


FULL PAPER

Open Access



Separation of magnetotelluric signals based on refined composite multiscale dispersion entropy and orthogonal matching pursuit

Xian Zhang^{1,2,3,4}, Jin Li^{2,3*} , Diqian Li^{1,4*}, Yong Li³, Bei Liu⁵ and Yanfang Hu^{1,4}

Abstract

Magnetotelluric (MT) data processing can increase the reliability of measured data. Traditional MT data denoising methods are usually applied to entire MT time-series, which results in the loss of useful MT signals and a decrease of imaging accuracy of electromagnetic inversion. However, targeted MT noise separation can retain part of the signal unaffected by strong noise and enhance the quality of MT responses. Thus, we propose a novel method for MT noise separation that uses the refined composite multiscale dispersion entropy (RCMDE) and the orthogonal matching pursuit (OMP) algorithm. First, the RCMDE is extracted from each segment of the MT data. Then, the RCMDEs for each segment are input to the fuzzy c-mean (FCM) clustering algorithm for automatic identification of the MT signal and noise. Next, the OMP method is utilized to remove the identified noise segments independently. Finally, the reconstructed signal consists of the denoised signal segments and the identified useful signal segments. We conducted simulation experiments and algorithm evaluations on electromagnetic transfer function (EMTF) data, simulated data and measured sites. The results indicate that the RCMDE can improve the stability of multiscale dispersion entropy (MDE) and multiscale entropy (ME) by analyzing the characteristics of the signal samples library, effectively distinguishing MT signals and noise. Compared with the existing technique of denoising entire time series, the proposed method uses the RCMDE as characteristic parameter and uses the OMP algorithm for noise separation, simplifies the multi-feature fusion, and improves the accuracy of signal-noise identification. Moreover, the denoising efficiency is accelerated, and the MT response in the low-frequency band is greatly improved.

Keywords: Magnetotelluric (MT), Refined composite multiscale dispersion entropy (RCMDE), Orthogonal matching pursuit (OMP), Noise separation

Introduction

Magnetotelluric (MT) sounding is one of the most mature electrical exploration techniques in recent years (Tikhonov 1950; Cagniard 1953) and is mainly used in geoelectrical structure exploration to measure the

orthogonal electric and magnetic field at the Earth's surface, mineral electrical exploration, and electromagnetic fracture monitoring (Becher and Sharpe 1969; Vallianatos 1996). Due to the wide frequency range of natural MT signals, artificial electromagnetic noise easily interferes with these signals. Thus, effectively suppressing noise can improve the signal-to-noise ratio (SNR) and ensure the quality of the MT response. However, the MT signal is a nonlinear, nonstationary and nonminimum phase and does not conform to the Fourier transform conditions (Hermance 1973). Therefore, strong electromagnetic noise will cause the excessive distortion of the apparent resistivity-phase curve and excessive concentration of the

*Correspondence: geologylj@163.com; lidiqian@csu.edu.cn

¹ Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment, Monitoring Ministry of Education, School of Geoscience and Info-Physics, Central South University, Changsha 410083, China

² Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China
Full list of author information is available at the end of the article

phase angle in polarization direction. For this reason, we hope to obtain a high-quality MT response under strong electromagnetic interference, which can provide technical support for subsequent inversion interpretation (Qi et al. 2020; Li et al. 2020a).

MT data processing methods, such as the remote reference (RR) method (Gamble et al. 1978; 1979) and the robust impedance estimation method (Egbert and Booker 1986; Jones et al. 1989), have been widely applied. Ritter et al. (1998) used indicators such as the transfer function between the magnetic field at the measured sites and reference site to judge the noise of each data segment and remove the noisy data segment, which did not participate in the next impedance estimation. Varentsov (2006) proposed the "RR magnetic field control" criterion, which uses the magnetic field transfer function to control the impedance estimation values. Although the RR method can eliminate the related noise, it relies on the selection of the reference sites. The robust impedance estimation method uses the measured field values and theoretical values to estimate the impedance, reducing the weight of the flight points and aligning the measured value with the estimated value, thereby achieving a better impedance estimate of the effect (Chave and Thomson 1989; 2004). Although the robust impedance estimation method can effectively reduce the dispersion of the apparent resistivity-phase curve and eliminate non-Gaussian noise in the MT data, the robust method is incapable of removing noise caused by the input and cannot eliminate the near-source interference with strong energy.

More novel MT signal processing methods have been applied to MT noise suppression. For example, the wavelet transform, which relies on the selection of the wavelet basis function, can effectively suppress local electromagnetic-related noise (Trad and Travassos 2000). However, with increasing scale, the spectral localization of the corresponding orthogonal basis function deteriorates, limiting the fine decomposition of MT data. The Hilbert–Huang transform (Huang et al. 1998) has been applied to electrical method exploration and can effectively suppress MT data with power frequency interference (Cai et al. 2009). The choice of the basis function has a stronger time–frequency characterization capabilities than the wavelet transform. Mathematical morphological filtering can effectively suppress large-scale interference and the baseline offset in the MT data and maintain the local characteristics of the target signal (Tang et al. 2012), but it is difficult to select the types and sizes of structural elements. Wang et al. (2017) treated the electric and magnetic field of time series independently; their proposed method replaces windows of noisy time series. In addition, they proposed a synthesis time-series method based on the interstation transfer function, which eliminates

the influence of anthropogenic noise. Variational mode decomposition (VMD) is a novel mode decomposition algorithm that has been applied for MT noise suppression. The K value, which is the number of modes in VMD, can only be manually selected. Li et al. (2020b) combined VMD with detrended fluctuation analysis (DFA) to adaptively select the K value and improved the denoising effect. The statistical analysis and time-series editing methods can directly and effectively improve the quality of MT data, but they will also destroy the effective signal contained in the noisy segments.

To the best of our knowledge, entropy was first introduced by Clausius when he studied the efficiency of the Carnot cycle in thermodynamics. Then, the idea of entropy is related to the degree of disorder in statistical physics and information theory. Entropy, such as the sample entropy (Richman et al. 2004), fuzzy entropy (Kosko 1986) and approximate entropy (Pincus 1991), is now used for feature extraction and can be used to assess the complexity of a system. On the basis of multiscale analysis, multiscale entropy (ME) is proposed to quantify the complexity of signals on multiple scales. Multiscale dispersion entropy (MDE) is a parameter to evaluate the multiscale dynamic complexity of time series (Zhang et al. 2018). The refined composite MDE (RCMDE) can increase the accuracy of entropy estimation and decrease the probability of facing situations where the entropy is undefined (Azami et al. 2017). The RCMDE can reduce the sensitivity of the ME and MDE to the signal length for the time series, which can combine the information of multiple coarse-grained sequences, reduce the standard deviation of entropy and improve the stability of numerical results. The Fuzzy c-mean (FCM) clustering algorithm is an unsupervised method for data analysis and modeling and is widely used in data classification and pattern recognition (Xu et al. 2019; Zhang et al. 2019). The input features are used to generate the clustering center, the Euclidean distance is calculated between the clustering points and the clustering centers, and the membership degree of the clustering centers is obtained to divide the types of input features automatically.

Sparse representation uses as few atoms as possible to represent the signal in a given over-complete dictionary to obtain a more concise representation of the signal and acquire the contained information and process the signal more conveniently. Based on the matching pursuit (MP) algorithm (Mallat and Zhang 1993), the orthogonal MP (OMP) algorithm is a kind of classic greedy algorithm (Pati et al. 1993). An atom is defined as a time domain signal, which is used to construct the over-complete dictionary. Any signal can be represented by a sparse linear combination of atoms. By using the Gram–Schmidt orthogonalization method to rotate the selected optimal

atom and atomic set, where the selected atom is orthogonal to the residual in each iteration, thereby accelerating the convergence speed of the algorithm. In view of the fact that the signals usually contain stationary and non-stationary components, an atomic library, namely, an over-complete dictionary (Cai and Wang 2011; Needell and Vershynin 2010), composed of sine, cosine and wavelet atoms, is designed to realize the adaptive and accurate matching of signals.

In the paper, based on inherent signal-noise characteristics in the respective time series, the RCMDE and OMP algorithm are used to realize MT noise identification and separation, respectively. First, we verified the stability of the RCMDE and the simulation denoising effect of the OMP algorithm. Then, we carry out simulation experiments on electromagnetic transfer function (EMTF) data, that is, an open source code for single-site robust MT estimation and RR analysis, and measured MT sites. Compared with the RR method and the OMP-based overall method (that is, the OMP method that does not involve noise identification processing), the proposed method can purposefully remove the identified noise and retain the low-frequency useful MT response. The fractal-entropy and clustering method is also compared, namely that fractal box dimension, Higuchi fractal dimension, fuzzy entropy and approximate entropy are extracted from MT time-series, and the signals and noise are automatically distinguish by using FCM clustering, while the wavelet threshold denoising method merely suppresses the identified strong interference (Li et al. 2018). The proposed method uses only the RCMDE and OMP algorithm, which improves the identification accuracy and denoising effect. The experimental results of apparent resistivity-phase curves, polarization direction, coherence, error and SNR at the measured sites showed that the denoised MT data can approach to the true MT field, and the MT response more truly reflects the underground electrical structural information.

Methods

It is well known that the MT signal is very weak and is often affected by strong electromagnetic interference, resulting in a serious decrease in data quality and some abnormal waveforms in the time-series waveform. Thus, improving the data quality and removing abnormal waveforms will contribute to the availability of data. In this section, the RCMDE and OMP methods are described in detail. The RCMDE as a characteristic parameter is used to quantitatively identify the signal and noise, and the OMP algorithm as a denoising method is merely used to eliminate noise. Moreover, the RCMDE is compared with the ME and MDE in the feature extraction of sample library signals. The denoising performance of the OMP

algorithm is compared with that of the MP algorithm in the simulation noisy data.

Dispersion entropy (DE)

The dispersion entropy (DE), proposed by Rostaghi and Azami in 2016, is a nonlinear dynamics method to characterize the complexity and irregularity of time series (Rostaghi and Azami 2016; Mitiche et al. 2018):

- (1) Suppose time-series x_j , ($j = 1, 2, \dots, N$) is mapped to c classes with integer indices from 1 to c . To realize this purpose, the normal cumulative distribution function (NCDF) maps x to $y = \{y_1, y_2, \dots, y_N\}$ from 0 to 1 as follows:

$$y_j = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad (1)$$

where σ and μ are the standard deviation (SD) and mean of time-series x , respectively. Then, a linear algorithm is applied to an integer that ranges from 1 to c for each y_i . For each member of the mapped signal, we use $z_j^c = \text{round}(c \times y_j + 0.5)$, where z_j^c denotes the j th member of the classified time series and round involves rounding a number up or down to the next digit. Note that, although this part is linear, the entire mapping method is nonlinear due to the use of the NCDF.

- (2) Each embedding vector $z_i^{m,c}$ are made with embedding dimension m and time delay d , according to the construction of the time-series $z_i^{m,c} = \{z_i^c, z_{i+d}^c, \dots, z_{i+(m-1)d}^c\}$, $i = 1, 2, \dots, N - (m-1)d$ (Bandt and Pompe 2002; Rostaghi and Azami 2016), which is mapped to a dispersion pattern $\pi_{v_0 v_1 \dots v_{m-1}}$, where $z_i^c = v_0$, $z_{i+d}^c = v_1, \dots, z_{i+(m-1)d}^c = v_{m-1}$. The number of possible dispersion patterns that can be assigned to each time-series $z_i^{m,c}$ is equal to c^m , since the signal has m members and each member can be an integers from 1 to c (Rostaghi and Azami 2016).
- (3) For each of the c^m potential dispersion patterns, the relative frequency is obtained as follows:

$$p(\pi_{v_0 v_1 \dots v_{m-1}}) = \frac{\text{Number}\{i | i \leq N - (m-1)d, z_i^{m,c} \text{ has type } \pi_{v_0 v_1 \dots v_{m-1}}\}}{N - (m-1)d}, \quad (2)$$

where $p(\pi_{v_0 v_1 \dots v_{m-1}})$ denotes the number of dispersion patterns $\pi_{v_0 v_1 \dots v_{m-1}}$ that are assigned to $z_i^{m,c}$, divided by the total number of embedded signals with embedding dimension m .

- (4) The DE value is derived from the definition of Shannon's entropy and is defined as follows:

$$DE(x, m, c, d) = - \sum_{\pi=1}^{c^m} p(\pi_{v_0 v_1 \dots v_{m-1}}) \cdot \ln(p(\pi_{v_0 v_1 \dots v_{m-1}})). \quad (3)$$

Refined composite multiscale dispersion entropy (RCMDE)

The MDE is the combination of the coarse-graining (Costa et al. 2005) with DE and then calculating the DE value of the coarse-graining sequence to obtain the DE at different scale. Instead, the mapping based on the NCDF used in the calculation of DE for the first temporal scale is maintained across all scales. The RCMDE is an improved MDE as follows.

For a scale factor τ , which has a different time series corresponding to different starting points of the coarse-grained process are created, the RCMDE value is defined as the Shannon entropy value of the average of the dispersion patterns of those shifted sequences (Azami et al. 2017). The k th coarse-grained time-series $x_k^{(\tau)} = \{x_{k,1}^{(\tau)}, x_{k,2}^{(\tau)}, \dots\}$ of u is as follows:

$$x_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{b=k+\tau(j-1)}^{k+\tau j-1} u_b, 1 \leq j \leq N, 1 \leq k \leq \tau. \quad (4)$$

Then, for each scale factor, the RCMDE is defined as follows:

$$RCMDE(x, m, c, d, \tau) = - \sum_{\pi=1}^{c^m} \bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}) \cdot \ln(\bar{p}(\pi_{v_0 v_1 \dots v_{m-1}})), \quad (5)$$

where $\bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}) = \frac{1}{\tau} \sum_{k=1}^{\tau} p_k^{(\tau)}$ with the relative frequency of the dispersion pattern π in the time-series $x_k^{(\tau)} (1 \leq k \leq \tau)$.

For the multiscale analysis, according to the entropy differences in the sample library signal (Li et al. 2018), which include the 150 sets of actual MT time-series sequences, the data sample length is 240. Among these 150 sets, 50 sets of actual MT time-series without interference are from electromagnetic interference-free areas in Qinghai Province, China, and the remaining 100 sets of measured MT time-series sequences (50 sets of square wave interference and 50 sets of triangle wave interference) are collected from the strong electromagnetic interference areas in Anhui Province, China. Thus, we predefined the parameter values: the embedding dimension m is 2, the number of classes c is 6, and the time lag d is 1 (Rostaghi and Azami 2016). The time scale factor τ was used to analyze the coarsening of time series. When $\tau = 1$, the result of coarsened data is the original time series. When $\tau = 2$, the coarsened time series is formed by calculating the average value of two continuous time

points, and so on. Therefore, the scale factor will influence the number of characteristic parameter values.

Figure 1 shows the results of ME, MDE and RCMDE using a set of sample library signals with different scale factor.

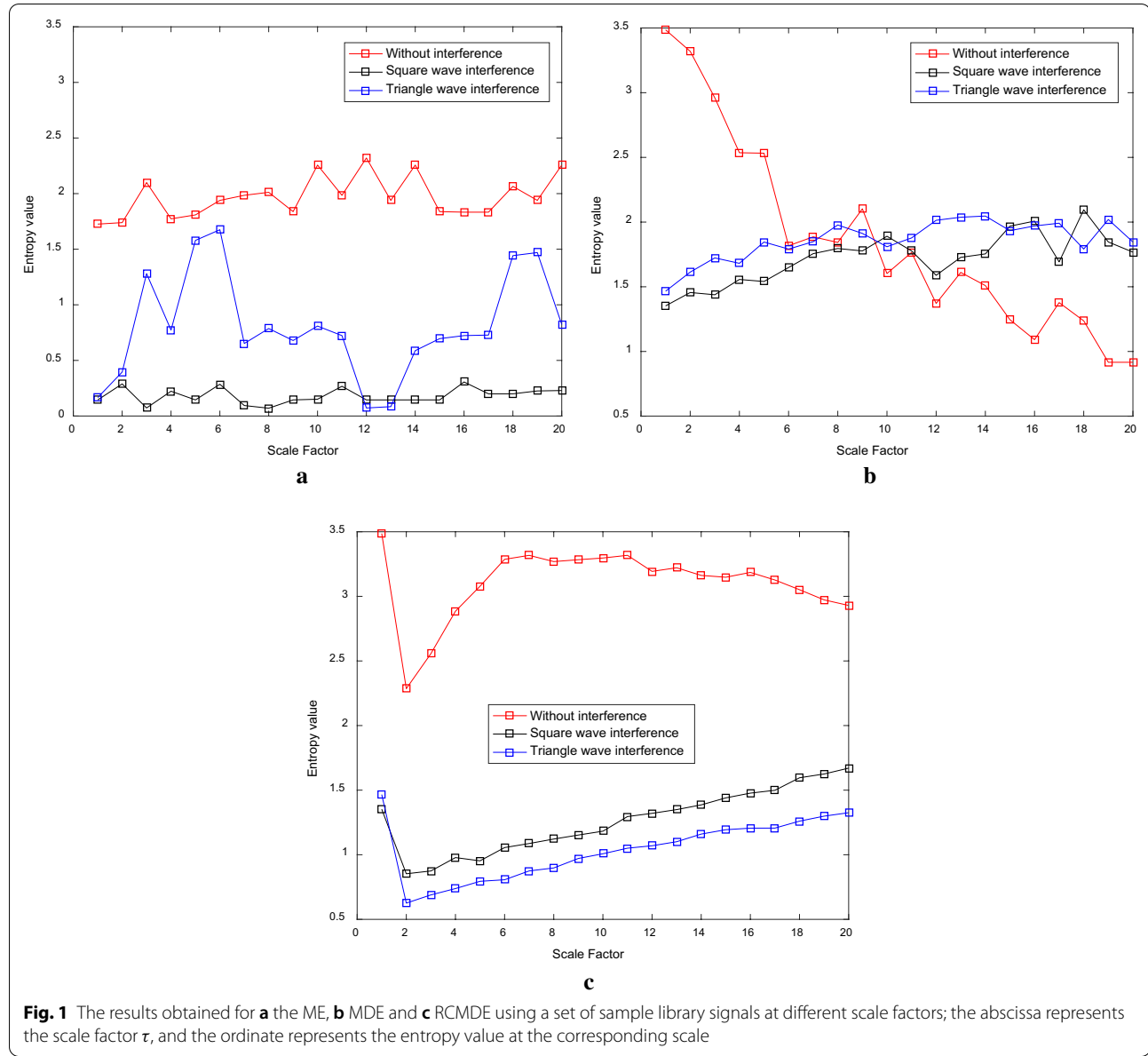
Figure 1 shows that a set of sample library signals have different entropy values at different scales. As the scale factor increases, the difference between the ME and MDE of a set of sample library signals decreases, and the crossover phenomenon and higher oscillations of the different interference curves are not conducive to classifying noise and signals. However, the RCMDE shows the importance of the refined composite technique to improve the stability of the results. Moreover, extracting appropriate characteristic parameters to describe the features of the MT signal is helpful to improve the FCM clustering effect. From Fig. 1(c), when the scale factor is 1, the RCMDE is the largest, and the single characteristic value will be meaningless, which cannot reflect the scale characteristic of the MT data. As the scale increases, these curves become more stable, and the scale factor will determine the number of characteristic parameter values. In the paper, the RCMDE can generate multiple feature parameter values to represent MT data at different scale factor, while many characteristic parameters will reduce the clustering accuracy and consume more time. Thus, the RCMDE can be used as appropriate feature vector analysis.

Denosing method

The MP algorithm at each iteration ensures only that the matching residual data are orthogonal to an atom, which is prone to local optimization and results in a low matching accuracy and a large amount of computation (Huang and Makur 2011; Jin et al. 2014). The OMP algorithm is based on the MP algorithm, ensuring full backward orthogonality between the matching residual and the selected waveforms at each iteration and ensuring the optimal approximation regarding all the selected subset of the dictionary after any finite number of iterations (Wang et al. 2013; Li et al. 2021).

Given a time-series signal x of length N . $D = (g_\gamma)_{\gamma \in \Gamma}$ is an over-complete dictionary, that is the Fourier atomic library and wavelet atomic library, which is used for sparse signal decomposition. g_γ is the γ th atom in the dictionary set Γ , and $\|g_\gamma\| = 1$.

