Journal of the Earth and Space Physics
Homepage: https://jesphys.ut.ac.ir
Vol. 51, No. 4, Winter 2026, P. 75-88
(Research Article)

Mahalanobis Distance Method for Enhancing Magnetotelluric Data Processing: A Case Study from the Rochechouart Impact Structure, France

Parnian-Khoy, H.[1] | Motavalli-Anbaran, S. H.[1] ✉ | Habibian Dehkordi, B.[1] | Sailhac, P.[2] | Mousapour Yasoori, M.[1] | Quesnel, Y.[3] | Lambert, Ph.[4]

1. Department of Earth Physics, Institute of Geophysics, University of Tehran, Tehran, Iran.
2. Department of Earth Sciences, Geosciences Paris-Saclay Laboratory (GEOPS), Université Paris-Saclay, Orsay, France.
3. CEREGE, Aix-Marseille Université, CNRS, IRD, INRAE, Aix-en-Provence, France.
4. CIRIR – Center for International Research and Restitution on Impacts and on Rochechouart, Rochechouart, France.

Corresponding Author E-mail: motavalli@ut.ac.ir

(Received: 11 Jan 2026, Revised: 7 Feb 2026, Accepted: 28 Feb 2025, Published online: 17 March 2026)

Abstract

One of the primary challenges in magnetotelluric data processing is the presence of noise and outliers (anomalous values). These disturbances often come from human-made sources such as power lines, electronic devices, and nearby infrastructure. They can significantly affect the results related to apparent resistivity and phase, leading to unreliable models of subsurface electrical resistivity.

To identify and remove these outlier and noisy components, the Mahalanobis distance method is proposed as an effective solution. This approach — applied here in a four-dimensional feature space comprising the real and imaginary parts of two components of the impedance transfer function — calculates the distance of each data point from the dataset, meanwhile accounting for variances, covariances, and correlations between variables, thereby enabling the detection of anomalous points more accurately than simpler (2D) approaches.

In this study, to identify outliers, we applied the Mahalanobis distance method to real data from a single MT station located at the Chassenon Forage site within the Rochechouart impact structure, France. The results demonstrate that this approach not only improves the accuracy of subsequent analyses and enables the extraction of more precise information from subsurface structures, but also reduces processing time by efficiently eliminating contaminated windows before final impedance estimation.

Keywords: Magnetotelluric, Outliers, Mahalanobis distance, Data processing.

1. Introduction

The magnetotelluric (MT) method is a passive geophysical technique that utilizes natural electromagnetic waves over a wide frequency range to characterize subsurface electrical resistivity structures (Vozoff, 1972). The foundational concepts of this method presented by researchers including Rikitake (1946), Price (1950), Kato and Kikuchi (1950), Tikhonov (1950), and Cagniard (1953). In magnetotelluric, the horizontal components of the electric field in two orthogonal directions and the magnetic field in three orthogonal directions are measured at the Earth's surface. Variations in these measured fields are then used to compute the impedance transfer function. The resulting impedance provides critical insights into the electrical anisotropy of the subsurface in a given area (Daud, 2010).

MT data processing faces significant challenges due to the presence of noise and outliers, which can bias impedance estimates and lead to unreliable subsurface models (Egbert & Booker, 1986; Chave et al., 1987; Ritter et al., 1998). Over the decades, various approaches have been developed to mitigate these issues, including robust estimation techniques (Egbert, 1997; Smirnov, 2003), remote reference methods (Gamble et al., 1979), coherence-based selection (Swift, 1967; Jones & Jödicke, 1984), polarization direction filtering (Weckmann et al., 2005), and more recently, automated pre-selection tools based on Mahalanobis distance (Platz & Weckmann, 2019) to detect and reject noisy segments or outliers prior to impedance calculation. Despite these advances, effective

Cite this article: Parnian-Khoy, H., Motavalli-Anbaran, S. H., Habibian Dehkordi, B., Sailhac, P., Mousapour Yasoori, M., Quesnel, Y., & Lambert, Ph. (2026). Mahalanobis Distance Method for Enhancing Magnetotelluric Data Processing: A Case Study from the Rochechouart Impact Structure, France. *Journal of the Earth and Space Physics*, 51(4), 75-88. DOI: http//doi.org/10.22059/jesphys.2026.409669.1007750

E-mail: (1) h.oseinparnian@ut.ac.ir | bhabibian@ut.ac.ir | mousapour@ut.ac.ir (2) pascal.sailhac@universite-paris-saclay.fr (3) quesnel@cerege.fr (4) lambertbdx@gmail.com

© Authors Retain the Copyright and Full Publishing Rights.
Publisher: University of Tehran Press.
DOI: http//doi.org/10.22059/jesphys.2026.409669.1007750
Print ISSN: 2538-371X
Online ISSN: 2538-3906

outlier rejection in single-station data remains an active area of research, particularly when remote reference data are unavailable or when the focus is on automated, multivariate approaches.

In this article, the stages of magnetotelluric data processing are outlined, with a detailed examination of outlier and noise removal using the novel Mahalanobis distance approach. To facilitate a better understanding of the method, an overview of magnetotelluric theory is first presented. The theoretical framework for the MT method in a layered Earth—fundamentally derived from the propagation of plane electromagnetic waves in a homogeneous, isotropic, linear, and lossy medium—was introduced by Cagniard (1953) and Keller and Frischknecht (1966). In such an environment, the electric and magnetic fields propagate orthogonally, and the ratio of the electric field to the magnetic field at any instant is called impedance.

$$Z_{xy} = \frac{E_x}{H_y} \quad (1)$$

In this frequency-domain equation, Z represents the impedance, E_x is the electric field strength in the north direction, and H_y is the magnetic field in the east direction. Based on Equation (1) and assuming the Earth behaves as a system responding to an input signal with a corresponding transfer function, the magnetic field serves as the input signal, while the electric field represents the output signal from the Earth system. In other words, the interaction of the magnetic field (as input) with the Earth's surface induces a secondary electric field within the subsurface (Ajithabh & Patro., 2023), which is recorded as the Earth's response to the input signal.

The impedance tensor, which is essentially a transfer function, is expressed as a complex impedance tensor in the following form:

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix} \quad (2)$$

The components of the impedance tensor are complex values with defined magnitude and phase at each frequency. The apparent resistivity (ρ) is related to the magnitude of the impedance, while its phase (φ) represents the phase difference between the electric and magnetic fields at each instant, calculated as follows:

$$\rho = \frac{1}{\omega \mu_0} |Z|^2 \quad (3)$$

$$\varphi = tan^{-1}\left(\frac{Im(Z)}{Re(Z)}\right) \quad (4)$$

$$Im(Z) = ImaginarypartofZandRe(Z) = RealPartofZ$$

where μ_0 is the magnetic permeability of free space, $\omega=2\pi f$ is the angular frequency, and $|Z|$ denotes the magnitude of the impedance. In practice, for the off-diagonal elements of the impedance tensor (typically Z_{xy} or Z_{yx}), apparent resistivity and phase curves are derived separately for each polarization, providing diagnostic information about subsurface conductivity structure.

The tipper transfer function also establishes a relationship between the horizontal magnetic field components (H_x, H_y) and the vertical magnetic field component (H_z), expressed as follows:

$$H_z = \begin{bmatrix} T_x & T_y \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix} \quad (5)$$

where T_x and T_y are the complex components of the tipper vector. The tipper vector is used to identify the direction of induction arrows and to understand lateral conductivity variations, ultimately improving horizontal resolution (Ajithabh & Patro., 2023). Upon closer examination of the tipper relationship, it can be inferred that the horizontal magnetic field components act as the input to the Earth system, while the vertical component represents the output. In other words, the horizontal components arise from the wave-generating source, whereas the vertical component results from induction processes within the Earth—a fact that can be readily demonstrated by examining the transmission and reflection relationships of electromagnetic fields.

