

*М. Ю. Смирнов*

## ОБРАБОТКА МАГНИТОТЕЛЛУРИЧЕСКИХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ РОБАСТНЫХ СТАТИСТИЧЕСКИХ ПРОЦЕДУР

Одним из необходимых условий успешной интерпретации магнитотеллурических данных является высокое качество определения компонент тензора импеданса по временным записям естественного электромагнитного поля (ЕЭМП) на поверхности Земли.

В настоящее время предложено несколько различных подходов к обработке магнитотеллурических данных. Основным недостатком большинства из них является то, что они дают неудовлетворительные результаты, когда в записях ЕЭМП присутствуют различного рода помехи.

Значительный успех в повышении качества обработки достигнут на пути использования робастной статистики. Первые исследования по применению робастных статистик для обработки магнитотеллурических данных появились в конце 80-х годов [1–3]. Было показано преимущество робастных статистических процедур перед стандартными подходами, использующими метод наименьших квадратов, в случае, когда наблюдаемые данные осложнены помехами [4]. В настоящее время робастные алгоритмы обработки МТ-данных [5, 6] активно развиваются.

В данной статье предлагается алгоритм обработки магнитотеллурических данных, основанный на использовании робастных процедур. Алгоритм состоит из трех частей: предварительная подготовка исходных данных, спектральное преобразование магнитотеллурических записей, определение компонент тензора импеданса. Спектральное преобразование осуществляется с использованием классического спектрального анализа на основе быстрого преобразования Фурье (БПФ) [7].

### Предварительная подготовка исходных данных

Одним из видов помех, которые необходимо убрать из исходных данных перед началом обработки, чтобы уменьшить ошибки определения компонент тензора импеданса, являются большие выбросы, присутствующие как в электрических, так и в магнитных компонентах. Для устранения больших выбросов из данных предлагается использовать метод авторегрессионного предсказания, который заключается в построении авторегрессионной (АР) модели процесса и вычисления ошибок предсказания. Если ошибка предсказания превышает некоторый порог  $\varepsilon_e$ , тогда значение в точке заменяется предсказанным значением.

АР-модель временного ряда с дискретным временем описывается выходом фильтра, выраженным следующим разностным уравнением:

$$x[n] = - \sum_{k=1}^p a[k]x[n-k] + u[n] = \sum_{k=0}^{\infty} h[k]u[n-k],$$

где  $x[n]$  — последовательность на выходе каузального фильтра, который формирует наблюдаемые данные;  $u[n]$  — входная возбуждающая последовательность;  $h[k]$  — импульсный отклик фильтра.

Предположим, что оценивание АР-параметров порядка  $p$  осуществляется по  $N$ -точечному сегменту данных. Оценка линейного предсказания вперед  $\hat{x}[n]$  для отсчета  $x[n]$  будет иметь форму

$$\hat{x}[n] = - \sum_{k=1}^p a_p[k] x[n-k],$$

где  $a_p$  — коэффициент линейного предсказания вперед порядка  $p$ . Ошибка линейного предсказания вперед определяется выражением

$$e_p[n] = x[n] - \hat{x}[n] = x[n] + \sum_{k=1}^p a_p[k] x[n-k].$$

Здесь предполагается, что ошибка предсказания является белым шумом, с тем чтобы эта модель соответствовала АР-модели.

Тогда, если с использованием некоторого алгоритма оценены коэффициенты  $a_p[k]$  для выбранного порядка  $p$  модели, то оценивая ошибку предсказания для  $(N+1)$ -го отсчета можно сказать, является  $(N+1)$ -й отсчет выбросом или нет. В том случае, если  $e_p[N+1] > \varepsilon_e$ , отсчет  $x[N+1]$  заменяется предсказанным значением  $\hat{x}[N+1]$ . Значение  $\varepsilon_e$  выбирается исходя из выборочной дисперсии ошибки предсказания:

$$D_p = \frac{1}{N-p-1} \sum_{p+1}^N |e_p[n]|^2.$$

Таким образом с окном длиной  $N$  отсчетов обрабатывается весь ряд данных со сдвигом в один отсчет. При этом, если это необходимо, корректируется  $(N+1)$ -й отсчет, который попадет в следующее окно. Для вычисления АР-коэффициентов используется алгоритм модифицированного ковариационного метода.

### Спектральное преобразование магнитотеллурических записей

Для того чтобы определить компоненты тензора импеданса, необходимо решить систему линейных уравнений, связывающих компоненты ЕЭМП в частотной области [8]:

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix}. \quad (1)$$

Пусть зарегистрированы временные последовательности компонент ЕЭМП  $e_x(t)$ ,  $e_y(t)$ ,  $h_x(t)$ ,  $h_y(t)$ . Поскольку решение системы (1) выполняется в частотной области, то необходимо произвести спектральное преобразование временных последовательностей. Для этого используется классический спектральный анализ, основанный на БПФ. Избыточная система уравнений для (1) в дальнейшем составляется из выборочных спектров, рассчитанных для отдельных сегментов данных длиной  $N$ , на которые разбивается весь массив данных. Перекрытие сегментов данных выбирается в диапазоне от 0 до 75 %.

Каждый сегмент данных обрабатывается в следующей последовательности.

1. Из сегмента удаляется среднее:

$$M = \frac{1}{N} \sum_{i=1}^N e_x(i); \quad \bar{e}_x(i) = e_x(i) - M; \quad i = 0, \dots, N-1.$$

2. Для уменьшения смещения оценок каждый сегмент обрабатывается спектральным окном Ханна:

$$\tilde{e}_x(i) = \bar{e}_x(i)h(i), \quad h(i) = 0,5 - 0,5 \cos \frac{2\pi i}{N-1}.$$

3. Осуществляется быстрое преобразование Фурье:  $\tilde{e}_x(t) \rightarrow E_x(\omega)$ .

Аналогичной обработке подвергаются все компоненты ЕЭМП.

Систему уравнений (1) удобно записать через спектральные плотности следующим образом, отдельно для каждого уравнения [7]:

$$\begin{cases} S_{E_x H_x} = Z_{xx} S_{H_x H_x} + Z_{xy} S_{H_y H_x}, \\ S_{E_x H_y} = Z_{xx} S_{H_x H_y} + Z_{xy} S_{H_y H_y}, \end{cases} \quad (2a)$$

$$\begin{cases} S_{E_y H_x} = Z_{yx} S_{H_x H_x} + Z_{yy} S_{H_y H_x}, \\ S_{E_y H_y} = Z_{yx} S_{H_x H_y} + Z_{yy} S_{H_y H_y}, \end{cases} \quad (2b)$$

и

$$\begin{cases} S_{E_x E_x} = Z_{xx} S_{E_x H_x} + Z_{xy} S_{E_x H_y}, \\ S_{E_y E_x} = Z_{yx} S_{E_y H_x} + Z_{yy} S_{E_y H_y}, \end{cases} \quad (3a)$$

$$\begin{cases} S_{E_y E_y} = Z_{yx} S_{E_y H_x} + Z_{yy} S_{E_y H_y}, \\ S_{E_x E_y} = Z_{yx} S_{E_x H_x} + Z_{yy} S_{E_x H_y}, \end{cases} \quad (3b)$$

где  $S_{XY}$  — сглаженные оценки спектральных плотностей.

Уравнения (2) соответствуют импедансной оценке тензора импеданса  $[Z]$ , а уравнения (3) — адмитансной оценке. Именно эти уравнения будут использоваться в дальнейшей обработке.

4. Вычисляются для каждого сегмента несглаженные авто и взаимные спектральные плотности компонент поля:

$$S_{XY}^n(\omega) = \frac{1}{N} X(\omega) Y^*(\omega),$$

где \* означает комплексное сопряжение;  $X, Y = E_x, E_y, H_x, H_y$ .

Для того чтобы исключить из дальнейшей обработки сегменты с большим уровнем помех или содержащие слабый полезный сигнал, выполняется отбраковка спектров. В качестве критерия отбраковки используются коэффициенты частной когерентности  $Co_{E_x H_y \cdot H_x}^2, Co_{E_y H_x \cdot H_y}^2$  [7]:

$$Co_{E_x H_y \cdot H_x}^2 = \frac{|S_{E_x H_y \cdot H_x}|^2}{S_{E_x E_x \cdot H_x} S_{H_y H_y \cdot H_x}},$$

где  $S_{XY \cdot Z}$  — условные спектральные плотности:

$$S_{XY \cdot Z} = S_{XY} - S_{ZY} S_{XZ} / S_{ZZ},$$

$S_{XY}$  — сглаженные спектральные плотности;  $X, Y, Z = E_x, E_y, H_x, H_y$ ;

$$S_{XY}(j) = \frac{1}{k+1} \sum_{m=-\frac{k}{2}}^{\frac{k}{2}} S_{XY}^n(j+m), \quad j = \frac{k}{2}, \dots, \frac{N}{2} - \frac{k}{2}.$$

При этом сглаживание для отдельного сегмента осуществляется по  $k + 1$  соседним спектральным частотам (эти сглаженные спектральные плотности дальше не используются).

Если когерентности сегмента превосходят некоторый наперед заданный порог, то спектральные плотности сохраняются для дальнейшей обработки. Таким образом процесс повторяется для каждого сегмента до тех пор, пока не будет обработан весь ряд данных.

В результате получается набор прошедших отбраковку несглаженных спектральных плотностей:

$$S_{E_z H_z i}^n(\omega), S_{E_z H_y i}^n(\omega), S_{E_y H_z i}^n(\omega), S_{E_y H_y i}^n(\omega), S_{H_z H_y i}^n(\omega),$$

$$S_{E_z E_z i}^n(\omega), S_{E_z E_y i}^n(\omega), S_{H_z H_z i}^n(\omega), S_{H_z H_y i}^n(\omega), i = 1, \dots, M.$$

От числа выборочных спектров  $M$ , полученных в результате спектрального анализа, существенно зависит скорость последующих вычислений. Как показала практика, оптимальным будет значение  $M$  в пределах от 50 до 200. Увеличить число  $M$  можно увеличением перекрытия сегментов данных. Чтобы уменьшить  $M$ , лучше всего усреднить выборочные спектры по  $k$  соседним сегментам, и тогда новое  $M' = M/k$ .

### Расчет компонент тензора импеданса

Классическим методом решения системы уравнений (1) является метод наименьших квадратов (МНК). Оценка МНК получается из преобразованной системы уравнений (2), если сглаживание выборочных спектров осуществляется усреднением по сегментам:

$$S_{XY}(\omega) = \frac{1}{M} \sum_{k=1}^M S_{X^k Y^k}^n(\omega).$$

Согласно теореме Маркова–Гаусса оценка МНК является оптимальной в классе несмещенных оценок и имеет минимальную дисперсию среди всех несмещенных оценок, если ошибки  $e_i$  распределены нормально. