Search for the optimal method for obtaining estimations for an automated magnetotelluric signal processing system

A A Lavrukhin¹, K V Lobov², A S Tukanova¹ and A S Onufriev¹

¹Omsk State Transport University, Omsk, Russia
²Omsk State Technical University, 11 Mira ave., Omsk, 644050, Russia

Abstract. The problem of obtaining estimates of impedance frequency characteristics in the magnetotelluric sounding method is considered. A variational algorithm is proposed for research through a variety of means of estimation from the point of view their accuracy and efficiency. The results of computational experiments on the processing of synthetic signals are presented, which represent indicators of the accuracy and efficiency of some estimation methods. The proposed approach can be used in the development of algorithms and software for the embedded system for recording and processing signals, as well as applied software for an automated system for processing signals from magnetotelluric sounding method.

1. Introduction

The class of problems in which it is necessary to obtain an estimated value (as well as the characteristics of this estimate) of a certain parameter of a process, phenomenon, or object under research are various kinds of inverse problems. These include, in particular, the problem of establishing an estimate of the impedance frequency response and confidence intervals of this characteristic in the magnetotelluric sounding method which is one of the most effective geoelectric methods (it has the best ratio of financial, logistic, ecological and other costs to the information result obtained). Magnetotelluric sounding (MT sounding) has practical application in high-resolution profiled and areal surveys of mineral deposits, in local and regional prospecting geological exploration works and for solving other geological problems. The depth range covered by this method ranges from several meters to hundreds of kilometers. The specificity of the MT sounding method consists in the signal recording of variations in the natural electromagnetic field (records of orthogonal electric and magnetic components through the instrumentality of special-purpose hardware—recording device), further processing of the received signals and interpretation of these signals in order to obtain the parameters of the geoelectric structure model [1].

Over the past 70 years, since the discovery of the MT sounding method by A. N. Tikhonov and through the further development, provided by L. Kanyar, M. N. Berdichevsky, V. I. Dmitriev and a number of other scientists, the method has been significantly improved and brought to a high-tech state. Today, there are impressive hardware and software, both in terms of capabilities and cost, which provide an established procedure of work with a fairly guaranteed result, subject to certain conditions. Nevertheless, there are unsolved problems associated with both the creation of modern hardware and software systems for MT sounding, and with the improvement of the methodology aimed at increasing the efficiency of the MT method, which will allow it to be applied at a new level, to work with very large amounts of data and perform flexible planning of the total operation cycle of work.
The problem introduced in this article rather of a scientific and applied nature than just a scientific one, since the result of the investigation should be used in the development of the latest hardware and software, both built-in and applied, which will form an automated system for carrying out a full cycle of works by the method of MT sounding.

2. Statement of the problem

According to the Tikhonov-Kanyar model, which is the basis of the MT method, the horizontal components of the electric and magnetic fields at the observation point are related by the following dependence [2]:

\[
\begin{align*}
E_x(\omega) &= Z_{xx}(\omega) \cdot H_x(\omega) + Z_{xy}(\omega) \cdot H_y(\omega); \\
E_y(\omega) &= Z_{yx}(\omega) \cdot H_x(\omega) + Z_{yy}(\omega) \cdot H_y(\omega),
\end{align*}
\]

(1)

where \(\omega\) is the frequency; \(E_x, E_y, H_x, H_y\) are the spectral estimates of the electric \((E)\) and magnetic \((H)\) fields components as functions of frequency, found over a certain time interval; \(Z_{xx}, Z_{yx}, Z_{xy}, Z_{yy}\) are the components of the main response function—the impedance tensor as a function of frequency. The impedance tensor looks like this:

\[
Z(\omega) = \begin{bmatrix}
Z_{xx}(\omega) & Z_{xy}(\omega) \\
Z_{yx}(\omega) & Z_{yy}(\omega)
\end{bmatrix}.
\]

(2)

Estimation of the impedance tensor (2) plays a key role in the method and further allows determining subsequent characteristics (for example, apparent resistivity) and proceeding to the obtaining of geoelectric sections. The importance of the intermediate estimation of the impedance tensor lies in the fact that this stage in the chain of all calculations from the initial signals to the resulting geoelectric model, on the one hand, indicates the quality of the field work (the specificity of which is that MT signals are recorded passively and can have different spectral power density, possibly in the presence of strong, including industrial noise, which means, at the best case, the operators will not introduce additional noise and complete the registration process for a sufficient time to accumulate the necessary information), and on the other hand, it can be obtained already in the field conditions by means of embedded computing facilities. The task is to obtain an accurate and stable (robust) estimate of the transfer function (2) connecting the components of the MT field in the frequency domain.