2. Methodology

Magnetotelluric data processing requires a sound understanding of electromagnetic wave theory and time-series analysis to accurately compute the impedance using various mathematical relationships based on linear systems. One of the most common challenges is that the five recorded components of electric and magnetic fields (Figure 1) are non-stationary signals that are perturbed by the presence of noise. A precise understanding of noise and its types is essential, as it typically originates from anthropogenic sources such as power lines, electronic equipment, and infrastructure. In magnetotelluric signal processing, through a

complex computational workflow, noise is ultimately removed, and the impedance and tipper transfer functions are estimated.

Given the processing of magnetotelluric data in the frequency domain, spectral analysis of raw data represents one of the most critical components of the processing workflow. Various techniques exist for transforming the recorded time series into spectral information in the frequency domain. Although these methods are theoretically equivalent, as they all derive Fourier components, significant practical differences arise depending on the signal characteristics and the underlying assumptions about the signal's nature (Hermance, 1973). In practice, physical insight into the results from each method plays a key role in determining their preferred application (Hermance, 1973).

A general overview of the processing workflow is presented in Figure 2. For a more detailed understanding, these steps are elaborated below, and ultimately, the Mahalanobis distance method is used to identify outliers and noisy data and is evaluated using real data.

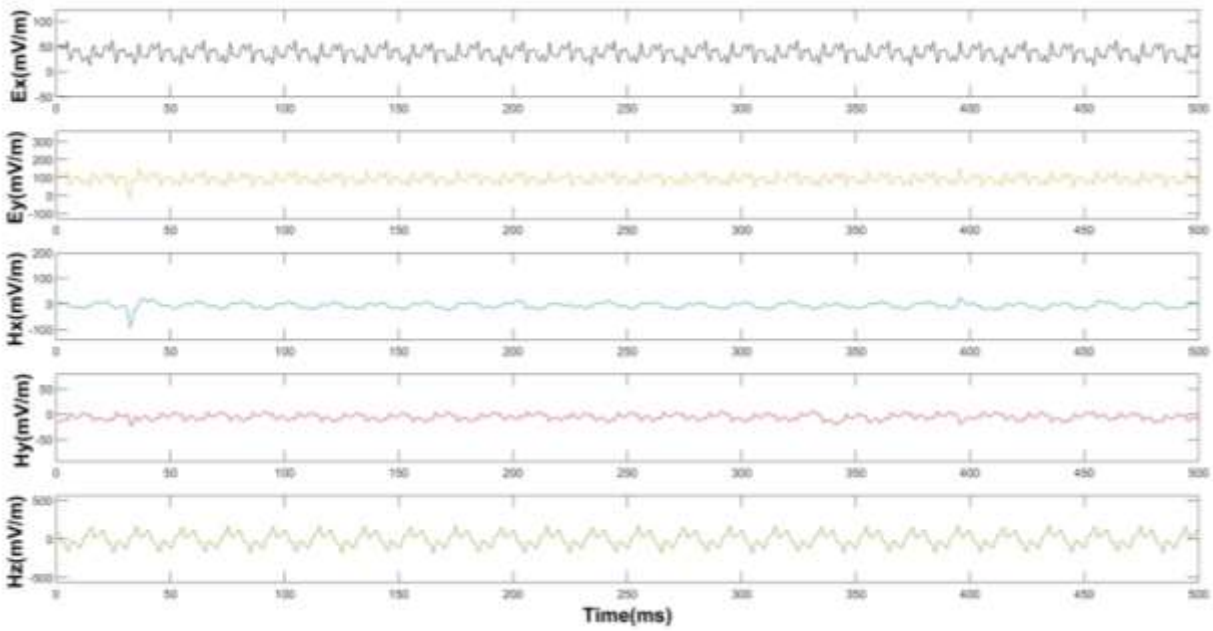

Figure 1. Raw time series of the five recorded magnetotelluric components (Ex, Ey, Hx, Hy, Hz). The displayed segment consists of 512 samples (500 ms), with the x-axis corresponding to time.

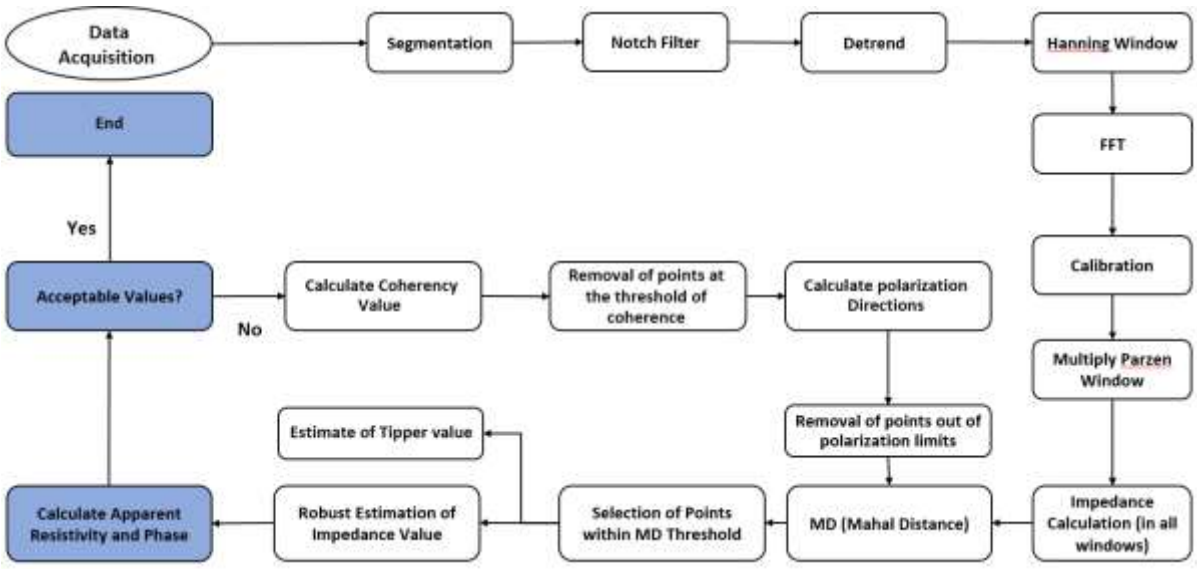

Figure 2. General schematic of the magnetotelluric data processing workflow.

The first step in magnetotelluric data processing is windowing the acquired data in the time domain. Given the presence of noise in many survey areas, a portion of the collected information is typically contaminated with noise. Therefore, during acquisition and considering the random nature of noise, magnetotelluric recording is conducted over longer time intervals to facilitate noise removal and enhance data quality. This interval can range from several hours to several days, depending on the study objectives. During this recording period, the electric and magnetic field components are continuously registered at a specified sampling rate. The data are then divided into smaller windows that contain a sufficient number of periods for the minimum required frequencies.

Due to the extended duration of magnetotelluric data acquisition, dividing the data into smaller windows introduces an approximately linear trend in the windows. The most critical aspect of this segmentation is determining an appropriate window length. This length is calculated based on the minimum window length. For this purpose, the empirical formula proposed by Borah et al. (2015) is employed.

$$WindowLength = MinimumWindowLength \times 2^n \quad (6)$$

where n = 1,2,3...

Building on the condition proposed by Borah et al. (2015), the number of windows can be calculated as follows:

$$\frac{Totalnumberofsamples}{MinimumWindowLength \times 2^n} \geq 20n \quad (7)$$

$$N = \frac{Totalnumberofsamples}{MinimumWindowLength \times 2^n}$$

By determining the maximum n that satisfies the aforementioned condition, the number of windows (N) can be calculated. It is essential to note that increasing the window length reduces the number of windows, resulting in less effective noise reduction. Conversely, decreasing the window length lowers frequency resolution. Therefore, selecting an optimal window length is the most critical aspect of this step, as neglecting it can significantly impact the final results.

After windowing, given the presence of noise associated with electrical currents in the survey environment, a notch filter is designed with center frequency Fc (50 Hz and its harmonics) and a specified bandwidth to remove these frequencies from the data.

In the subsequent step, the linear trend is removed from the magnetotelluric data (detrending), and a Hanning window is then applied to prevent spectral leakage in the frequency domain caused by windowing. All data must now be transformed to the frequency domain using the Fourier transform to enable the remaining processing steps to be performed in the frequency domain. Subsequently, the magnetic field data recorded by the coils must be calibrated using the calibration coefficients for each coil and converted to nanoTesla (nT) units.

Due to the slow varying nature of the impedance function in the frequency domain, spectral values at a target frequency are averaged using the Parzen window (Jenkins & Watts., 1968). The advantage of employing the Parzen window for averaging is that it assigns lower weights to values farther from the target (central) frequency. The Parzen window is defined as follows:

$$PV(f) = \begin{cases} 1 & |f_t - f| = 0 \\ \left(\frac{sin(u)}{u}\right)^4 & 0 < |f_t - f| < f_r \\ 0 & |f_t - f| \geq f_r \end{cases} \quad (8)$$

$$u = {(\pi|f_t - f|)}/{f_r}$$

where f_t represents the target frequency and $f_r$radius of the Parzen window. The radius of the Parzen window is determined based on the sampling frequency (Borah et al., 2015). Once the Parzen window values are specified, a weighted averaging is performed at each target frequency for the spectral values. Finally, after computing the averaged spectral values, the components of the impedance tensor and the tipper vector are calculated (Simpson & Bahr, 2005) as follows:

$$\langle E_x H_x^* \rangle = \sum ((E_x \times H_x{}^*) \times PV) / \sum sum(PV) \quad (9)$$

$$Z_{xx} = \frac{\langle H_y H_y^* \rangle \langle E_x H_x^* \rangle - \langle H_y H_x^* \rangle \langle E_x H_y^* \rangle}{\langle H_x H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_x H_y^* \rangle \langle H_y H_x^* \rangle} \quad (10)$$

$$Z_{xy} = \frac{\langle H_x H_x^* \rangle \langle E_x H_y^* \rangle - \langle H_x H_y^* \rangle \langle E_x H_x^* \rangle}{\langle H_x H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_x H_y^* \rangle \langle H_y H_x^* \rangle} \quad (11)$$

$$Z_{yx} = \frac{\langle H_y H_y^* \rangle \langle E_y H_x^* \rangle - \langle H_y H_x^* \rangle \langle E_y H_y^* \rangle}{\langle H_x H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_x H_y^* \rangle \langle H_y H_x^* \rangle} \quad (12)$$

$$Z_{yy} = \frac{\langle H_x H_x^* \rangle \langle E_y H_y^* \rangle - \langle H_x H_y^* \rangle \langle E_y H_x^* \rangle}{\langle H_x H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_x H_y^* \rangle \langle H_y H_x^* \rangle} \quad (13)$$

$$T_x = \frac{\langle H_z H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_z H_y^* \rangle \langle H_y H_x^* \rangle}{\langle H_x H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_x H_y^* \rangle \langle H_y H_x^* \rangle} \quad (14)$$

$$T_y = \frac{\langle H_z H_y^* \rangle \langle H_x H_x^* \rangle - \langle H_z H_x^* \rangle \langle H_x H_y^* \rangle}{\langle H_x H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_x H_y^* \rangle \langle H_y H_x^* \rangle} \quad (15)$$

The bracket terms represent averaged auto and cross spectral values. After computing the impedance tensor and tipper vector across all windows, the next step is to remove noise and outliers in the frequency domain using statistical methods. Common approaches include the coherence test, examination of electric and magnetic fields polarization direction, and calculation of the Mahalanobis distance. In this study, the Mahalanobis distance method is implemented in a four-dimensional feature space to robustly identify and remove outliers. This approach not only significantly improves data quality and analytical accuracy but also reduces computational processing time.

The coherence test is briefly employed to assess the linear relationship between the measured and predicted electric fields, with low values indicating the presence of incoherent noise (Jones & Jodicke., 1984). Similarly, field polarization directions can serve as a simple tool for detecting certain types of noise (Fowler et al., 1967).

The main method under investigation is the Mahalanobis distance, a multivariate measure for outlier detection. Originally introduced by Mahalanobis (1936), this approach computes the distance of a point from the distribution center while accounting for correlations and variances, rendering it scale-independent. This property makes the Mahalanobis distance a powerful tool for identifying and removing outliers and noisy data in magnetotelluric processing.

Following the application of these analytical methods, the final impedance results are computed using the robust estimation method. This statistical method is applicable to datasets with a normal distribution but is contaminated by outliers and is not sensitive to non-Gaussian errors. Various robust estimation methods have been developed, and robust estimation methods with maximum likelihood are commonly used in MT data processing. The maximum likelihood method is an iteratively reweighted adjustment technique that prevents the estimated values from being controlled by a small number of data points (Egbert & Booker, 1986). In this study, the Huber M-estimator was utilized for robust estimation (Ritter at el., 1998). To use the Huber method, an initial value is required, which is acquired using the Jackknife estimator (Efron., 1982; Chave et al., 1987).

3. Mahalanobis Distance (MD)

The Mahalanobis distance is a highly effective tool for the automatic rejection of noisy data in the magnetotelluric method (Platz & Weckmann, 2019). Since only a single impedance value exists at each target frequency, noisy windows must be removed at various stages, ultimately yielding one final window from the remaining ones in the last step.

If the real and imaginary parts of the impedance components across all windows are plotted on a complex scatter diagram for a given frequency (Figure 3), a cluster with scatter around the mean value is observed, representing the impedance estimates. In this diagram, values far from the cluster center (outliers) are removed using the Mahalanobis distance method.

The Mahalanobis distance for a given frequency across different windows is calculated as follows (Platz & Weckmann, 2019):

$$MD_i = \sqrt{(x_i - \mu) C_x^{-1} (x_i - \mu)^T} \quad (16)$$

where μ represents the mean and C_x the covariance of the dataset x. The index i denotes calculations across different windows.

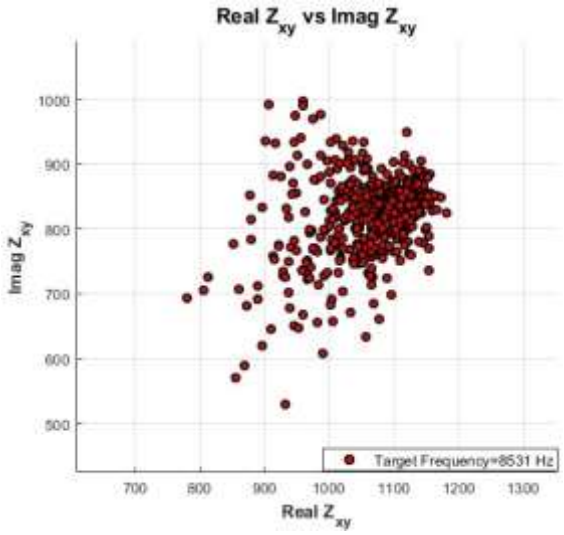

Figure 3. Complex scatter diagram showing the real and imaginary parts of Z_{xy} impedance component for all windows at a frequency of 8531 Hz. Axes are in (mV/m)/nT.

In magnetotelluric processing, the Mahalanobis distance method is applied to remove outliers, considering the electric field polarization direction. Based on Equation (2), the impedance components Z_{xx} and Z_{xy} are derived using the Ex component, while Z_{yx} and Z_{yy} are derived using the Eycomponent. To form the data matrix x, for example in the Exelectric field direction, the following structure is used, which results in a four-dimensional feature space consisting of the real and imaginary parts of two impedance components:

$$x = \begin{matrix} X1 & X2 & X3 & X4 \end{matrix} \begin{bmatrix} Real(Z_{xx}1) & Imag(Z_{xx}1) & Real(Z_{xy}1) & Imag(Z_{xy}1) \\ \vdots & \vdots & \vdots & \vdots \\ Real(Z_{xx}n) & Imag(Z_{xx}n) & Real(Z_{xy}n) & Imag(Z_{xy}n) \end{bmatrix} \quad (17)$$

This four-dimensional approach allows the Mahalanobis distance to account for correlations between all four variables simultaneously, improving outlier detection compared to bivariate method.

Subsequently, the mean vector μ and the covariance matrix C of the data matrix x are computed as follows:

$$\mu = [\mu_1 \mu_2 \mu_3 \mu_4] \quad (18)$$

$$C_x = \begin{bmatrix} Cov(X1,X1) & Cov(X1,X2) & Cov(X1,X3) & Cov(X1,X4) \\ Cov(X2,X1) & Cov(X2,X2) & Cov(X2,X3) & Cov(X2,X4) \\ Cov(X3,X1) & Cov(X3,X2) & Cov(X3,X3) & Cov(X3,X4) \\ Cov(X4,X1) & Cov(X4,X2) & Cov(X4,X3) & Cov(X4,X4) \end{bmatrix} \quad (19)$$

where $X_1 = Real(Z_{xx})$, $X_2 = Imaginary(Z_{xx})$, $X_3 = Real(Z_{xy})$ and $X_4 = Imaginary(Z_{xy})$.

With the data matrix **x**, mean μ, and covariance C now available, the Mahalanobis distance values are readily computed using the matrix form of the equation above. The four-dimensional formulation ensures that the covariance structure between all impedance components is properly considered in the outlier detection process.

Finally, by setting a threshold for the Mahalanobis distance, impedance matrix values exceeding this threshold are removed. It should be noted that in the steps described, the Mahalanobis distance is calculated for a single frequency across different windows.

4. Case study

To validate the effectiveness of the proposed method, the Mahalanobis distance approach was applied to real magnetotelluric data from two stations — the Chassenon Forage station and the Forêt Beaulieu station — both located within the Rochechouart impact structure in France (Quesnel et al., 2021). These datasets represent typical field conditions with a moderate level of environmental and anthropogenic noise. Although a detailed noise analysis is beyond the scope of this study, the presence of noise is reflected in the error bars of the apparent resistivity and phase curves shown in the results. Therefore, these data provide an appropriate case study for evaluating the performance of the proposed outlier rejection technique under realistic conditions.

At the Chassenon Forage station, a sampling frequency of 65,536 Hz, a target frequency of 8,531 Hz, and an acquisition duration of one minute were selected to evaluate the method's performance under E_x polarization. At the Forêt Beaulieu station, a sampling frequency of 4,096 Hz, a target frequency of 40 Hz, and an acquisition duration of 25 minutes were used. All other processing parameters — including the polarization type (Ex) and the three Mahalanobis distance thresholds (MD = 1.5, MD = 3, and MD = 5) — were kept identical to those used for the Chassenon Forage station to ensure direct comparison.

These thresholds were tested to assess their influence on outlier removal and overall data

quality. The processed results, including apparent resistivity and phase curves over the available frequency range (determined by each station's sampling rate), were systematically compared. Additionally, the computational time required for each threshold was recorded to evaluate processing efficiency.

The resulting apparent resistivity, phase, and complex scatter diagram for thresholds of 1.5, 3, and 5 are presented in Figures 4 to 9 (subplots a to c), respectively.

First, the threshold of 1.5 was examined for the Chassenon Forage station. The complex scatter diagrams for the Z_{xy} and Z_{xx} impedances at 8,531 Hz are shown in Figures 4a and 4b. The processing time was 37 seconds, with 343 outliers and 136 inliers identified. The resulting apparent resistivity and phase curves show smooth variations, indicating stable subsurface resistivity responses.

The same threshold was then applied to the Forêt Beaulieu station at 40 Hz. The corresponding complex scatter diagrams for the Z_{xy} and Z_{xx} impedances are presented in Figures 5a and 5b. The processing time was 45 seconds, with 610 outliers and 139 inliers detected. The resulting apparent resistivity and phase curves also show smooth and consistent variations, confirming the effectiveness of the method under different acquisition conditions.

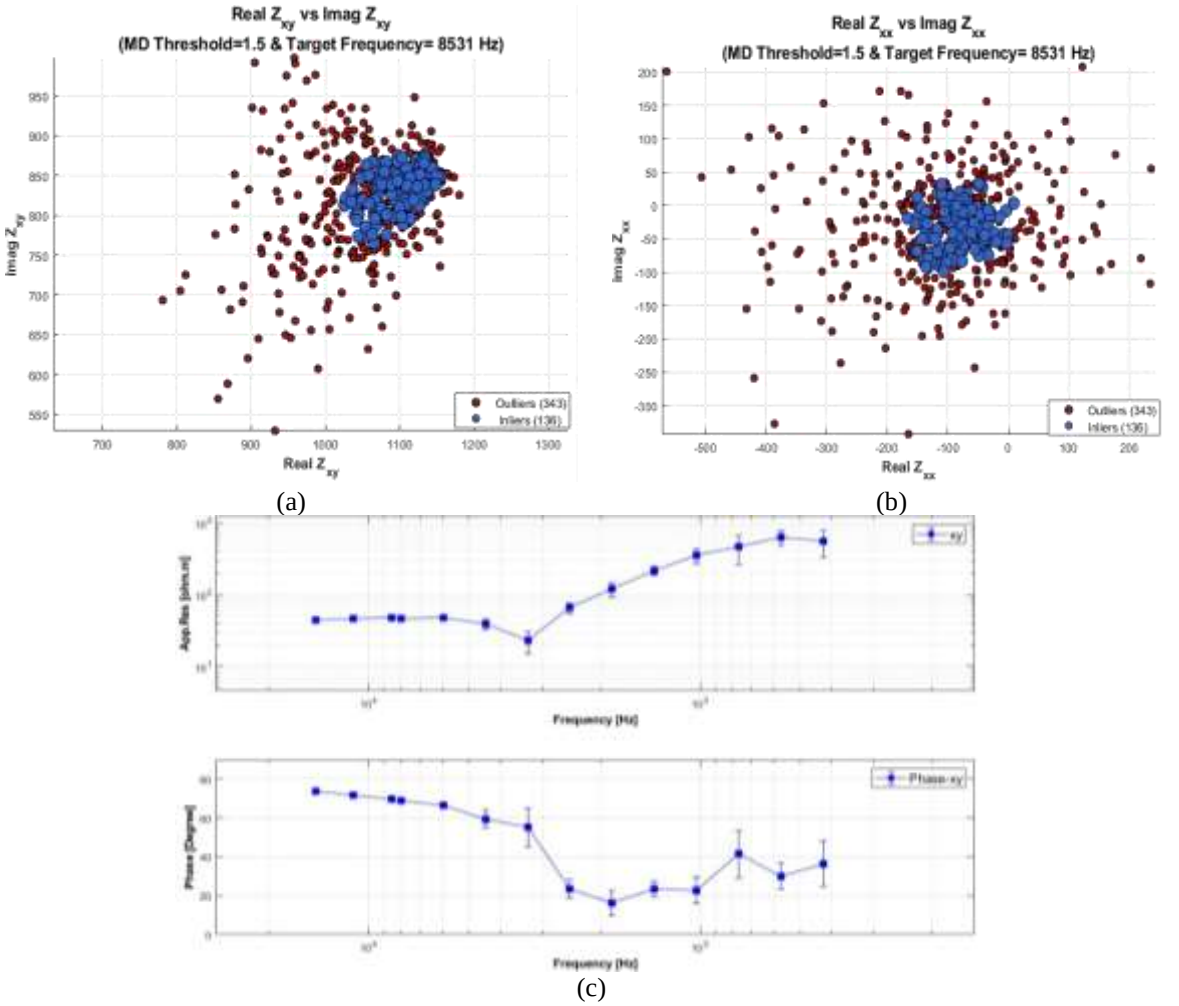

Figure 4. Chassenon Forage station, Threshold = 1.5. (a) Complex scatter diagram of real vs. imaginary parts of Zxy impedance. (b) Complex scatter diagram of real vs. imaginary parts of Zxx impedance. (c) Apparent resistivity and phase curves for Ex polarization.

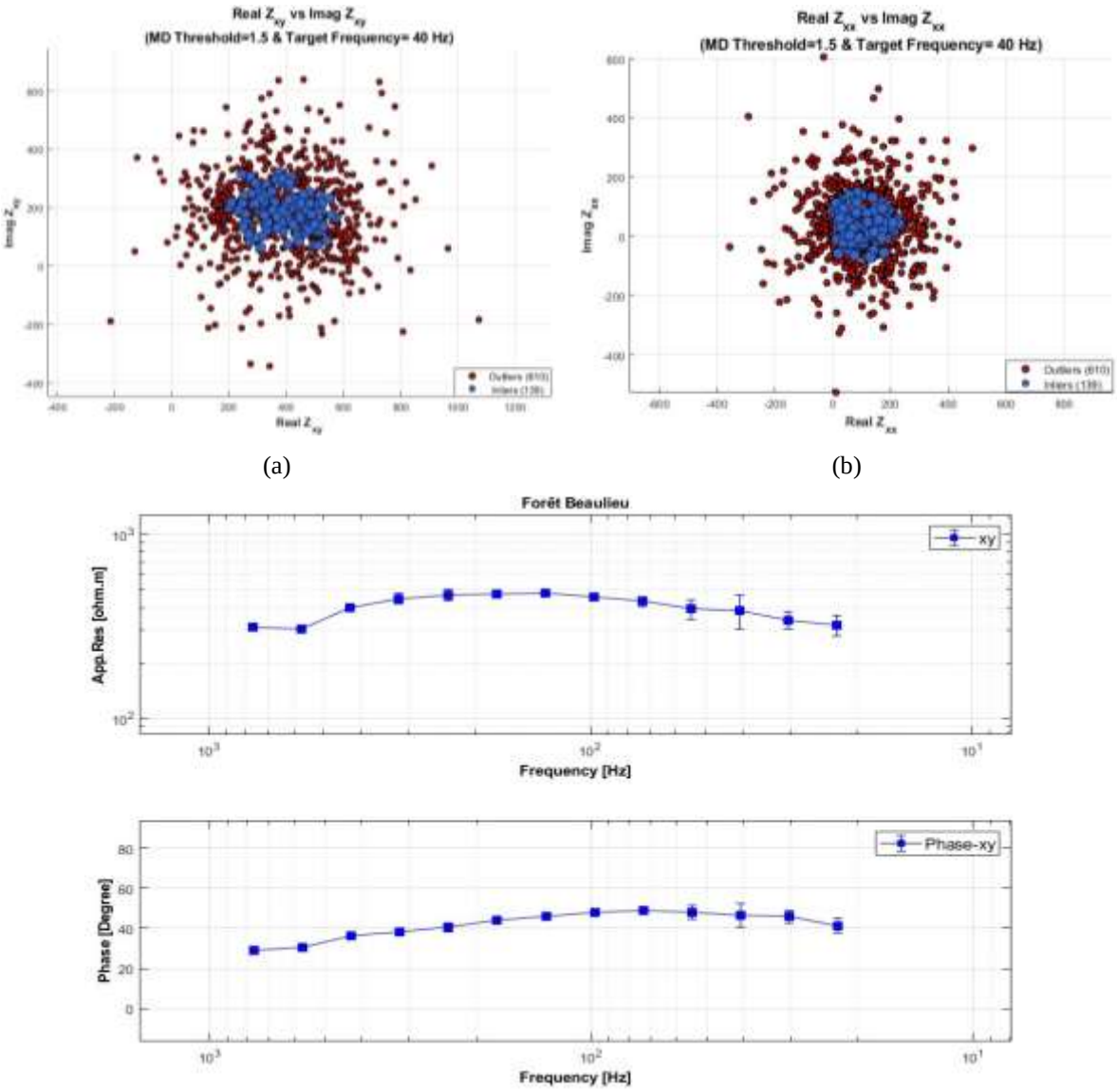

Figure 5. Forêt Beaulieu station, Threshold = 1.5. (a) Complex scatter diagram of real vs. imaginary parts of Zxy impedance. (b) Complex scatter diagram of real vs. imaginary parts of Zxx impedance. (c) Apparent resistivity and phase curves for Ex polarization.

Next, a threshold of 3 was applied to the Chassenon Forage station, with the results shown in Figure 6 (a–c). The processing time increased to 65 seconds, with 138 outliers and 341 inliers detected. The apparent resistivity and phase curves remained smooth and showed little variation compared to the threshold of 1.5.

The same threshold was then applied to the Forêt Beaulieu station (Figure 7). In this case, the processing time increased to 91 seconds, with 333 outliers and 416 inliers identified. Similar to the previous threshold, the apparent resistivity and phase curves remained stable and exhibited only minor differences compared to the results obtained with a threshold of 1.5.

