Interval Approach to Magnetotelluristic Data Processing

A A Lavrukhin¹, A S Tukanova¹
¹Omsk State Transport University, 35 Marksa pr., Omsk, 644046, Russia

E-mail: lavruhinaa@gmail.com

Abstract. The article presents a new approach to estimate the frequency characteristics of the impedance tensor for processing magnetotelluristic data. The approach is based on the applying of interval analysis methods when solving a system of linear equations. As a reference method, to compare with, a combined robust algorithm is used (with discarding data by the coherence criterion, median estimating, and weighting least squares method). This algorithm is compared with the results of the proposed interval computational algorithm that is based on the method of J. Rohn, implemented in the intvalpy Python library. Computational experiments on the data processing were performed using natural magnetotelluristic field data. The interval approach can be successfully applied to the processing of magnetotelluristic data.

1. Introduction

The magnetotelluristic (MT) method refers to the technologies of deep electrical prospecting and is based on the study of natural electromagnetic fields with varying periods of oscillations. The method includes registration, processing, analysis, inversion, and interpretation of variations in the electromagnetic field. The implementation of the MT is in studying the internal structure of the Earth, platform cover, upper mantle, electrical conduction zones, exploration for oil and ore deposits, geothermal sources [1, 2].

The components of the electromagnetic field are recorded using grounded electrical lines and induction sensors for some time—from 10 minutes to 5 days (depending on the modification of the method, frequency range and investigated depths). Measurements of the strengths of the electric and magnetic fields are carried out at frequencies from $10^{-4}$ to $10^{4}$ Hz and make it possible to study the geoelectric structure of the Earth at depths from 10 meters to 100 kilometers. After completing the procedure for registering the variations of the natural electromagnetic field, a set of records of signals of the electric and magnetic fields strengths $E_x(t), E_y(t), H_x(t), H_y(t)$ is obtained. The data processing carried out during the registration process consists of recalculating the values of the electric and magnetic fields strengths, measured at different periods (frequencies), into the functions of the apparent resistance, which are subsequently used in the construction of geoelectric sections. This procedure can be performed during field or office work [2].

A condition for the successful interpretation of the recorded magnetotelluristic data is an accurate and stable estimation of the impedance tensor components. The impedance tensor is a magnetotelluristic response connecting the electromagnetic field components in the frequency domain, performed at the data processing stage [3].

When processing MT data, various methods based on the provisions of the modern theory of probability and mathematical statistics are widely used [4]. The problem with these methods is that the nature of noise and interference in MT data often leads to violation of accepted distribution law.
Various robust approaches to overcome this make the partial exclusion of part of the registered data or reducing their influence on the resulting estimate [5-9]. Interval analysis methods can open new possibilities in obtaining impedance estimates [10, 11]. Interval methods seem to be an effective means of processing data obtained during field measurements and can be used in combination with robust algorithms.

This work aims to consider and investigate the applicability of the interval approach to the processing of magnetotelluric data. To carry out numerical experimental studies, the authors used nature field geophysical data.

2. Problem statement

To obtain impedance tensor estimates, the Tikhonov-Cagniard identification model is used, which connects the electric and magnetic components of the MT field and the unknown components $Z_{xx}, Z_{xy}, Z_{yx}, Z_{yy}$ of the impedance tensor $Z$ for a certain frequency value $\omega$ by linear relations [1]:

\[
H_x Z_{xx} + H_y Z_{xy} = E_x \quad (1)
\]
\[
H_x Z_{yx} + H_y Z_{yy} = E_y \quad (2)
\]

where $E_x, E_y, H_x, H_y$ are the complex spectral values of the MT field components at frequency $\omega$, considered in a certain time window (the set of values $E_x, E_y, H_x, H_y$ found for the time window will be called one independent measurement).

Obtained values of the components of the impedance tensor lead to the calculation of apparent resistances and the phase characteristics (so-called amplitude and phase curves of magnetotelluric sounding). Transformation and further inversion of the obtained frequency curves make it possible to build a model of the geological-geophysical section at the sounding site.

The problem of processing MT data can be formulated as follows: according to the recorded data $E$ and $H$, which are related to each other by the model (1)—(2), it is necessary to find the optimal estimates of the components of the impedance tensor $Z$, having the minimum bias and minimum variance, or to find the numerical intervals, within the found values of the $Z$ components.

