quad (B21)$$

Thus, each diagonal element on the left-hand side of eq. (26) can be expressed as

$$\frac{2}{\sigma_k^2}+\frac{\lambda}{n_d}\sum_{i=1}^{n_d}\frac{\rho'(d_i)}{2d_i}\frac{|\varepsilon_{i,(k)}|^2}{\sigma_k^4}\quad(k=1,\ldots,q), \quad (B22)$$

and, as a result, eq. (26) can be transformed into

$$2\tilde{C}^{-1} + \frac{\lambda}{n_d} \sum_{i=1}^{n_d} \frac{\rho'(d_i)}{2d_i} \tilde{C}^{-1} \mathcal{D}\left[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i{}^H\right] \tilde{C}^{-1} = 0, \tag{B23}$$

where $\mathcal{D}: \mathbb{C}^{q \times q} \to \mathbb{C}^{q \times q}$ denotes the operator that extracts the diagonal elements from an input matrix. By calculating traces of both sides of eq. (B23) after multiplying $\tilde{C}^{1/2}$ from the left- and right-hand sides, the following can be obtained:

$$2q + \frac{\lambda}{n_d} tr\left(\sum_{i=1}^{n_d} \frac{\rho'(d_i)}{2d_i} \tilde{C}^{-1/2} \mathcal{D}\left[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i{}^H\right] \tilde{C}^{-1/2}\right) = 0. \tag{B24}$$

The trace in the second term can be rearranged to

$$\sum_{i=1}^{n_d} \frac{\rho'(d_i)}{2d_i} \boldsymbol{\varepsilon}_i{}^H \tilde{C}^{-1} \boldsymbol{\varepsilon}_i = \sum_{i=1}^{n_d} \frac{\rho'(d_i)}{2d_i} \boldsymbol{r}_i{}^T \boldsymbol{C}^{-1} \boldsymbol{r}_i = \sum_{i=1}^{n_d} \frac{\rho'(d_i) d_i}{2}. \tag{B25}$$

By substituting eq. (B25) into eq. (B24), the following equation for the Lagrange multiplier is obtained:

$$\lambda = -4q n_d \left(\sum_{i=1}^{n_d} \rho'(d_i) d_i\right)^{-1}. \tag{B26}$$

By substituting eq. (B26) into eq. (B23), followed by the multiplication of $\tilde{C}$ from both the left- and right-hand sides, the following is obtained:

$$\tilde{C} \sum_{i=1}^{n_d} \rho'(d_i) d_i - q \sum_{i=1}^{n_d} \frac{\rho'(d_i)}{d_i} \mathcal{D}\left[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i{}^H\right] = 0, \tag{B27}$$

which can be transformed into an equation for $\tilde{C}$:

$$\tilde{C} = \frac{q \sum_{i=1}^{n_d} \frac{\rho'(d_i)}{d_i} \mathcal{D}\left[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i{}^H\right]}{\sum_{i=1}^{n_d} \rho'(d_i) d_i}. \tag{B28}$$