The general goal of the research consist in to obtain two algorithms: a fast algorithm that can be applied in the software built into the recorder and an accurate algorithm that can be used in the final processing of data. In the first case, the limitation is imposed by the autonomy of the recorder, and therefore by a rigid power limit.

For the achievement of this goal, it is necessary to carry out a comparative study of the estimation accuracy of the impedance tensor components by the use different types of estimates and different options of the algorithm. In the present article, such a variational algorithm will be constructed and its main metrics will be selected, which will allow at a later date to build an algorithm that is optimal in terms of computational costs with acceptable accuracy and an algorithm that is optimal in accuracy for using them in an automated system for recording and signals processing in the MT sounding method.

Currently, a number of methods have been proposed that allow determining the components of the transfer function (2) with varying accuracy: the classical least squares method and its variations (methods of the median least-squares, functional least-squares, truncated least squares, etc.), the search for minima and maximum differently robust functions (for example, loss, weight, likelihood functions), calculation of estimates L, S, W and other types [3], as well as the remote base method. The development of algorithms for estimating components based on the listed methods is the subject of the scientific papers of M. Smirnov [4], S. Vagin [5], D. Larsen [6], A. Chave [7], G. Egbert [8], A. Jones [3], T. Gamble [9] and other researchers.
When developing methods and data processing algorithm in the MT sounding problem, you can work with two types of data arrays: with actual data and synthetic. The processing of actual data makes it possible to carry out comparative tests of various hardware and software complex, to compare the effectiveness of the MT sounding method with other types of exploration. When processing actual data, it is imperative that some reference estimates obtained by other methods, already tested or giving a repeatable result under a variety of approaches, are required, since the crucial problem is the inability to directly observe and obtain the true parameters of geoelectric sections (such information can only be provided by expensive drilling). Synthetic data allow us to get away from this problem and transparently investigate the algorithm for solving the inverse problem by comparing the solution result with a previously known reference object. In geophysics, the most famous such projects are COMDAT [10] and SimPEG [11].

To research the methods, a general variational algorithm was developed, the diagram of which is shown in figure 1. It includes particular algorithms, signed with numbers, which can be used either individually or in the order of their sequence from top to bottom:

**Figure 1. General variable algorithm for obtaining estimates of Z and their characteristics.**
• (1) Algorithm based on the classical least squares method (CLS method) in two arbitrary time intervals.
• (2) Algorithm based on the CLS method over the entire set of time intervals.
• (3) Algorithm based on the median least squares method.
• (4) Algorithm based on Siegel's median estimate.
• (5) Algorithm based on the median Siegel estimate and on the W-estimate.
• (6) Algorithm based on the weighted least squares method (WLS method) using various weight functions.
• (7) Approximation of the estimate by an artificial neural network.

For a comparative analysis of the results obtained, we will use several numerical indicators: the mean normalized bias of an estimate, the normalized standard deviation, the mean normalized confidence interval [12]. Normalization when averaging the indicators is required due to the fact that the estimate of the impedance characteristic along the frequency axis changes up to several orders of magnitude. On the basis of these indicators, in the future, conclusions can be drawn about the effectiveness of the estimation of a particular method. The estimation efficiency together with the computational efficiency will make it possible to make a separate choice of method for a complete computational algorithm, as well as for a fast simplified one.

While using synthetic signals, it is possible to compare the results obtained from noisy data with the results from clear data, which makes it possible to determine the bias of the estimates obtained.

3. The theory
This paper presents the processing of MT data, which is limited to obtaining estimates of the transfer functions connecting the components of the MT field in the frequency-domain (impedance tensor). It is proposed to apply and compare several methods for estimating the components of the impedance tensor.

To put the case that at each frequency there are \( N \) time intervals for recording the MT field components and the corresponding spectral estimates \( E_x, E_y, H_x, H_y \), obtained using various methods of spectral analysis (Fourier transform, wavelet transform, complex demodulation) at each frequency. Estimates of the impedance tensor components are calculated based on selected spectral densities (by coherence).

A simple method for finding estimates of the impedance tensor components is the least squares (LS) method, based on minimizing the sum of the squared residuals between the left and right sides of the equations of system (2), then such a solution can be written in matrix form [13]:

\[
Z_{LS} = \left( EH^* \right)^{-1}
\]

or by one of the formulas that are used in MTs [14]:

\[
Z_{MT} = \frac{(H_x H_x^*) (E_x H_y^*) - (H_x H_y^*) (E_x H_x^*)}{(H_x H_x^*) (H_y H_y^*) - (H_x H_y^*) (H_y H_x^*)}.
\]