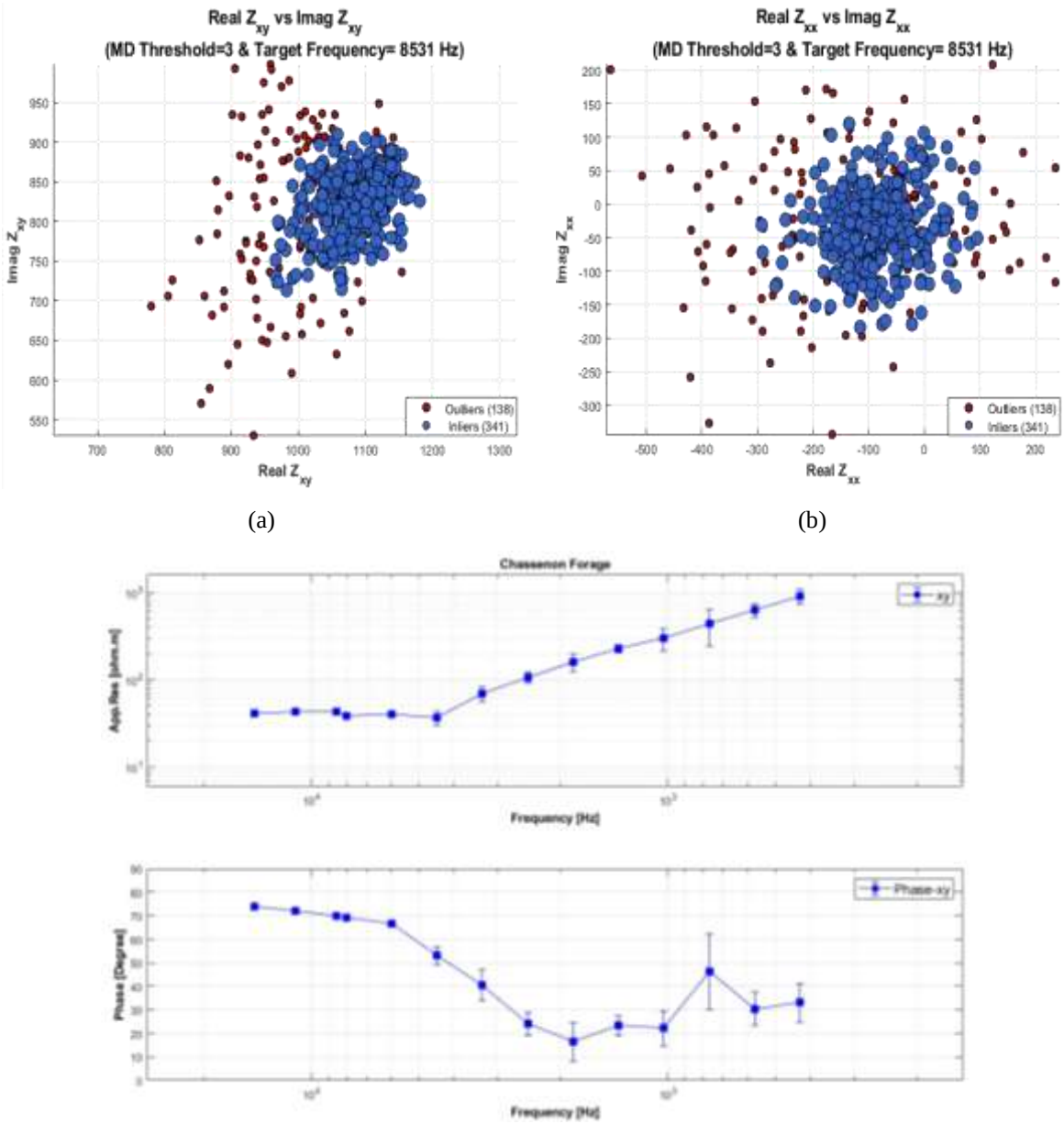

Figure 6. Chassenon Forage station, Threshold = 3. (a) Complex scatter diagram of real vs. imaginary parts of Zxy impedance. (b) Complex scatter diagram of real vs. imaginary parts of Zxx impedance. (c) Apparent resistivity and phase curves for Ex polarization.

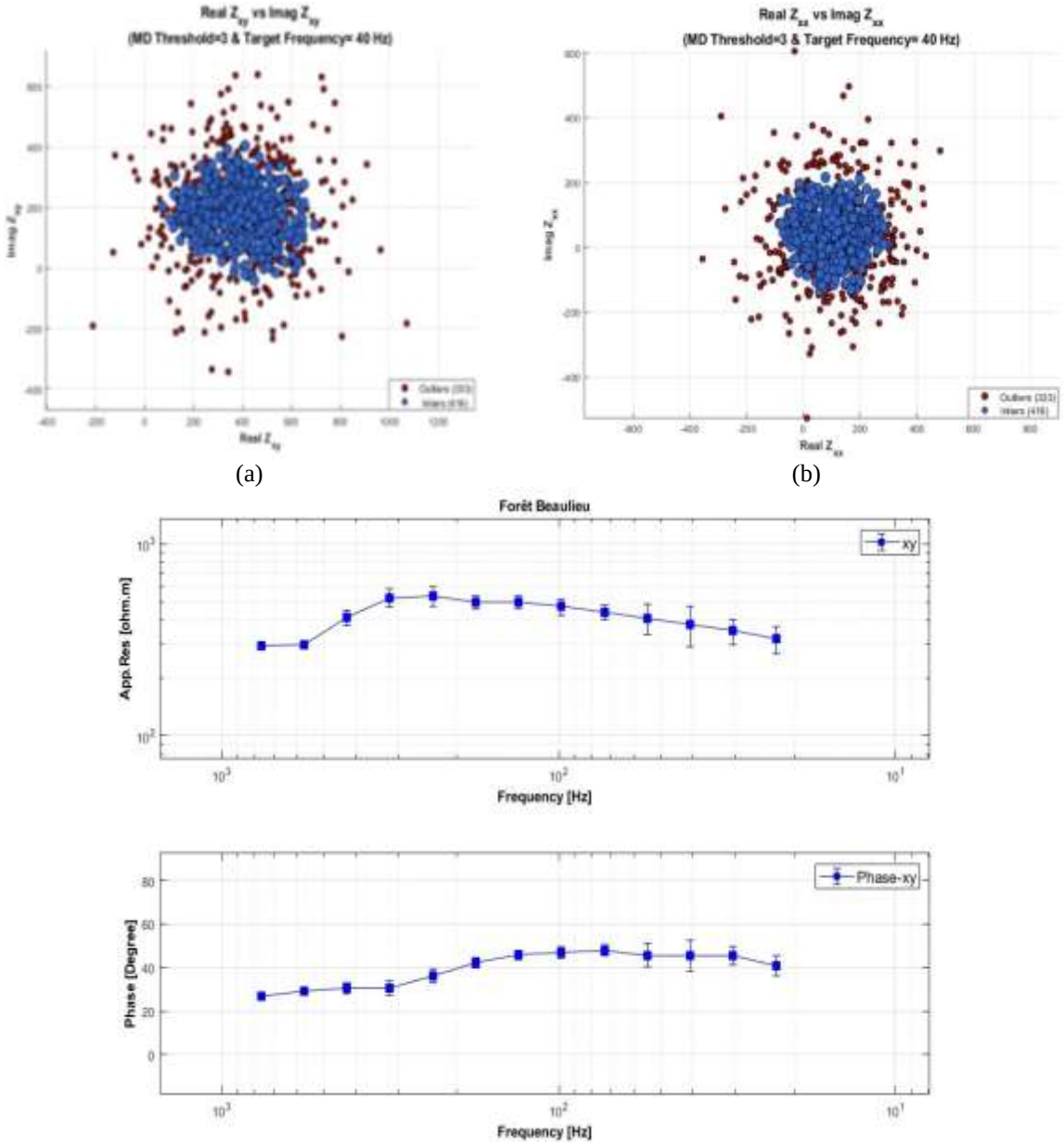

Figure 7. Forêt Beaulieu station, Threshold = 3. (a) Complex scatter diagram of real vs. imaginary parts of Zxy impedance. (b) Complex scatter diagram of real vs. imaginary parts of Zxx impedance. (c) Apparent resistivity and phase curves for Ex polarization.

In the final step, both stations were evaluated using a threshold value of 5. The results for the Chassenon Forage station are presented in Figure 8 (a–c). At this threshold, the processing time increased to 90 seconds, with 42 outliers and 437 inliers identified.

The corresponding results for the Forêt Beaulieu station are shown in Figure 9 (a–c). In this case, the processing time reached 121 seconds, with 41 outliers and 708 inliers detected.