Initialized the signal residual $R^0 = x$ and reconstructed signal $\bar{x}_n = 0$, we selected the atom set ψ_0 as the empty set, the number of iterations $n = 1$, and the maximum number of iterations as M . The following steps are repeated until the iteration stop condition is met (Tropp and Gilbert 2007):



- (1) Select the atom g_γ that most closely matches the analysis signal f from the dictionary. They meet the following conditions:

$$\left| \langle f, g_\gamma^n \rangle \right| = \sup_{\gamma \in \Gamma} \left| \langle f, g_\gamma^n \rangle \right|,$$

where $\left| \langle f, g_\gamma^n \rangle \right|$ is the inner product of f and g_γ^n , and $\sup_{\gamma \in \Gamma} \left| \langle f, g_\gamma^n \rangle \right|$ is the upper limit.

- (2) Update the selected atom set $\psi_n = \psi_{n-1} \cup \{g_\gamma^n\}$.
- (3) Find the projection coefficient according to the least squares method $u_n = (\psi_n^T \psi_n)^{-1} \cdot \psi_n^T x$. Thus, the reconstructed signal is $\bar{x}_n = \psi_n x_n$, and the residual signal is $R^n = x - \bar{x}_n$.
- (4) Update the number of iterations $n = n + 1$. Judge whether the energy ratio $\|R^n\|_2 / \|x\|_2$ of the residual signal to the original signal is less than the given value. If this condition is not satisfied, return to step (1). If this condition is satisfied, the reconstructed signal $x = \bar{x}_n$ is obtained, and the residual signal is $R = R^n$.

According to typical MT large-scale strong interference type and 20 dB white Gaussian noise simulated in MATLAB, we constructed triangle wave interference for analysis, and the length of the noisy signal was 2048. Figure 2 shows the denoising effects of the MP and OMP methods. Among them, the atomic library consists of sine (sin), discrete cosine transform (dct) atoms, symlets (sym) and Daubechies (db) wavelet atoms.

To estimate the denoising effects of the MP and OMP methods for the noisy signal, the normalized cross-correlation (NCC), SNR, mean square error (MSE) and the runtime were used for the quantitative analysis. For the definitions of these parameters, refer to Li et al. (2020b).

Table 1 shows the comparison between the denoising performance of the MP and OMP methods.

As seen from Fig. 2 and Table 1, comparing the denoising effects of the MP and OMP methods with the same number of atoms and iterations, we found that the MP method still has residual noise in the reconstructed signal, while the OMP method shows the excellent characteristics in the NCC, SNR, MSE and runtime. Since the OMP method ensures that the residue is orthogonal to all selected atoms, the convergence speed is faster than that of the MP method. Thus, the OMP method is more efficient and more suitable for MT noise separation.

Experiments

Algorithm steps

The algorithm steps of the proposed method are as follows:

Step 1: Input MT data, and divide the MT data at equal intervals of 240. Because the low-frequency data collected by the V5-2000 instrument are usually sampled at 24 Hz, we hope to make a judgment every 10 s.

Step 2: Extract the RCMDE with a scale factor of 2 for each segment of the MT data; that is, each segment data has two RCMDE values;

Step 3: Input all the RCMDE values into the FCM clustering algorithm, and automatically identify MT signal and noise; that is, one part of is the MT signal and the other part is noise;

Step 4: Retain the part that is marked as a signal, and use the OMP method to denoise the data identified as noise;

Step 5: Combine the denoised MT signal and the data segment identified as a signal to obtain a reconstructed signal.

Clustering analysis of the sample library

In this section, we applied FCM clustering analysis to the sample library signals by extracting the RCMDE. The FCM clustering algorithm obtains the membership

degree of each sample to all the clustering centers by optimizing the objective function, thereby determining the type of the samples to achieve automatic classification.

Figure 3 shows the FCM clustering effect of the sample library signals.

We calculate the Euclidean distance from each sample point to the cluster center and select the length of the farthest distance point as the radius of the pink circle, including all the sample points with the longest and shortest distance in the same type of sample. Specifically, the two pink circles effectively divide the sample signals into different types and accurately identify the MT signals with and without interference. According to verification with the FCM clustering effect, the MT signal without interference can be divided, which is represented by blue points. Subsequently, the identification and purposeful removal of interfered MT signal will be a critical way to improve reconstructed signals.

Numerical simulation analysis of the EMTF data

The numerical simulation analysis is based on the EMTF open source software (Eisel and Egbert 2001). The open source code package provides two $100 \Omega \cdot m$ uniform half-space time-series data (test1.asc and test2.asc). Each data point in the time series has five columns of signal with a length of 40,000 and the sampling rate is 1 Hz. The five columns of data represent the x, y, and z directions of the magnetic field and the x and y directions of the electric field for the observed data. The correlation between the signals of the two time series provided by the open source package is close to 1 (Egbert and Livelybrooks 1996). Furthermore, typical strong interference is added to the original signal of the EMTF data (test1.asc) to assess the performance of the OMP-based overall method and the proposed method.

Figure 4(a) and (b) shows the signal-noise separation effect of adding square wave interference to the original Hx-channel time-series of the EMTF data and adding triangle wave interference to the original Ey-channel time-series of the EMTF data, respectively. Figure 4(c) shows the comparison of the apparent resistivity-phase curves and error obtained by the RR method, the OMP-based overall method and the proposed method.

Considering that the Ey and Hx data affect the change in the ρ_{yx} curve, we add strong interference to the same position for the Ey and Hx data; that is, strong interference is added at the same time period. As seen from Fig. 4(a) and (b), the noise and signals are accurately identified, and the proposed method can extract the contour of the noise. Comparing the original data (Hx and Ey) with the reconstructed data (Hx and Ey) of the proposed method, the NCCs are 0.9524 and 0.9683, respectively, and the frequency spectra of the original data and

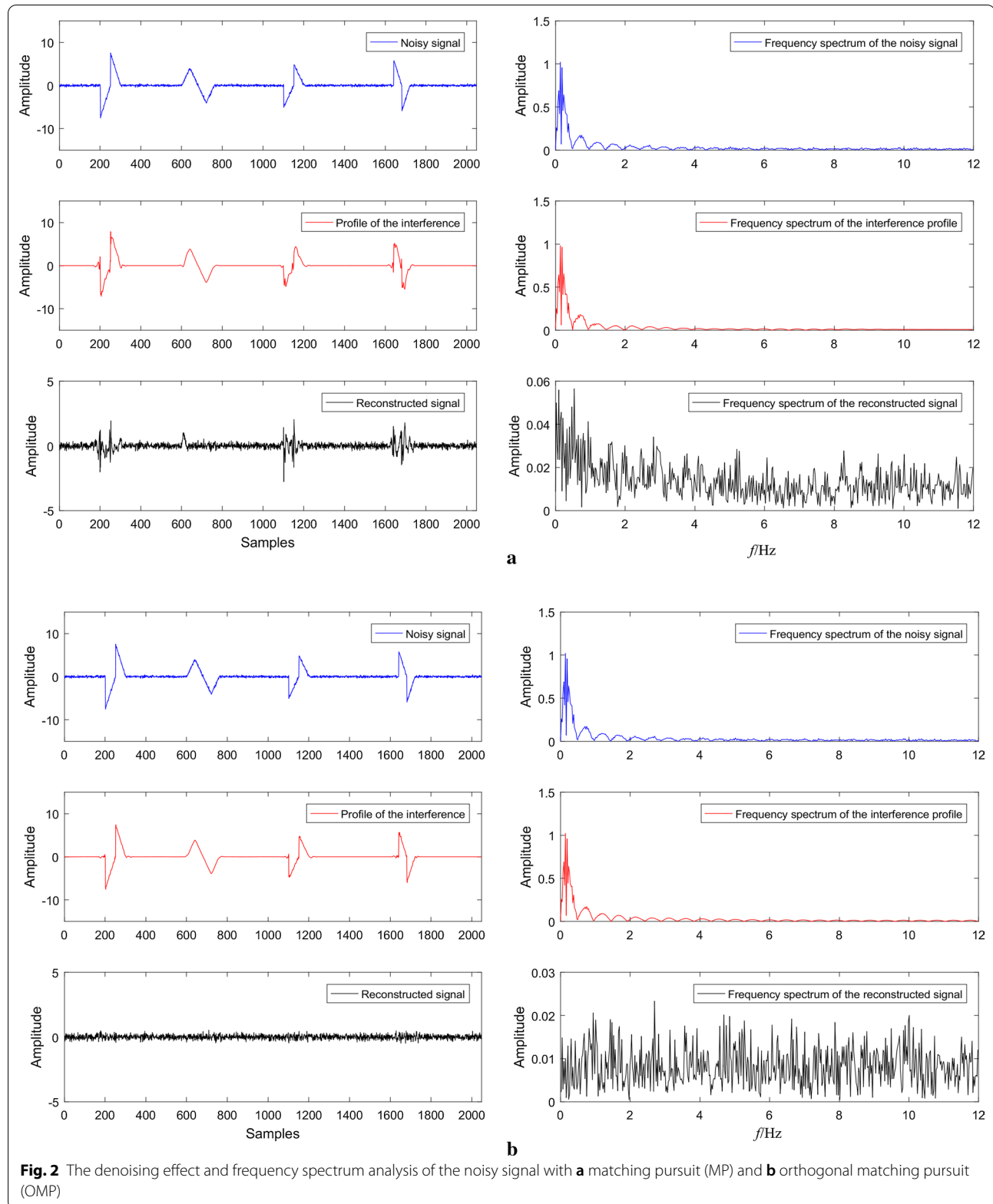
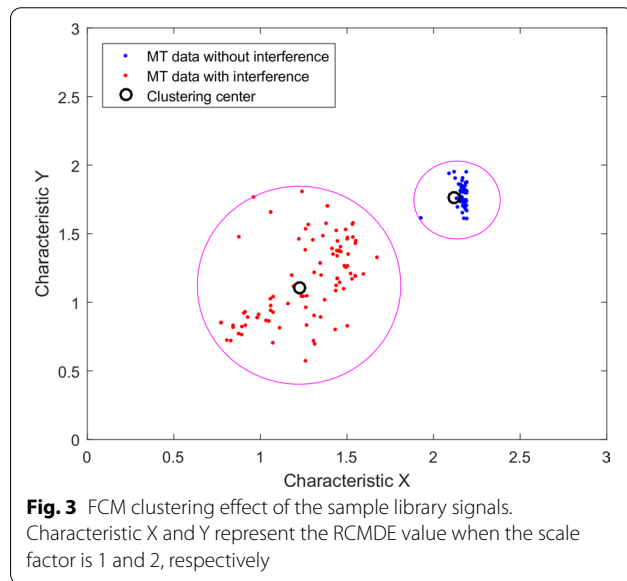


Table 1 The denoising performance of the MP and OMP methods

Methods	NCC	SNR (dB)	MSE	Runtime (s)
MP	0.9848	15.1976	0.0651	2.45
OMP	0.9989	26.3956	0.0049	0.59



reconstructed data are also similar. In contrast, the NCCs of the data reconstructed by the OMP method are 0.1162 and 0.4325, respectively, and the low-frequency useful signal is basically lost.

As shown in Fig. 4(c), due to the noise added by the Hx and Ey channels, ρ_{yx} and P_{yx} in curve 2 are greatly deformed. Curve 3 is obtained from the RR method; that is, test2.asc of the EMTF time-series data is used as reference data to suppress noisy data. Due to the high noise energy, which is added to the relevant noise data, there are some frequency jumps in the low-frequency part of ρ_{yx} , which does not yield the ideal effect. The apparent resistivity-phase curve 4 is still disordered, indicating that although the OMP-based overall method can suppress strong interference, the useful signal is lost due to the filtering of the entire time series. Analyzing the apparent resistivity-phase and error curve obtained by the proposed method, it can be seen that the smoothness of curve 5 is basically similar to the shape of curve 1 obtained by the original noise-free data. This experiment shows that the proposed method has a significant denoising effect on the above-mentioned tested EMTF noise-added data.

Noise separation of the measured MT data

To verify the effectiveness of the proposed method, the measured MT signal with typical strong interferences is used for analysis. The measured MT signal is collected by the V5-2000 instrument from the Luzong ore concentration area in Anhui Province, China.

Figure 5 shows the signal-noise identification and targeted denoising effect by comparing with the OMP-based overall method and the proposed method on the MT data with square and triangle wave interference.

As shown in Fig. 5, although the OMP-based overall method eliminates the large-scale strong interference, there is residual noise in the signal part with interference, and part of the useful signal is also filtered. The proposed method can identify and merely eliminate the identified noise and reserve the useful MT signal, thereby avoiding underprocessing and overprocessing with the OMP-based overall method and improving the reliability of the reconstructed signal.

Results

Apparent resistivity-phase curve analysis of the measured sites

In this section, we compare the apparent resistivity-phase curves of the original data, the RR method, the OMP-based overall method, the fractal-entropy and clustering method (Li et al. 2018) and the proposed method. The measured sites (D37890, EL22189 and EL22174) are collected from ore concentration area (Anhui Province, China) that are affected by square wave and triangle wave interference in the time series.

Figure 6 shows the comparison of the apparent resistivity-phase curves of the measured MT site D37890.

As shown in Fig. 6, the apparent resistivity-phase curve of the original data (curve 1) gradually increases to $10^4 \Omega \cdot m$ at frequencies of 10–0.03 Hz, and the corresponding phase curve is attached to 0° . While the apparent resistivity curve drops sharply at 0.03 Hz, the result shows a typical near-source effect. This is because the original data are affected by the large-scale strong interference and periodic interference, so the MT response of this site cannot objectively reflect the information of the underground electrical structure.

Curve 2 is the data filtered by the OMP-based overall method. The entire low-frequency band of the apparent resistivity curve decreases, and the phase curve is seriously disturbed. Although the large-scale strong interference is eliminated, the loss of low-frequency useful signals and the influence of residual noise still make this method unable to provide an effective response. Curve 3 is the fractal-entropy and

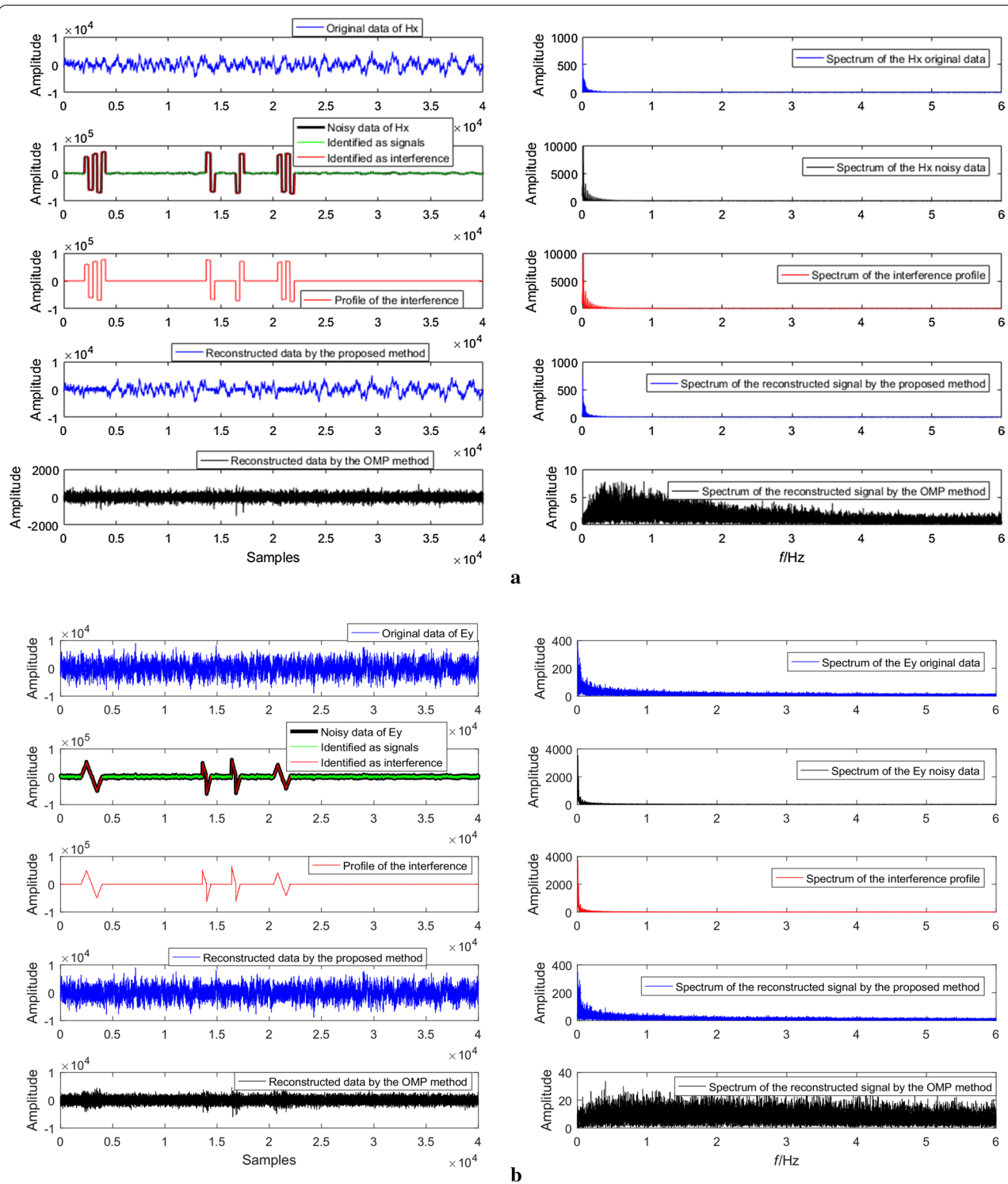
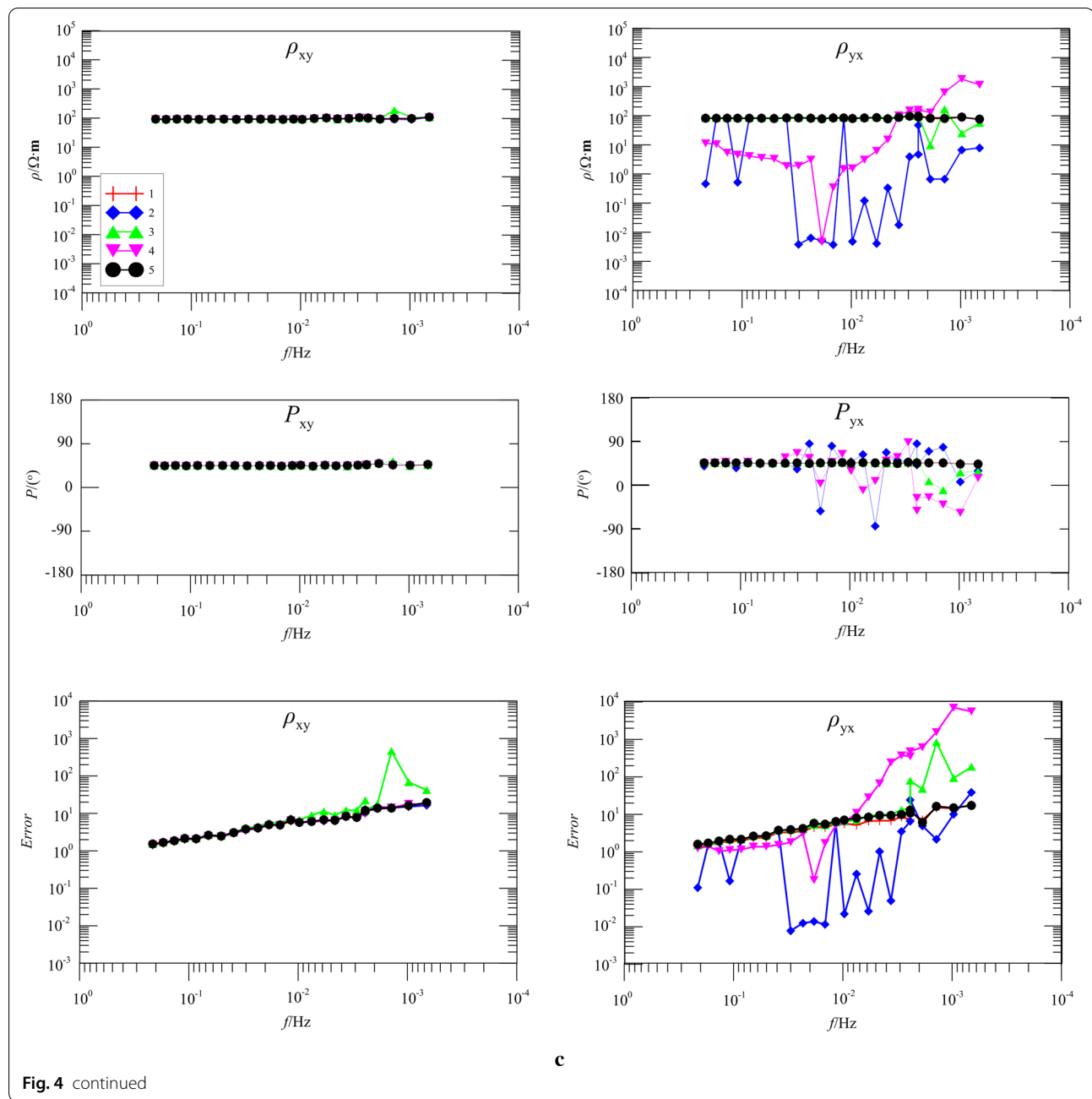


Fig. 4 The EMTF data are disturbed by a noisy signal with **a** square wave interference of the Hx and **b** charge and discharge triangle wave interference of the Ey; **c** is a comparison of the apparent resistivity-phase and error curves. Curve 1 is the original data, curve 2 is the noisy data, curve 3 is the RR method, curve 4 is the OMP-based overall method, and curve 5 is the proposed method



clustering method, which extracts four types of feature parameters for FCM clustering, and uses the wavelet threshold denoising method for targeted denoising. The result shows that near-source interference is suppressed, but ρ_{yx} of curve 3 at 3–0.03 Hz has not been alleviated. Considering that this method takes a long time to calculate the characteristic parameters, for the wavelet threshold denoising method, different wavelet bases and decomposition layers need to be defined in

advance, which leads to the robustness of the method. By analyzing curve 4 obtained by the proposed method, it can be seen that this method can purposefully suppress the identified MT interference segments and preserve the useful MT signal segments. As the apparent resistivity curve becomes more continuous, the phase curve also becomes more stable, and the MT response can reflect the underground electrical structure information more accurately and objectively.

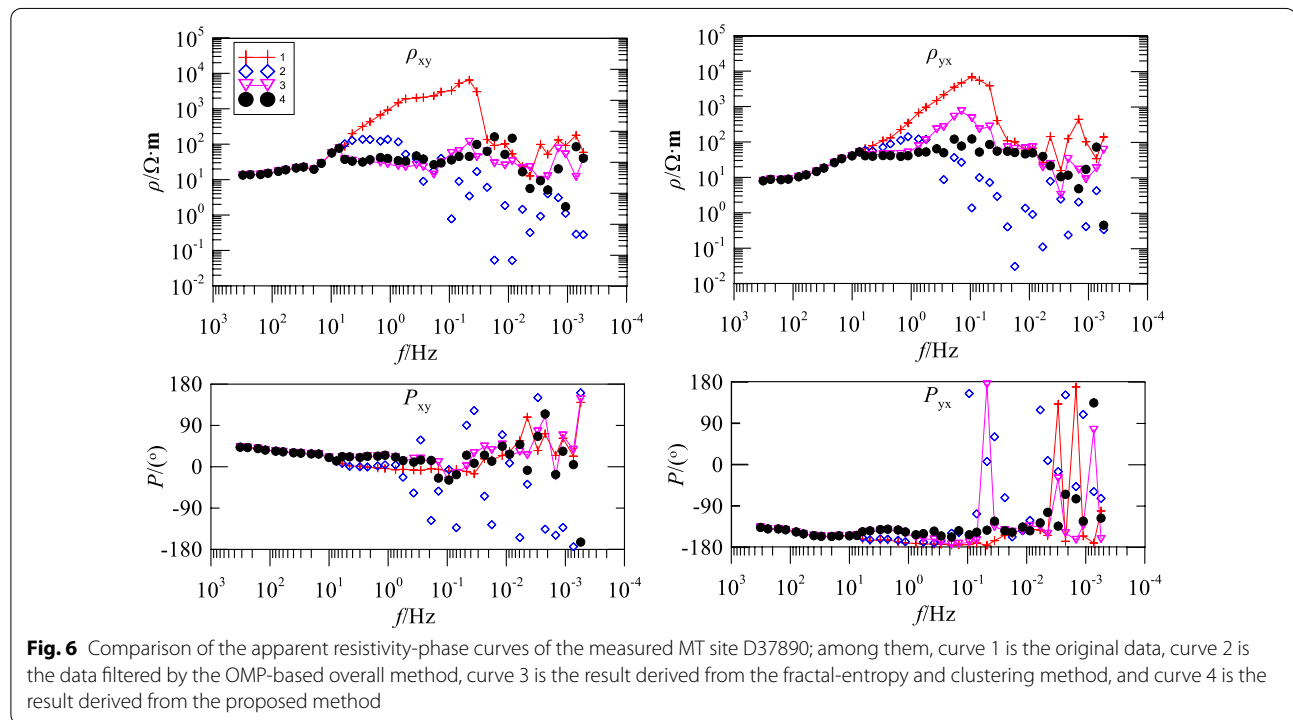
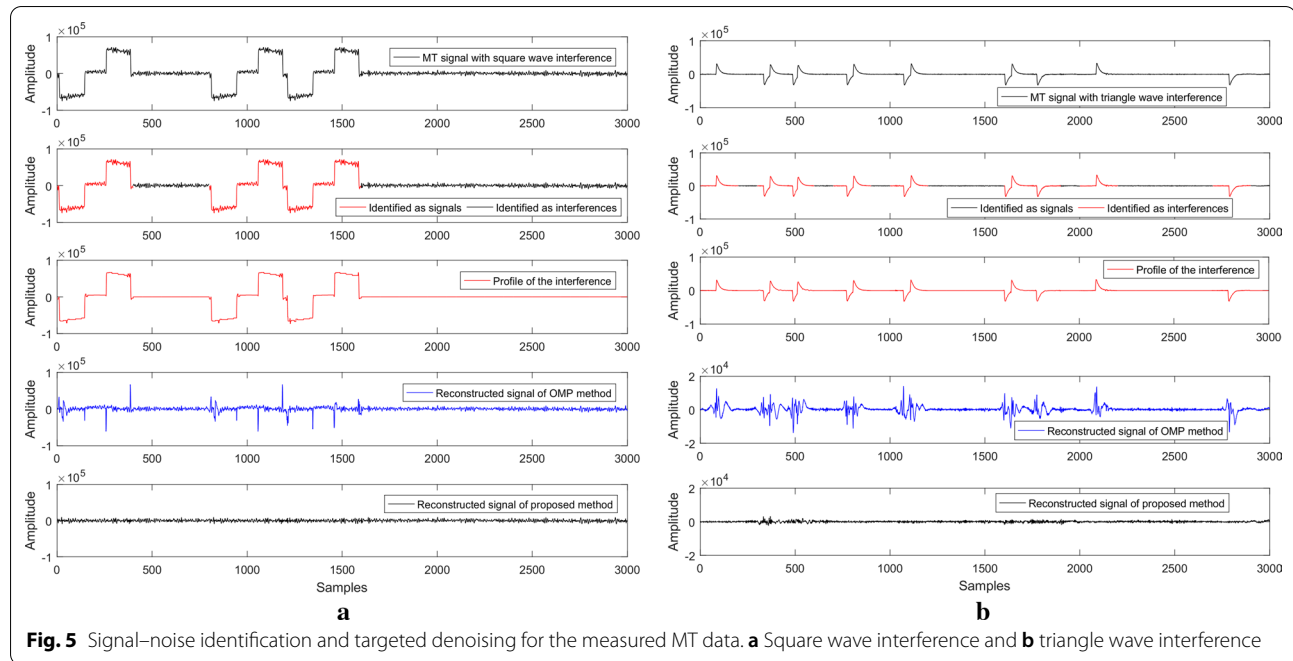


Figure 7 shows the comparison of the apparent resistivity-phase curves of the measured audio MT (AMT) sites (EL22189 and EL22174).

By analyzing curve 1 in Fig. 7(a) and (b), it can be seen that because the original measured site is subjected to strong electromagnetic noise, the obtained MT apparent resistivity-phase curve is seriously distorted, and its

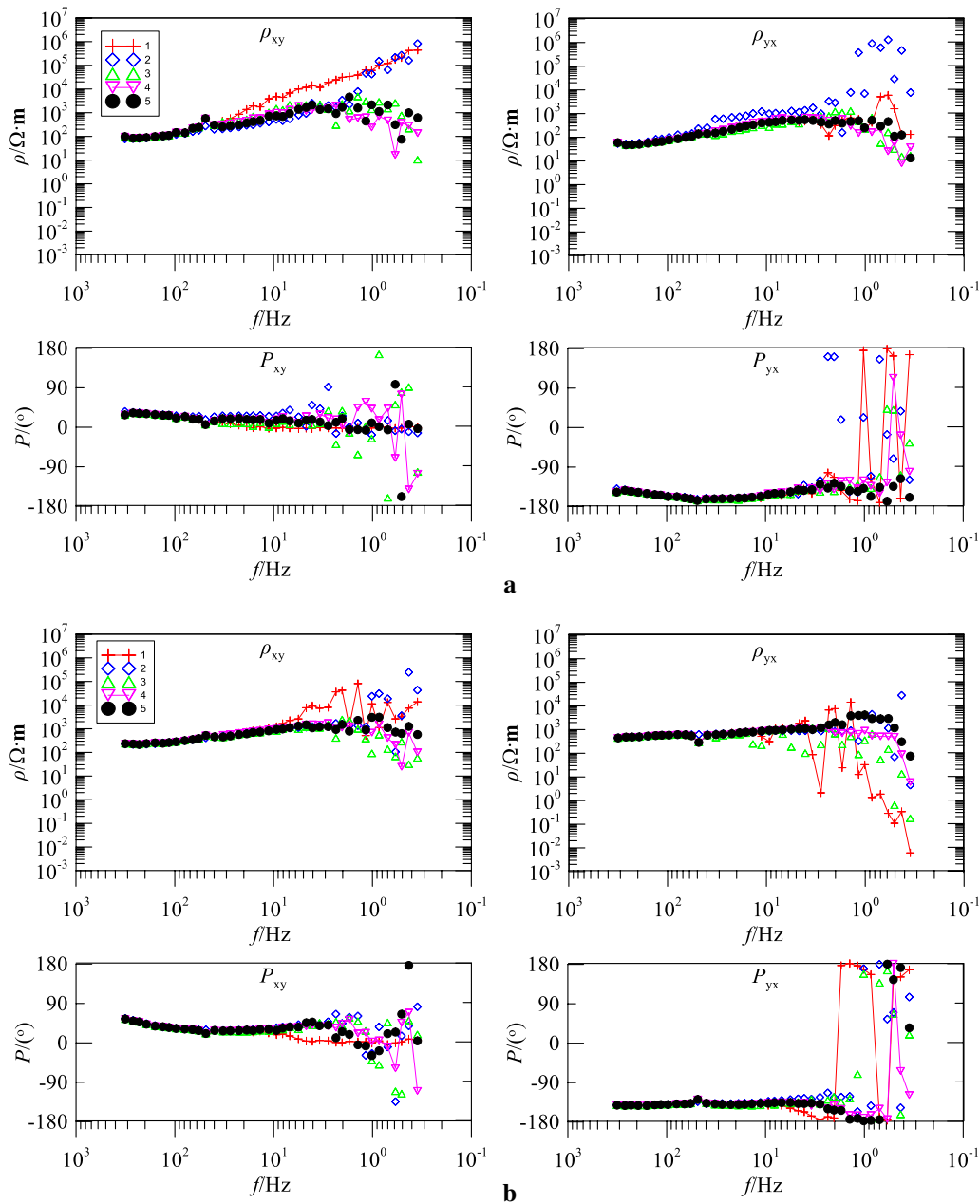


Fig. 7 Comparison of the apparent resistivity-phase curves of the measured AMT site **a** EL22189 and **b** EL22174. Among them, curve 1 is the original data, curve 2 is the result obtained by the RR method, curve 3 is the data filtered by the OMP-based overall method, curve 4 is the result derived from the fractal-entropy and clustering method, and curve 5 is the result derived from the proposed method

response cannot objectively represent the underground electrical structure information. Curve 2 shows that the effect of this method is not ideal. The low frequency still has a 45° increase and amplitude jump of 3–0.3 Hz because the RR method depends on the choice of reference site and the measured site distance; that is, a suitable reference site is difficult to select. The effect of curve 3 is not satisfactory. The OMP-based overall method is

used to process the entire time series, and the low-frequency useful information is lost, resulting in a serious drop in the low-frequency band of the resistivity curve, and the corresponding phase is disordered. Curve 4 has a good effect in the mid-frequency band, but the frequency point drops sharply in the low-frequency band. This is because weak noise may also be recognized as a useful signal when the fractal-entropy and clustering

method is used for identification, and the selection of wavelet denoising parameters is also limited. Compared with other methods, the effect of curve 5 is the best, the apparent resistivity-phase curve is more continuous and smoother, and the frequency point is less scattered.

Polarization direction analysis

To verify the effectiveness of the proposed method, the polarization direction of the electromagnetic field (Weckmann et al. 2005) is introduced to evaluate the quality of MT data. Figure 8 shows the comparison of the electromagnetic polarization direction (scatter plot and histogram) at 2 Hz and 4.2 Hz for the measured site EL22174 in the electric and magnetic fields, respectively.

Analyzing the scatter plot and histogram in Fig. 8, it can be seen that at 2 Hz, the electric field polarization directions of the original data are concentrated at 40° and 80°, and at 4.2 Hz, the magnetic field polarization directions are concentrated at -80° and 80°, which indicates that the original data are affected by strong electromagnetic noise. Although the OMP-based overall method has been improved, the effect is not good, and the polarization point still has scattering at some angles. The randomness of the polarization direction and polarization point obtained by the proposed method are more scattered, which is in line with the polarization characteristics of the natural field.

Summary

Since the theory of MT sounding was proposed, the problem of noise has plagued the majority of MT researchers. The large-scale strong interference that often appears in the time series causes the apparent resistivity-phase curve to be discontinuous, and the polarization direction is highly concentrated in a certain direction. Therefore, MT noise suppression is very important.

Based on the existing technology in MT noise suppression, reliable MT sounding data can be obtained through editing, filtering and identifying signals and noise in the time-frequency domain. These techniques can provide effective data for geological exploration and interpretation. In this study, Fig. 1 shows that the ME and MDE not only involve the process of coarse-graining time domain sequences, but also calculate the SE and DE values at different scales. They are mainly used to analyze time series with increasingly coarse temporal resolutions. For typical interference, the obtained ME values and MDE values are unstable, so the signal and noise cannot be distinguished accurately. Moreover, the RCMDE is an improvement over the MDE algorithm, using refined composite technology to obtain a better consistency, a more stable entropy value and a faster calculation speed. However,

we use only the RCMDE with a scale factor of 2 for FCM clustering to improve the accuracy and efficiency.

By observing a large amount of measured strong interference data, we found that the useful MT signal is very weak, and the energy of the noise is very strong. For this reason, Fig. 2 reasonably simulates the real MT noisy signal for analysis by adding large-scale strong interference and compares the effects of the MP and OMP algorithm. However, when the amplitude of the noise and the signal are very similar, the proposed method will have difficulty distinguishing between noise and the signal. Therefore, the proposed method mainly focuses on the identification of large-scale strong interference. In the MP algorithm, if the vertical projection of the residual signal on the selected atoms is nonorthogonal, it will trim the result of each iteration and require many iterations to converge. With the OMP algorithm, all the selected atoms are orthogonal in every step of the decomposition. With the same number of atoms and iterations as in Fig. 2, the runtime of the OMP algorithm is superior to that of the MP algorithm in signal-noise separation.

Next, we excavated the feature parameter (RCMDE) for FCM clustering analysis (Fig. 3), divided the MT signal and noise through its scale characteristics with high precision, and used the OMP algorithm to separate noise. To verify the feasibility of the proposed method, we designed experimental instructions, such as sample library signals (Fig. 1), simulated signals (Fig. 2), EMTF data (Fig. 4) and measured MT data (Fig. 5). As a numerical simulation analysis of EMTF data, we added large-scale strong interference to the known signal in the same time period of the Ey-channel and Hx-channel, resulting in a change in the ρ_{yx} curve (Fig. 4). By comparing with the RR method and the OMP-based overall method, the MT response obtained by the reconstructed signal is basically consistent with the response of the original undisturbed signal in the EMTF data.

Simple feature parameters and rapid denoising methods are further applied to the measured MT data, which can improve the effect and efficiency of the existing technology as shown in Fig. 5. By comparing the RR, OMP, signal-noise identification and separation methods with the apparent resistivity-phase curves. Figures 6 and 7 show that the proposed method effectively improves the multiple frequency point information in the low-frequency band, and the entire low-frequency curve becomes smoother and more stable. The polarization direction (Fig. 8) further illustrates that the result obtained by the proposed method is closer to the measured MT data than those obtained by the other methods.

Although the “natural” signals or “true” MT responses are unknown in the obtained data, we can be sure that abrupt and strong energy data in the collected time-series

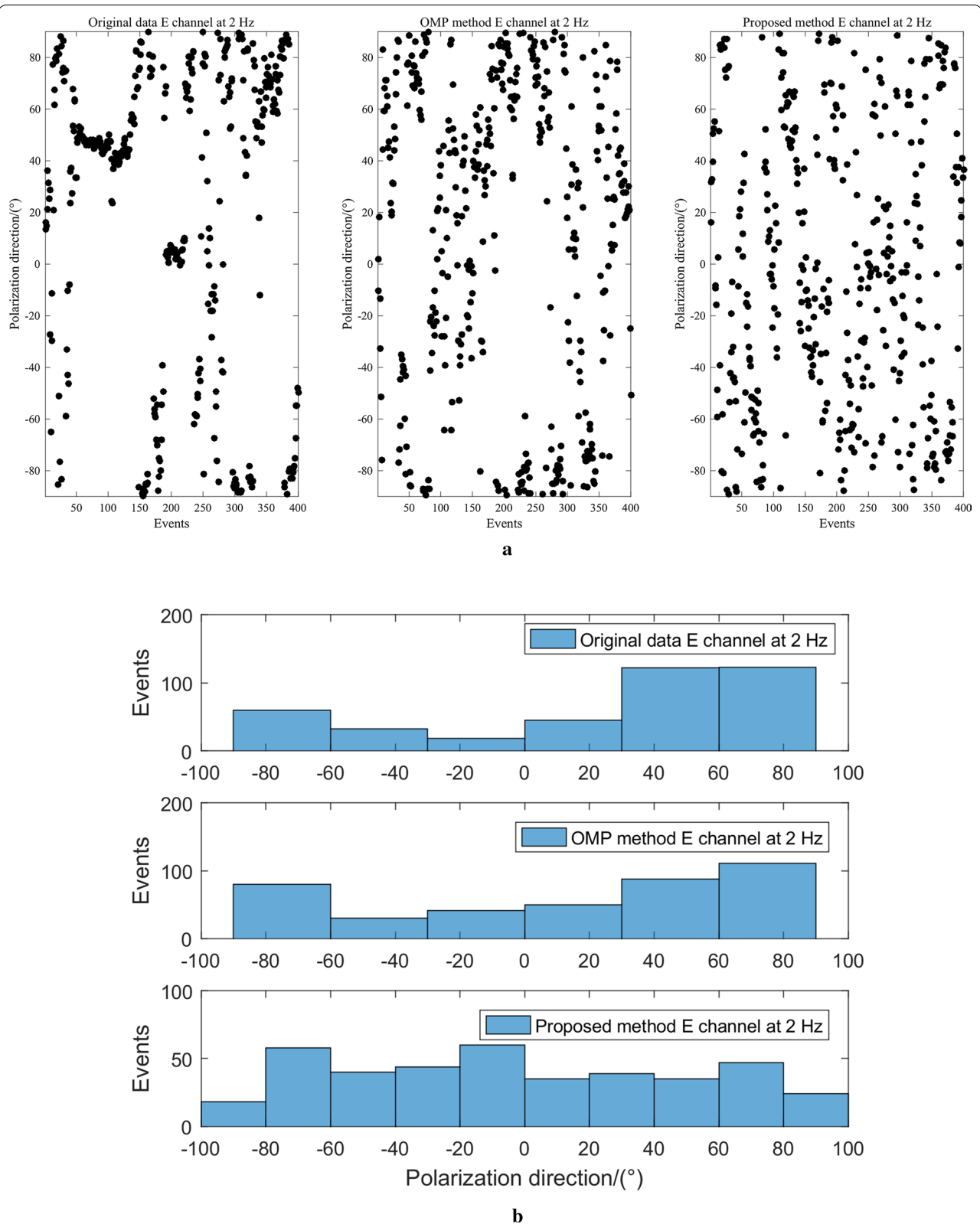


Fig. 8 Comparison of the polarization direction (scatter plot and histogram) for site EL22174: **a** scatter plot of the electric field data at 2 Hz; **b** histogram of the magnetic field data at 2 Hz; **c** scatter plot of the electric field data at 4.2 Hz, and **d** histogram of the magnetic field data at 4.2 Hz

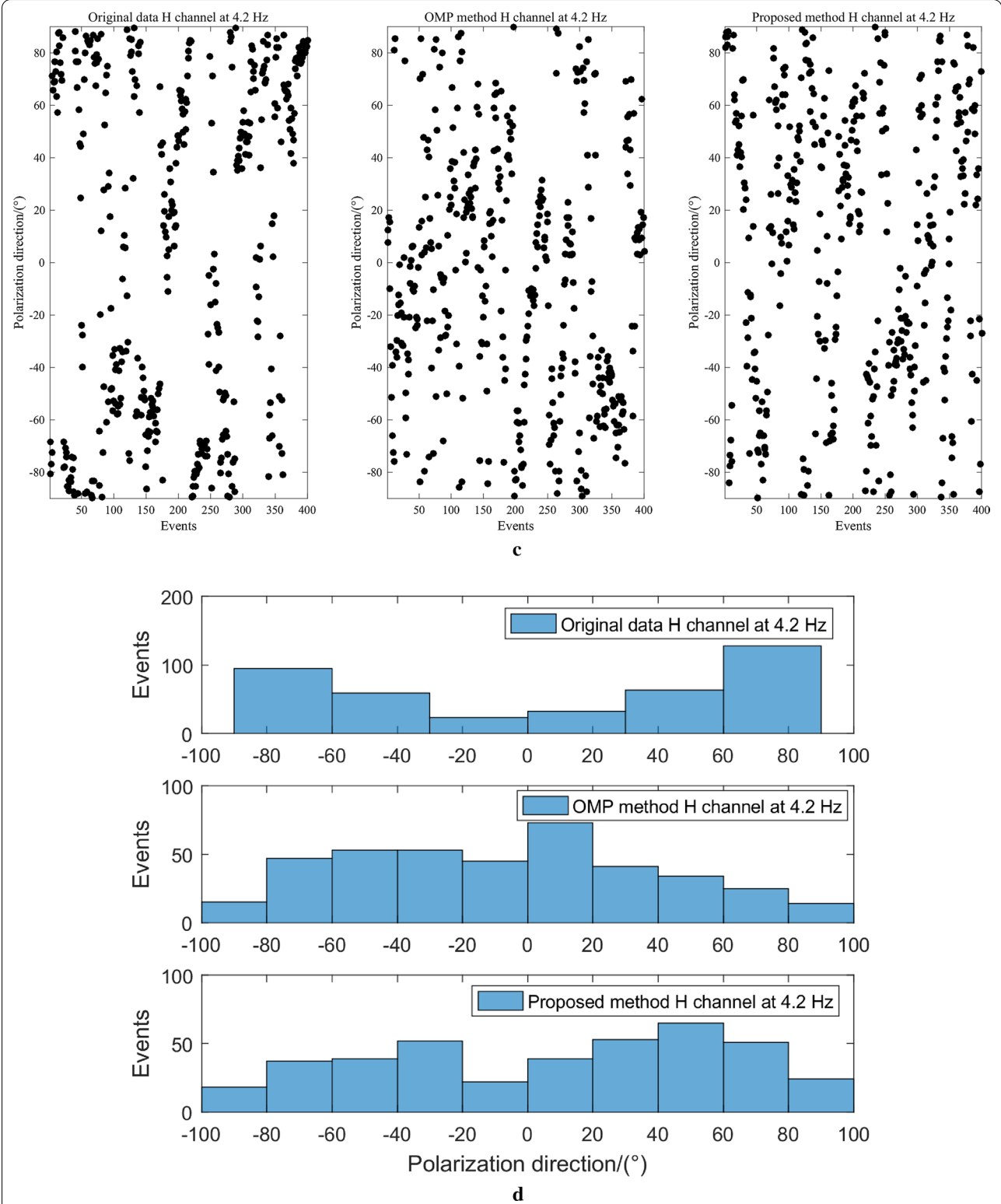


Fig. 8 continued

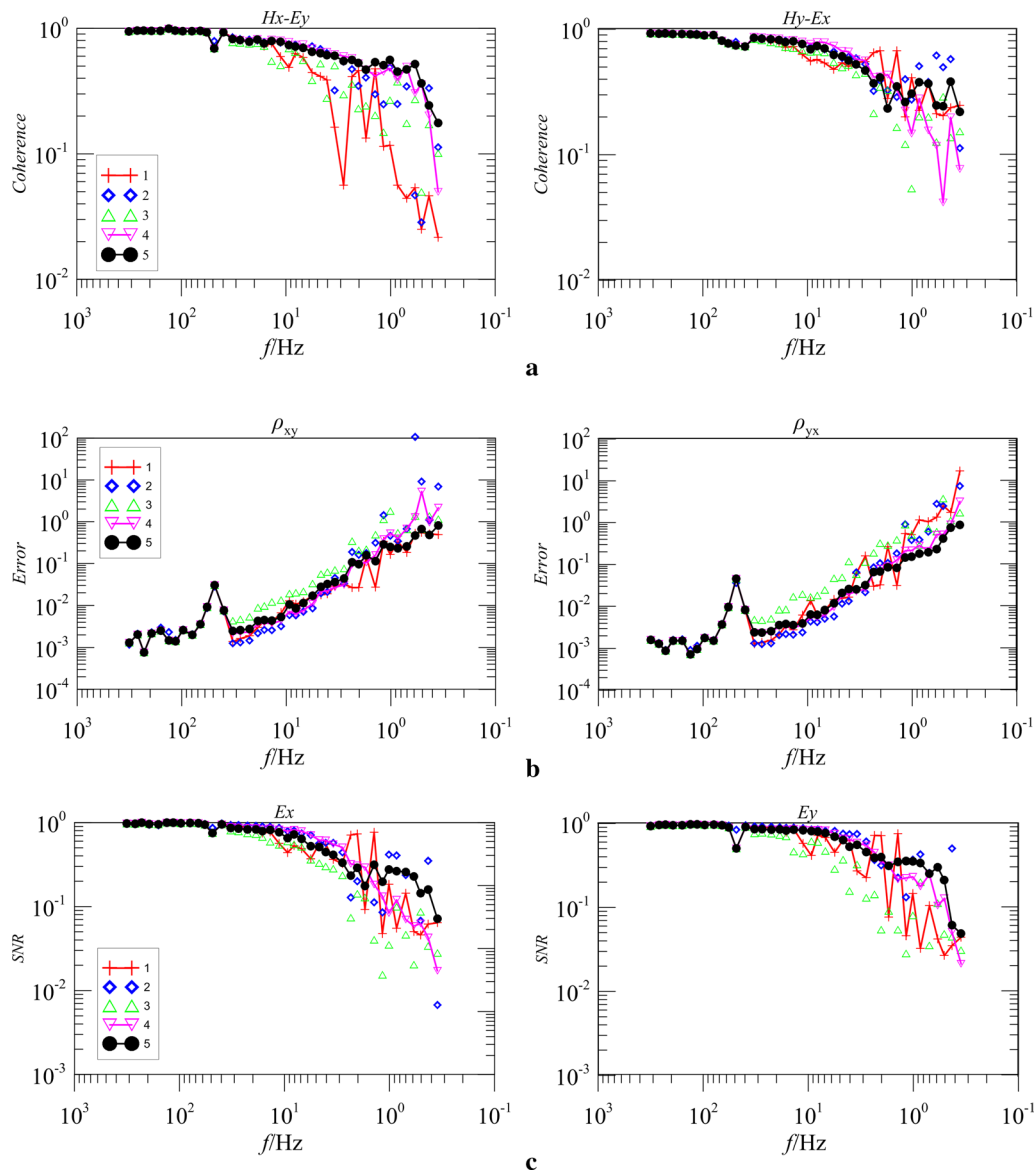


Fig. 9 Comparison of the coherence (a) error (b) and SNR (c) curves for site EL22174. Among them, curve 1 is the original data, curve 2 is the result obtained by the RR method, curve 3 is the data filtered by the OMP-based overall method, curve 4 is the result derived from the fractal-entropy and clustering method, and curve 5 is the result derived from the proposed method

data are definitely not the “natural” MT signals. To further illustrate the effect of the proposed method, the coherence, error and SNR are synthetically evaluated, as shown in Fig. 9 for site EL22174. The coherence is used to measure the degree of coherence between two fields. The two fields of orthogonal components ($Hx-Ey$ and $Hy-Ex$) are linearly correlated; that is, the coherence is 1. Otherwise, they are uncorrelated. The stronger the MT data noise is, the worse the coherence, and the closer the coherence value will be to 0.

As seen from Fig. 9, the two fields of the original time-series are affected by strong interference at 20–0.3 Hz, leading to a simultaneous reduction in the coherence and SNR and the increase in the error. The result of the original data shows a lower data quality. Comparing the proposed method with RR, OMP, and fractal entropy and clustering methods, it can be seen that the coherence of low-frequency $Hx-Ey$ components and the SNR of Ex and Ey field data are improved, and the error of the ρ_{xy} and ρ_{yx} is reduced. Thus, these results provide a basic for

concluding that the proposed methods can improve the quality of MT data.

This method will fail when the amplitudes of the noise and signal are very ambiguous because the RCMDE is not sensitive to the amplitude of similar waveforms. In addition, finding the optimal atom matching with the OMP algorithm and improving the multiscale analysis parameters are the focus of future research.

Conclusions

We have proposed a novel noise separation method for MT data using RCMDE and the OMP algorithm. As a robust feature parameter, the RCMDE can generate multiple feature parameter values to distinguish MT signals and noise. The OMP algorithm is used as a rapid denoising method. Combined the RCMDE and the OMP algorithm, we improved the efficiency and accuracy of MT feature extraction, identification and noise separation. The experimental results show that the identified strong interference is purposefully eliminated, the useful MT signals are bounteously preserved, and the quality of the MT data is improved. The apparent resistivity-phase curve obtained by using the proposed method becomes more continuous and smoother, and the polarization direction becomes more scattered and random. This method will further provide an innovative technology route for MT signal processing and obtain a high-precision MT response for subsequent electromagnetic inversion.

Abbreviations

MT: Magnetotelluric; AMT: Audio magnetotelluric; RCMDE: Refined composite multiscale dispersion entropy; MP: Matching pursuit; OMP: Orthogonal MP; VMD: Variational mode decomposition; SNR: Signal-to-noise ratio; MSE: Mean square error; NCC: Normalized cross-correlation; DE: Dispersion entropy; RR: Remote reference; FCM: Fuzzy c-mean; ME: Multiscale entropy; MDE: Multiscale dispersion entropy; DFA: Detrended fluctuation analysis; SD: Standard deviation; NCDF: Normal cumulative distribution function; SE: Sample entropy.

Acknowledgements

We would like to thank Cong Zhou, Zhimin Xu and Guang Li for their helpful discussions. We would also like to thank the anonymous reviewers and editor for providing critical comments that improve the manuscript greatly.

Authors' contributions

XZ wrote the manuscript, designed and analyzed the experiments; JL and DL conceived the idea and helped revise the manuscript; YL helped analyze the experimental results; BL and YH helped discuss the algorithm and experiment. All authors read and approved the final manuscript.

Funding

This research was supported by the National Key R&D Program of China (No. 2018YFC0807802, No. 2018YFE0208300), the National Natural Science Foundation of China (No. 42074084, No. 41874081), the Open Research Fund Program of Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education (No. 2020YSJS06), the Key Laboratory of Geophysical Electromagnetic Probing Technologies of Ministry of Natural Resources (No. KLGEPT201905), and the Natural Science Foundation of Hunan Province (No. 2018JJ2258).

Availability of data and materials

The datasets and MATLAB code used during the current study are available from the corresponding authors on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

Author details

¹ Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment, Monitoring Ministry of Education, School of Geoscience and Info-Physics, Central South University, Changsha 410083, China. ² Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China. ³ Key Laboratory of Geophysical Electromagnetic Probing Technologies of Ministry of Natural Resources, Institute of Geophysical and Geochemical Exploration, Chinese Academy of Geological Science, Langfang 065000, China. ⁴ Hunan Key Laboratory of Nonferrous Resources and Geological Hazard Exploration, Changsha 410083, China. ⁵ College of Mathematics and Physics, Hunan University Arts and Science, Changde 415000, China.

Received: 6 August 2020 Accepted: 11 March 2021

Published online: 23 March 2021

References

- Azami H, Rostaghi M, Abásolo D, Escudero J (2017) Refined composite multiscale dispersion entropy and its application to biomedical signals. *IEEE T Bio-Med Eng* 64:2872–2879
- Bandt C, Pompe B (2002) Permutation entropy: a natural complexity measure for time series. *Phys Rev Lett* 88:174102
- Becher WD, Sharpe CB (1969) A synthesis approach to magnetotelluric exploration. *Radio Sci* 4(11):1089–1094
- Cagniard L (1953) Basic theory of the magnetotelluric method of geophysical prospecting. *Geophysics* 18(3):605–635
- Cai JH, Tang JT, Hua XR, Gong YR (2009) An analysis method for magnetotelluric data based on the Hilbert-Huang transform. *Explor Geophys* 40(2):197–205
- Cai TT, Wang L (2011) Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE T Inform theory* 57(7):4680–4688
- Chave AD, Thomson DJ (1989) Some comments on magnetotelluric response function estimation. *J Geophys Res* 94(10):14215–14225
- Chave AD, Thomson DJ (2004) Bounded influence magnetotelluric response function estimation. *Geophys J Int* 157(3):988–1006
- Costa M, Goldberger AL, Peng CK (2005) Multiscale entropy analysis of biological signals. *Phys Rev E* 71:021906
- Egbert GD, Booker JR (1986) Robust estimation of geomagnetic transfer functions. *Geophys J Roy Astr Soc* 87(1):173–194
- Egbert GD, Livelybrooks DW (1996) Single station magnetotelluric impedance estimation: coherence weighting and the regression M-estimate. *Geophysics* 61(4):964–970
- Eisel M, Egbert GD (2001) On the stability of magnetotelluric transfer function estimates and the reliability of their variances. *Geophys J Int* 144:65–82
- Gamble TM, Goubau WM, Clarke J (1978) Magnetotelluric data analysis: removal of bias. *Geophysics* 43(6):1157–1169
- Gamble TM, Goubau WM, Clarke J (1979) Magnetotelluric with a remote magnetic reference. *Geophysics* 44(1):53–68
- Hernance JF (1973) Processing of magnetotelluric data. *Phys Earth Planet Interiors* 7(3):349–364
- Huang H, Makur A (2011) Backtracking-based matching pursuit method for sparse signal reconstruction. *IEEE Signal Proc Lett* 18(7):391–394

- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc A Math Phys Eng Sci* 454:903–995
- Jin W, Wang L, Zeng X, Liu Z, Fu R (2014) Classification of clouds in satellite imagery using over-complete dictionary via sparse representation. *Pattern Recogn Lett* 49(1):193–200
- Jones AG, Chave AD, Egbert GD, Auld D, Bahr K (1989) A comparison of techniques for magnetotelluric impedance estimation. *J Geophys Res* 94(10):14201–14213
- Kosko B (1986) Fuzzy entropy and conditioning. *Inform Sci* 40(2):165–174
- Li G, Liu XQ, Tang JT, Deng JZ, Hu SG, Zhou C, Chen CJ, Tang WW (2020a) Improved shift-invariant sparse coding for noise attenuation of magnetotelluric data. *Earth Planets Space* 72:45
- Li J, Zhang X, Gong JZ, Tang JT, Ren ZY, Li G, Deng YL, Cai J (2018) Signal-noise identification of magnetotelluric signals using fractal-entropy and clustering algorithm for targeted de-noising. *Fractals* 26(2):1840011
- Li J, Zhang X, Tang JT (2020b) Noise suppression for magnetotelluric using variational mode decomposition and detrended fluctuation analysis. *J Appl Geophys* 180:104127
- Li J, Peng YQ, Tang JT, Li Y (2021) Denoising of magnetotelluric data using K-SVD dictionary training. *Geophys Prospect* 69(2):448–473
- Mallat SG, Zhang Z (1993) Matching pursuit with time-frequency dictionaries. *IEEE Trans Signal Process* 41(12):3397–3415
- Mitiche I, Morison G, Nesbitt A, Hughes-Narborough M, Stewart BG, Boreham P (2018) Classification of partial discharge signals by combining adaptive local iterative filtering and entropy features. *Sensors* 18:406
- Needell D, Vershynin R (2010) Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE J-STSP* 4(2):310–316
- Pati YC, Rezaiifar R, Krishnaprasad PS (1993) Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *Proceedings of 27th Asilomar Conference on Signals Systems and Computers* 40–44
- Pincus SM (1991) Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci* 88:2297–2301
- Qi J, Zhang L, Zhang K, Li L, Sun J (2020) The application of improved differential evolution algorithm in electromagnetic fracture monitoring. *Adv Geo-Energy Res* 4:233–246
- Richman JS, Lake DE, Moorman JR (2004) Sample entropy. *Numerical Computer Method, Part E*, pp 172–184
- Ritter O, Junge A, Dawes G (1998) New equipment and processing for magnetotelluric remote reference observations. *Geophys J Int* 132(3):535–548
- Rostaghi M, Azami H (2016) Dispersion entropy: a measure for time series analysis. *IEEE Signal Proc Lett* 23:610–614
- Tang JT, Li J, Xiao X, Zhang LC, Lv QT (2012) Mathematical morphology filtering and noise suppression of magnetotelluric sounding data. *Chin J Geophys* 55(5):1784–1793
- Tikhonov AN (1950) On determining electrical characteristics of the deep layers of the Earth's crust. *Dokl Akad Nauk SSSR* 73:295–297
- Trad DO, Travassos JM (2000) Wavelet filtering of magnetotelluric data. *Geophysics* 65(2):482–491
- Tropp JA, Gilbert AC (2007) Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans Inform Theory* 53(12):4655–4666
- Vallianatos F (1996) Magnetotelluric response of a randomly layered earth. *Geophys J Int* 125(2):577–583
- Varentsov IM (2006) Arrays of simultaneous electromagnetic sounding: design, data processing and analysis. *Methods Geochem Geophys* 40:259–273
- Wang H, Campaña J, Cheng JL, Zhu GW, Wei WB, Jin S, Ye GF (2017) Synthesis of natural electric and magnet Time-series using Inter-station transfer functions and time-series from a Neighboring site (STIN): applications for processing MT data. *J Geophys Res-Sol Ea* 122(8):5835–5851
- Wang JB, Wang SX, Yin HJ, Zhang R (2013) A self-adaption denoising method using orthogonal matching pursuit. *SEG Technical Program Expanded Abstracts*
- 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(3):635–652
- Xu Y, Chen R, Li Y, Zhang P, Yang J, Zhao X, Liu M, Wu D (2019) Multispectral image segmentation based on a fuzzy clustering algorithm combined with Tsallis entropy and a gaussian mixture model. *Remote Sens* 11:2772
- Zhang H, Liu J, Chen L, Chen N, Yang X (2019) Fuzzy Clustering algorithm with non-neighborhood spatial information for surface roughness measurement based on the reflected aliasing images. *Sensors* 19:3285
- Zhang YD, Tong SG, Cong FY, Xu J (2018) Research of feature extraction method based on sparse reconstruction and multiscale dispersion entropy. *Appl Sci* 8:888

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)