Estimates of the components of the impedance tensor $Z$ can be found on the basis of one of the following approaches:

- in the time domain based on the solution of the equations of convolution of the registered MT field signals and impulse response [12];
- in the frequency domain based on the solution of a system of linear algebraic equations written for the spectral values of the MT field signals and unknown values of the impedance tensor components, followed by the calculation of confidence intervals [4];
- in the frequency domain based on the solution of the interval system of linear algebraic equations written for the intervals of spectral values of MT field signals and unknown intervals of the impedance tensor components (this approach has no description in the sources).

The procedure for obtaining impedance estimates in the frequency domain can be represented as a sequence of the following steps [4, 7, 13, 14]:

- obtaining spectral estimates—the transition from records of MT field time series to their spectral densities;
- rejection of data (by the criterion of partial or multiple coherence or by the jackknife method)—search and exclusion of spectral estimates, the presence of which can significantly worsen the results of processing and interpretation;
- obtaining impedance estimates—calculating point estimates or interval estimates of the components of the impedance tensor based on the solution of a system of linear equations or an interval system of linear equations (1)—(2), written for those spectral densities that have not been rejected;
• evaluation of the quality of the obtained solutions—calculation and analysis of the numerical characteristics of the obtained impedance estimates.

There are several approaches to obtaining spectral estimates. These methods are Fourier transform, wavelet transform, complex demodulation with filtering. These methods represent the operation of convolution of electromagnetic field signals and various taper functions (Hann, Hamming, Blackman, Kaiser, Bartlett) [9, 15], wavelet functions (Haar, Shannon-Kotelnikov, Morlet, Daubechies, Meyer, Battle-Lemarie) [8, 16], or impulse response of a narrow-band filter.

The criterion for rejecting unsatisfactory data is the level of coherence. The partial coherence coefficient [17] (the coherence $Coh_{ExHy}^2$ is determined similarly) is:

$$Coh_{ExHy}^2 = \left| E_x H_y' \right|^2 \left[ (E_x E_x') (H_y H_y') \right]^{-1}$$

(3)

Coherence coefficients show how the spectral densities of the components of the electromagnetic field $E_x$ and $H_y$; $E_x$ and $H_y$ are related. If the value of the coherence coefficients significantly differs from 1, then spectral estimates are rejected (estimates corresponding to highly contaminated experimental data).

The matrix form for the system of equations (1) and (2), where each row is an independent measurement, is

$$\begin{bmatrix}
H_1 \\
H_2 \\
\vdots \\
H_N
\end{bmatrix}
\cdot Z =
\begin{bmatrix}
E_1 \\
E_2 \\
\vdots \\
E_N
\end{bmatrix}$$

(4)

or abbreviated

$$H \cdot Z = E$$

(5)

The identification model (8) for each independent measurement (time interval) is a system of linear algebraic equations or an interval system of linear algebraic equations written concerning the point values or interval values of the impedance tensor components. Obtaining point and interval impedance estimates consists of solving a redundant system of equations for $n$ unknown values or interval values of the $Z$ components.

There are several approaches to obtaining point estimates of impedance. For this purpose, one can use various methods, which include

- methods for solving systems of linear equations: Gauss method, Gauss-Seidel method, Newton method, simple iteration method, etc. [18];
- methods based on minimization/maximization of various functionals (residuals/errors, losses, weight, likelihood): least-squares method; generalized least squares method; weighted least squares method using different weight functions or their arithmetic/geometric/quadratic mean. Weight functions can be calculated based on residuals (functions of Huber, Thomson, Tukey, Rousseeuw, Merrill-Schwepp, etc.), diagonal elements of the projection matrix, derived from the logarithmic likelihood function. [9, 19, 20, 21, 22, 23, 24];
- median methods (median least squares, the methods of Siegel, Andrews, Theil-Sen, Hodges-Lehmann, Simon, Tukey, etc.) [21, 22];
- two-step and three-step methods based on combinations of different assessment methods [25, 4].

There are also several approaches to obtaining interval impedance estimates. Various methods for this purpose include the interval Gauss method, interval Gauss-Seidel method, interval Newton method, interval Jacobi method, Krawczyk method, Moore method, Rohn method, Rump method, Hansen-Blick-Rohn method, and others [8].
This article sets the tasks:
• get an interval form for a system of linear algebraic equations;
• propose an algorithm for the reference point solution;
• propose an algorithm for the interval solution;
• investigate the applicability of the interval approach and the rule for choosing the width of the intervals;
• compare the results of the interval approach with one of the robust methods.

3. Theoretical part
We propose the following combined robust algorithm for finding point estimates of the components of the impedance tensor, which includes two stages of estimation:
• at the first stage, the estimates of $Z_1$ should be calculated from the spectral densities of the electromagnetic field signals that have passed the rejection by the criterion of partial coherence, using the median method;
• at the second stage, the estimates of $Z_2$ should be calculated using the weighted least squares method according to the estimates of $Z_1$ obtained in the first stage.

The median estimate with a cutoff point of 50 % can be calculated as follows (the formula is given only for $Z_{xy}$) [22, 26, 27]:

$$Z_{xy} = \text{med}_{k_1=1,\ldots,N_1} \text{med}_{k_2=1,\ldots,N_2} \frac{(H_x \cdot H_y^*)(E_y \cdot H_y^*) - (H_y \cdot H_y^*)(E_y \cdot H_x^*)}{(H_x \cdot H_y^*)(H_y \cdot H_y^*)} = \text{med}_{k_1=1,\ldots,N_1} \text{med}_{k_2=1,\ldots,N_2} Z_{0,k_1,k_2}$$

where the spectral values $E_x, E_y, H_x, H_y$ are averaged over any two independent measurements $k_1$ and $k_2$.

Estimates of the impedance tensor components obtained using the weighted least squares method can be represented in matrix form as follows [4, 25]:

$$Z_2 = (w \cdot E \cdot H^*)^{-1}$$

where Huber weight function [4] is:

$$w = \begin{cases} 
1, & |r| \leq r_0; \\
\frac{r_0}{|r|}, & |r| > r_0 
\end{cases}$$

and values of $r$ represent the residuals:

$$r = (Z_0, k_1, k_2 - Z_1) d^{-1}$$

Expressions (8) and (9) include the parameter of the weighting function $r_0 = 1.345$ and the scale parameter calculated based on the absolute median deviation [20] is:

$$d = S_{MAD} = 1.4826 \cdot \text{med}|r - \text{med} r|$$

The confidence intervals for the estimates of the components of the impedance tensor $Z$ are

$$[Z; Z] = [Z_{\text{mean}} - 1.96S_{MAD} r^{-0.5}; Z_{\text{mean}} + 1.96S_{MAD} r^{-0.5}]$$

where $Z_{\text{mean}} = n^{-1} \sum_{i=1}^{n} Z_{ii}$ is the sample mean; $n$ is the sample size; $S_{MAD}$ is the absolute median deviation; $\Phi_0 = 1.96$ is the value of the Laplace function with the reliability of 0.95.

To search for interval estimates, we represent model (4) in interval form [10] (using the notation accepted in interval analysis):
where the parameters $h$ and $e$ define the boundaries of the intervals with the central point values $E$ and $H$.

There are two problems with the interval approach:

- numerical methods for solving systems of linear interval equations currently work only with real numbers;
- it is necessary to define a rule for setting the boundaries of the intervals of the coefficients of the system of equations (parameters $h$ and $e$).

To solve the first problem, we pass from complex numbers to real and imaginary ones, rewriting equation (1) in the form of equations for real and imaginary parts (equation (2) can be written by analogy). Magnetic field spectral density matrix is:

$$
\begin{bmatrix}
H_1 - h_1, H_1 + h_1 \\
H_2 - h_2, H_2 + h_2 \\
... \\
H_N - h_N, H_N + h_N
\end{bmatrix} \cdot Z = 
\begin{bmatrix}
E_1 - e_1, E_1 + e_1 \\
E_2 - e_2, E_2 + e_2 \\
... \\
E_N - e_N, E_N + e_N
\end{bmatrix}
$$

(12)

where the parameters $h$ and $e$ define the boundaries of the intervals with the central point values $E$ and $H$.