Least squares method is sensitive to spikes and noise present in MT signals. This problem can be solved by applying various robust methods that reduce the deviation of estimates from the true values of parameters. The most flexible is the maximum-likelihood method (M-estimates method), which consists in minimizing the sum of less rapidly growing residual functions \( r = [E] - [Z][H] \) [15]:

\[
\sum_{i=1}^{N} \chi(r_i) \rightarrow \min,
\]
where $\chi(r)$ is the loss function (the function of the discrepancy between the true value of the parameter and its estimate) [3]:

$$\chi(r) = \begin{cases} 
\frac{r^2}{2}, & |r| \leq r_0; \\
|r| - \frac{r^2}{2}, & |r| > r_0.
\end{cases}$$

To solve problem (5), the influence function $\psi(r)$ (influence function of various noises and interference on the estimate) [16] is introduced, which is the derivative of the loss function $\psi(r) = \frac{d\chi(r)}{dr}$:

$$\psi(r) = \begin{cases} 
r, & |r| < r_0; \\
n_0 \cdot \text{sign}(r), & |r| \geq r_0,
\end{cases}$$

and the search for its root is performed, for example, by iterative numerical methods, such as Newton's method, where the result obtained by formula (3) can be considered as an initial approximation.

An equivalent form of search [15] is the introduction of the Huber and Thompson weight functions with the parameters $r_0 = 1.5$ and $\alpha = 0.8$, respectively [16]:

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

and

$$w(r) = \exp\left(-\exp\left(\alpha(|r| - \alpha)\right)\right).$$

Given that the median is the most robust estimate, consider two median methods with a maximum cutoff point $\epsilon = 50\%$ (the data pollution percentile). The least squares median method (LSM method) is to minimize the medians of the sum of squares of the residuals [16]:

$$\text{med} \sum_{i=1}^{N} r_i^2 \rightarrow \min.$$

The repeating median method proposed by Siegel can be represented as follows formula [17]:

$$Z_S = \text{med}\left(\text{med}(Z_0)\right).$$

The above estimates are not automatically invariant (equivariant) with respect to scale, which in most cases acts as an interfering factor. In this case, the estimates of the impedance tensor components are accompanied by the calculation of additional estimates and the procedure for normalizing the residuals $r/d$. It is most expedient to determine the estimate of the scale parameter through the absolute median deviation [16]:

$$d = S_{MAD} = 1.483 \cdot \text{med}|r|.$$

The efficiency of various estimation methods can be increased by introducing a one-step $W$-estimate, which is a weighted average—the sum of the products of each estimate by its weight, divided by the sum of the weights [16]:

$$Z_W = \frac{\sum_{i=1}^{N} Z_i w_i}{\sum_{i=1}^{N} w_i}.$$

Parameters calculated based on robust statistics can be used to determine confidence intervals, for example, using the absolute median deviation (9). Then, with the average value of the estimate $\hat{Z}$, the confidence interval is determined by the inequality:
We also note a number of studies with other methods of obtaining estimates, which can be used to supplement the algorithm in the future.

In [18], an algorithm for obtaining estimates of the maximum likelihood type was implemented, based on the search for the maximum of the logarithmic likelihood function, a logarithmic function of parameters for a given set of residuals, taking into account the distribution of residuals for the recorded MT data.

A. Chave and A. Jones in [3] presented several efficient algorithms for finding M-estimates: based on the weighted least squares method and on the basis of calculating limited influence functions—diagonal elements of the influence matrix, followed by weighing the results obtained using various weight functions.

Larsen [12] proposed a component estimation algorithm based on the WLS method using different types of weighting functions for the time and frequency domain.

In [19], a method is proposed for finding MT transfer functions by determining the impulse response in the time domain.

4. The research results

A series of numerical experiments were carried out using synthetic data from the COMDAT project. As a result, derived estimate, some of which are shown in figure 2. The estimates are obtained using the Fourier transform, for noiseless data—in figure 2 (a) (CLS method and Siegel's method), for noisy data—in figure 2 (b) (CLS method and Siegel's method) and in figure 2 (c) (CLS method and WLS method).

Figure 2. Results of the computational experiment.
Numerical indicators of the accuracy and efficiency of estimation of some methods are shown in table 1.

**Table 1. Evaluation performance indicators according to the results of the computational experiment.**

| Method              | Mean normalized bias of the estimate | Normalized standard deviation |
|---------------------|--------------------------------------|-----------------------------|
| CLS method          | 0.203                                 | 0.267                       |
| WLS method          | 0.148                                 | 0.183                       |
| Siegel's median score | 0.164                               | 0.217                       |

Only a small part of the results is presented, since the purpose of this publication is to show the methodology for finding the optimal way to obtain estimates.