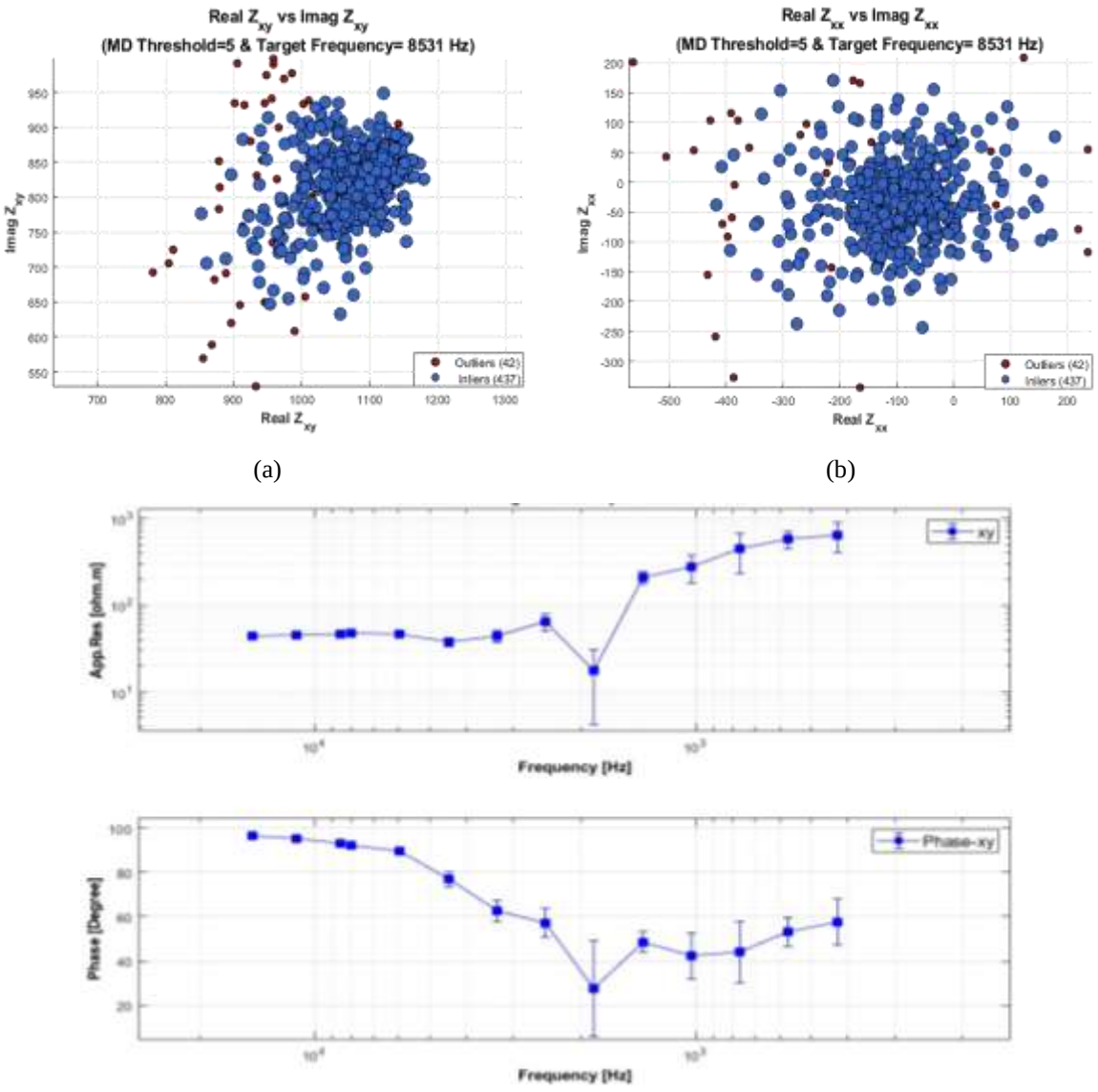

Figure 8. Chassenon Forage station, Threshold = 5. (a) Complex scatter diagram of real vs. imaginary parts of Zxy impedance. (b) Complex scatter diagram of real vs. imaginary parts of Zxx impedance. (c) Apparent resistivity and phase curves for Ex polarization.

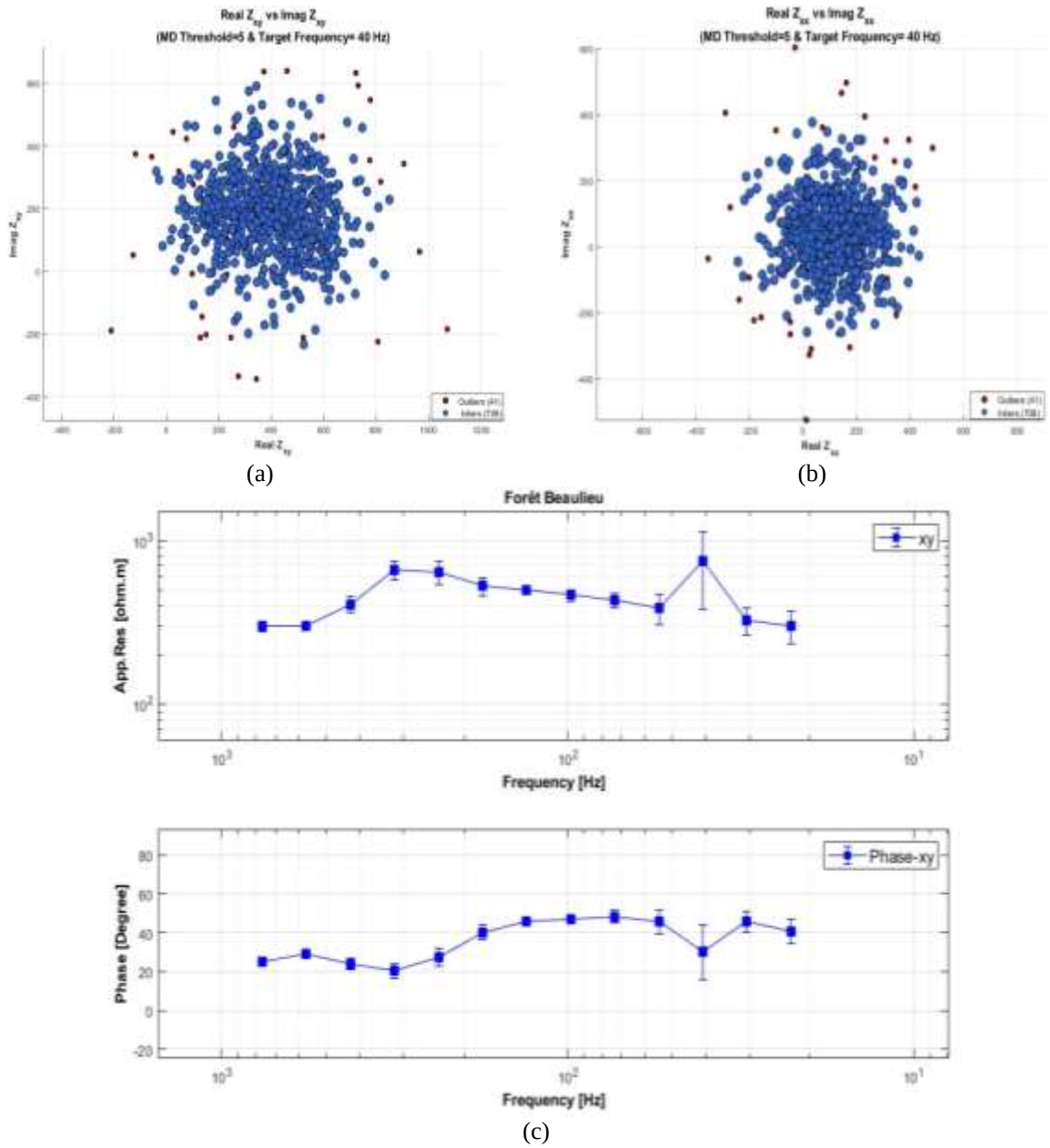

Figure 9. Forêt Beaulieu station, Threshold = 5. (a) Complex scatter diagram of real vs. imaginary parts of Zxy impedance. (b) Complex scatter diagram of real vs. imaginary parts of Zxx impedance. (c) Apparent resistivity and phase curves for Ex polarization.

A comparative analysis of the three tested thresholds indicates that selecting an appropriate Mahalanobis distance value plays a critical role in both processing efficiency and data quality. A very low threshold (MD = 1.5) removes a large number of data windows, which in some cases reduces the natural variation of apparent resistivity across frequencies. In contrast, a high threshold (MD = 5) allows noisy and outlier-contaminated windows to remain in the dataset, leading to out-of-range values, increased error bars, and deviations from the expected smooth resistivity trend.

The intermediate threshold (MD = 3) provides a more balanced result. Although the associated error bars are slightly higher than those for MD = 1.5, the apparent resistivity and phase curves show consistent and smooth variations while preserving data integrity. Moreover, choosing an optimal threshold reduces processing time, which is particularly beneficial for long-duration and low-frequency MT surveys. Overall, the

results demonstrate that MD = 3 offers a suitable compromise between noise suppression, computational efficiency, and geological reliability of the final impedance estimates.

5. Conclusions

This study presented a comprehensive overview of magnetotelluric data processing, from initial time-series windowing and spectral transformation to final impedance estimation. Within this framework, the critical challenge of noise and outlier removal was addressed through the application of the Mahalanobis distance method.

The four-dimensional Mahalanobis distance — simultaneously incorporating the real and imaginary parts of two impedance components (Z_{xx} and Z_{xy}) — demonstrates clear advantages over conventional two-dimensional approaches. By forming a multivariate feature space, it more effectively captures inter-component correlations and improves the detection and rejection of outliers, leading to more stable and reliable impedance estimates.

The proposed method was tested on field data from two stations, Chassenon Forage and Forêt Beaulieu (Rochechouart impact structure, France), using three distinct thresholds: MD = 1.5, 3, and 5. The results from both stations confirm that:

- The four-dimensional implementation provides more reliable outlier detection than conventional two-dimensional approaches, reducing noise contamination while preserving the essential geophysical signal.
- Threshold selection plays a critical role in balancing data integrity and processing efficiency. A low threshold (MD = 1.5) removes a large number of data windows, which may limit natural variations in the apparent resistivity curves. A high threshold (MD = 5) retains most windows but allows noisy or outlier-contaminated estimates to influence the final resistivity and phase responses, sometimes leading to larger error bars and less stable trends.
- An intermediate threshold (MD = 3) provides the most balanced performance for both stations, yielding smooth and geophysically consistent resistivity and phase curves with reasonable processing time.

Overall, the four-dimensional Mahalanobis distance method improves the robustness and efficiency of MT data processing by enhancing multivariate outlier rejection in impedance space. This leads to more reliable apparent resistivity and phase estimates, which are essential for accurate subsurface interpretation. In addition, the reduction in processing time makes the method particularly suitable for long-duration and low-frequency MT datasets.

Future work will involve applying this approach to additional MT profiles from the Rochechouart dataset, including sites located near power lines where anthropogenic noise levels are higher. Such data will provide a valuable opportunity to further evaluate the performance of the four-dimensional Mahalanobis distance method under more challenging noise conditions.

References

Ajithabh, K. S., & Patro, P. K. (2023). SigMT: An open-source Python package for magnetotelluric data processing. *Computers & Geosciences*, 171, 105270. https://doi.org/10.1016/j.cageo.2022.105270

Borah, U. K., Patro, P. K., & Suresh, V. (2015). Processing of noisy magnetotelluric time series from Koyna-Warna seismic region, India: A systematic approach. *Annales Geophysicae*, 58(2), G0222. https://doi.org/10.4401/ag-6690

Cagniard, L. (1953). Basic theory of the magneto-telluric method of geophysical prospecting. *Geophysics*, 18(3), 605-635. https://doi.org/10.1190/1.1437915

Chave, A. D., Thomson, D. J., & Ander, M. E. (1987). On the robust estimation of power spectra, coherences, and transfer functions. *Journal of Geophysical Research: Solid Earth*, 92(B1), 633–648. https://doi.org/10.1029/JB092iB01p00633

Daud, Y. (2010). Electromagnetic method: Success story in geothermal exploration & possibility for hydrocarbon exploration. College Textbook, Depok.

Egbert, G. D. (1997). Robust multiple-station magnetotelluric data processing. *Geophysical Journal International*, 130(2), 475–496. https://doi.org/10.1111/j.1365-246X.1997.tb05661.x

Egbert, G. D., & Booker, J. R. (1986). Robust estimation of geomagnetic transfer functions. *Geophysical Journal International*, 87(1), 173–194. https://doi.org/10.1111/j.1365-246X.1986.tb04552.x

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia.

Fowler, R., Kotick, B., & Elliot, R. (1967). Polarization analysis of naturally and artificially geomagnetic micropulsations. *Journal of Geophysical Research*, 72, 2871–2883. https://doi.org/10.1029/JZ072i011p02871

Gamble, T. D., Goubau, W. M., & Clarke, J. (1979). Magnetotellurics with a remote magnetic reference. *Geophysics*, 44(1), 53–68. https://doi.org/10.1190/1.1440923

Hermance, J. F. (1973). Processing of magnetotelluric data. https://doi.org/10.1016/0031-9201(73)90060-5

Jenkins, G. M. (1968). Spectral analysis and its applications. Holden-Day, Inc., San Francisco.

Jones, A., &Jodicke, H. (1984). Magnetotelluric transfer function estimation improvement by a coherence-based rejection technique. SEG Technical Program Expanded Abstracts. https://doi.org/10.1190/1.1894081

Kato, Y., & Kikuchi, T. (1950). Scientific Reports of Tohoku University, Series 5, Geophysics, 2, 139.

Keller, G. V., & Frischknecht, F. C. (1966). Electrical methods of geophysical prospecting. Pergamon Press, New York, NY.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India, 2(1), 49–55.

Platz, A., &Weckmann, U. (2019). An automated new pre-selection tool for noisy magnetotelluric data using the Mahalanobis distance and magnetic field constraints. *Geophysical Journal International*, 218(3), 1853–1872. https://doi.org/10.1093/gji/ggz197

Price, A. T. (1950). Quarterly Journal of Mechanics and Applied Mathematics, 3, 385.

Quesnel, Y., Sailhac, P., Lambert, P., Lambert, A., & Dyment, J. (2021). Multiscale geoelectrical properties of the Rochechouart impact structure, France. Geochemistry, Geophysics, Geosystems, 22(12), e2021GC010036. https://doi.org/10.1029/2021GC010036

Rikitake, T. (1946). Bulletin of the Earthquake Research Institute, Tokyo University, 24, 1.

Ritter, O., Junge, A., & Dawes, G. J. (1998). New equipment and processing for magnetotelluric remote reference observations. *Geophysical Journal International*, 132(3), 535–548. https://doi.org/10.1046/j.1365-246X.1998.00440.x

Simpson, F., & Bahr, K. (2005). Practical magnetotellurics. Cambridge University Press. https://doi.org/10.1017/CBO9780511614095

Smirnov, M. Yu. (2003). Magnetotelluric data processing with a robust statistical procedure having a high breakdown point. *Geophysical Journal International*, 152(1), 1–7. https://doi.org/10.1046/j.1365-246X.2003.01733.x

Swift, C. M. (1967). A magnetotelluric investigation of an electrical conductivity anomaly in the southwestern United States. Ph.D. Thesis, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts.

Tikhonov, A. N. (1950). DokladyAkademiiNauk S.S.S.R., 73, 295.

Vozoff, K. (1972). The magnetotelluric method in the exploration of sedimentary basins. *Geophysics*, 37, 98–141. https://doi.org/10.1190/1.1440228

Weckmann, U., Ritter, O., & Haak, V. (2005). Effective noise separation for magnetotelluric single site data processing using a frequency domain selection scheme. *Geophysical Journal International*, 161(3), 635–652. https://doi.org/10.1111/j.1365-246X.2005.02608.x