The vector of spectral densities of the electric field is:

$$
E' = \begin{bmatrix} \text{Re}E_x & \text{Im}E_x \end{bmatrix}^T
$$

(14)

The vector of the estimated components of the impedance tensor is:

$$
Z_n = \begin{bmatrix} \text{Re}Z_{xx} & \text{Im}Z_{xx} & \text{Re}Z_{xy} & \text{Im}Z_{xy} \end{bmatrix}^T
$$

(15)

For simplicity, we consider only the right-hand sides of equation (12) in interval form. This statement can be acceptable, because in the common cases the electric components of the field are more susceptible to the influence of noise, and the magnetic ones can be cleaned of noise, for example, using the remote base method. The intervals included in the left side of (12) degenerate into point values, and the vector represented the intervals that are included in the right side of (12) is:

$$
\begin{bmatrix}
[-e_1 + e_1] \\
[-e_2 + e_2] \\
... \\
[-e_N + e_N]
\end{bmatrix}^T = \begin{bmatrix} r_1 & r_2 & ... & r_N \end{bmatrix}^T
$$

(16)

where $r_1$, $r_2$, ..., $r_N$ are the interval numbers that have a zero center and their radius is equal to the corresponding values of the residuals $r_1, r_2, ..., r_N$.

Thus, by analogy with the form (4), the redundant system of equations (1) is written in matrix form for $N$ independent measurements, and interval additions (the interval vector of residuals) are introduced for it:

$$
\begin{bmatrix}
H_{r1} \\
H_{r2} \\
... \\
H_{rN}
\end{bmatrix} \cdot Z_n = 
\begin{bmatrix}
E_{r1} \\
E_{r2} \\
... \\
E_{rN}
\end{bmatrix} + \beta
$$

(17)

where $\beta$ is parameter for regularization.

We propose the following algorithm for finding interval estimates of the components of the impedance tensor, which includes two stages of estimation:
• at the first stage, the estimates of $Z_1$ should be calculated using the combined robust method, from the spectral densities of signals of the electromagnetic field, which have been rejected according to the criterion of partial coherence;
• at the second stage, the estimates of $Z_2$ should be calculated, using the Rohn interval method [28, 29] (with the width of the intervals corresponding to the value of the residuals obtained based on the estimates of $Z_1$).

To compare the accuracy of the obtained solutions, we will use the coefficient $\alpha$, which is equal to the ratio of the width of the interval solution to the width of the confidence interval:

$$\alpha = \frac{\Delta Z_{\text{interval}}}{\Delta Z_{\text{conf}}}$$

(18)

4. Results of computational tests and discussion
The authors investigated the algorithms proposed for consideration and analysis using the signals obtained in the Tevrizovsky district of the Omsk region ($57°\ 32'\ 3.84''\ N;\ 72°\ 19'\ 12.65''\ E$). The point estimates with appropriate confidence intervals of the component $Z_{xy}$ of impedance tensor using the first algorithm are shown in figure 1. The interval estimates of the component $Z_{xy}$ of impedance tensor using the second algorithm are shown in figure 2. To compare the joint arrangement of point and interval solutions, the impedance characteristics are combined and shown in figure 3. The ratios of the width of the interval estimates to the width of the confidence intervals, plotted for the entire frequency range, are shown in figure 4.

To obtain interval solutions, the authors used the intvalpy library for Python, developed by S. P. Shary and A. S. Androsov.

The resulting point solution can be considered quite satisfactory and, after smoothing it with a spline (as is customary in the practice of data processing), it can be used in the further construction of geo-sections. Nevertheless, the confidence intervals show high accuracy only in the ranges of periods from $2\cdot10^{-2}$ s to $7\cdot10^{-1}$ s and from $2\cdot10^{1}$ s to $2\cdot10^{2}$ s. In the dead band, in which the signal-to-noise ratio is too small (from $1\cdot10^{0}$ s to $2\cdot10^{1}$ s), the accuracy of the estimates obtained is significantly reduced. In the low frequency region (with periods of more than $2\cdot10^{1}$ s), the accuracy of estimates also decreases due to a decrease in the amount of accumulated data.

The interval solution is generally similar to the point solution. At the same time, the ranges of solution accuracy turned out to be wider: from $1\cdot10^{-2}$ s to $7\cdot10^{-1}$ s and from $1\cdot10^{1}$ s to $3\cdot10^{2}$ s, in the dead band, there is also a widespread. A characteristic feature can be considered insufficient stability of solutions in the dead band, as evidenced by the significant mutual discrepancy of the intervals obtained at close frequencies (at periods from $7\cdot10^{-1}$ s to $5\cdot10^{0}$ s). The interval algorithm also shows insufficient insensitivity to correlated noise (values at periods of $9.5\cdot10^{-2}$ s, $2\cdot10^{1}$ s, $4.5\cdot10^{1}$ s), while the robust algorithm gave unbiased solutions at these periods.

![Figure 1. Point estimates of impedance $Z_{xy}$ with confidence intervals.](image-url)
If we exclude the dead band from consideration, almost everywhere the relative accuracy of the interval solution $\alpha$ is from 0.3 to 1. The regularization coefficient $\beta$ was set equal to 0.25 for high and
middle frequencies and 0.5 for low frequencies. Changing the coefficient $\beta$ allows you to adjust the relative accuracy of the solutions obtained.

5. Conclusion
Interval analysis can be a rather useful addition for solving the problem of estimating MT impedance, apparent resistivities, and geoelectric sections. Unlike classical statistical methods, which operate separately with point estimates and their confidence intervals, it is possible to obtain interval solutions directly. The interval approach can be successfully applied to the processing of MT data. With a sufficiently high signal-to-noise ratio, interval solutions can be more accurate. With strong data noise, the interval solutions may not be completely satisfactory and unstable, and to solve this problem, further development of the interval approach is necessary.

6. References
[1] Berdichevsky M N and Dmitriev V I 2009 Models and Methods of Magnetotellurics (Berlin: Springer) p 563
[2] Zhdanov M S 2009 Geophysical Electromagnetic Theory and Methods (Salt Lake City: Elsevier) p 868
[3] Berdichevsky M N, Dmitriev V I, Novikov D B and Pastutsan V V 1997 Analysis and Interpretation of Magnetotelluric Data (Moscow: Dialogue-MSU) p 161
[4] Chave A D and Jones A G 2012 The Magnetotelluric Method: Theory and Practice (New York: Cambridge University Press) p 603
[5] Larsen J C, Randall L M, Manzella A, Fiordelisi A and Rieven S 1996 Geoph. J. Int. 124 801
[6] Müller A 2000 Geoph. J. Int. 142 257
[7] Smirnov M Yu 2003 Geoph. J. Int. 152 1
[8] Garcia X and Jones A G 2008 Geoph. 73 223
[9] Chave A D 2017 Surv. Geophys. 38 837
[10] Shary S P and Sharaya I A 2013 J. Comp. Techn. 18 (3) 80
[11] Shary S P 2012 Automation and Remote Control 73 (2) 310
[12] Nowozynski K 2004 Geophys. J. Int. 158 32
[13] Electrical prospecting vol 1 2018 ed I N Modin, A G Yakovlev (Tver: Polipress) p 274
[14] Vagin S A 2009 Geophysical issues 42 100
[15] Oppenheim A and Schafer R 2010 Discrete-Time Signal Processing (Pearson)
[16] Percival D B and Walden A T 2006 Wavelet Methods for Time Series Analysis (Cambridge University Press)
[17] Bendat J S and Piersol A G 2011 Random data: analysis and measurement procedures (New York, Wiley) p 486
[18] Mathews J H and Fink K K 2004 Numerical Methods Using Matlab (Pearson)
[19] Huber P J 2004 Robust statistics (New York, John Wiley and Sons) p 308
[20] Hampel F R, Ronchetti E M, Rousseuw P J and Stahel W A 1986 Robust statistics: the approach based on influence functions (New York, John Wiley and Sons) p 468
[21] Kobzar A I 2006 Applied Mathematical Statistics (Moscow: Fizmatlit) p 816
[22] Rousseuw P J and Leroy A M 1987 Robust regression and outlier detection (New York: Wiley) p 347
[23] Chave A D, Thomson D J and Ander M E 1987 J. of Geoph. Research 92 633
[24] Egbert G D and Booker J R 1986 Geophys. J. R. 87 173
[25] Silvestrov A N and Chinaev P I 1987 Identification and Optimization of Automatic Systems (Moscow: Energoatomizdat) p 200
[26] Sims W E, Bostick F X and Smith H W 1971 Geoph. 36 938
[27] Siegel A F 1982 Biometrika 69 242
[28] Rohn J 1996 Reliable Computing 2 167
[29] Shary S P 2017 J. Comp. Techn. 22 (2) 150