Shown in figure 1, the algorithm has a large number of variants, and at some stages its additional adjustment is possible. So, for example, the stage of rejection of spectral estimates by the coherence criterion has at least two parameters: the minimum threshold requirement of admissible coherence and the minimum fraction of the taken into account data with coherence not lower than a preset threshold. These two parameters of this stage of the algorithm, among other things, depend on the results of the performance indicators for evaluating the subsequent stages of the algorithm. Nonetheless (nevertheless), general patterns are reflected in the values given in table 1.

Different algorithms of robust estimation can improve the efficiency of estimation of impedance by tens of percent in comparison with CLS methods, but for the final choice of the optimal algorithm, it is necessary to carry out computational experiments on a large array of various synthetic data and consider various combinations of algorithm options.

Additional advantages can be provided by the use of artificial neural networks and the transfer of algorithms that are critical from the point of view of computational costs, which will make it possible to implement preliminary estimation in the embedded software of the automated system recorder. To date, it is known to use neural networks to clear signals from noise [20], as well as to interpret the results at subsequent stages of solving the inverse MT problem, which are beyond the scope of this publication.

**5. Conclusions**

The authors have developed a technique for finding the optimal method for obtaining impedance estimates in the MT sounding problem, which is based on the indicators of the accuracy and efficiency of estimates: the mean normalized bias of an estimate, the normalized standard deviation, the mean normalized confidence interval. The methodology provides for the study of a number of assessment methods in their various combinations with the aim of rejecting methods that are poorly suited to a given task, and selecting recommended methods that give relatively good results. The algorithm has passed partial computational tests and needs to be expanded in terms of the number of estimation methods used and tested on a variety of synthetic and actual MT data.

Let's note the tasks list to be investigated in the near future:

- Supplementing the considered algorithm with other methods.
- Selection of the type, software implementation, research and training of an artificial neural network based on the proposed set of methods, which is planned to be considered to determine the criteria for goal-setting in its training. Practical hardware implementation of a trained neural network in a signal recorder and in applied software.
- Implementation of a hardware base for joint distributed preprocessing of data in real time.
6. References

[1] Berdichevsky M N and Dmitriev V I 2009 Models and methods of magnetotellurics (Berlin: Springer)

[2] Berdichevsky M N, Dmitriev V I, Novikov D B and Pastutsan V V 1997 Analysis and interpretation of magnetotelluric data (Moscow: Dialogue-MSU)

[3] Chave A D and Jones A G 2012 The magnetotelluric method: theory and practice (New York: Cambridge University Press)

[4] Smirnov M Yu 2003 Magnetotelluric data processing with a robust statistical procedure having a high breakdown point Geophys. J. Int. 152 pp 1–7.

[5] Vagin S A 2009 Algorithm of processing magnetotelluric sounding data using Matlab Geophysical issues 42 pp 100–114.

[6] Larsen J C 1989 Transfer functions: smooth robust estimates by least squares and remote reference methods Geophys. J. Int. 99 pp 645–663.

[7] Chave A D 1987 On the robust estimation of power spectra, coherences, and transfer functions Journal of Geophysical Research 92 pp 633–648.

[8] Egbert G D 1986 Robust estimation of geomagnetic transfer functions Geophys. J. R. 87 pp 173–194.

[9] Gamble T D, Goubau W M and Clarke J 1979 Magnetotellurics with a remote magnetic reference Geophysics 44 pp 53–68.

[10] Ernst T, Sokolova E Yu, Varentsov I M and Golubev N G 2001 Comparison of two techniques for magnetotelluric data processing using synthetic data sets Acta Geophysica Polonica 53 pp 213–243.

[11] Heagy L J, Cockett R, Kang S, Rosenkjaer G K and Oldenburg D W 2017 A framework for simulation and inversion in electromagnetics Computers & Geosciences 107 pp 1–19.

[12] Bendat J S and Piersol A G 2011 Random data: analysis and measurement procedures (New York, Wiley).

[13] Chave A D and Thomson D J 2004 Bounded influence magnetotelluric response function estimation Geophys. J. Int. 159 pp 988–1066.

[14] Sims W E 1971 The estimation of magnetotelluric impedance tensor elements from measured data Geophysics 36 pp 938–942.

[15] Huber P J 2004 Robust statistics (New York, Wiley).

[16] Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W A 1986 Robust statistics: the approach based on influence functions (New York, Wiley).

[17] Siegel A F 1982 Robust regression using repeated medians Biometrika 69 pp 242–244.

[18] Chave A D 2017 Estimation of the magnetotelluric response function: the path from robust estimation to a stable maximum likelihood estimator Surv. Geophys. 38 pp 837–867.

[19] Nowozynski K 2004 Estimation of magnetotelluric transfer function in the time domain over a wide frequency band Geophys. J. Int. 158 pp 32–41.

[20] Manoj C and Nagarajan Nandini 2003 The application of artificial neural networks to magnetotelluric time-series analysis Geophys. J. Int. 153 pp 409–423.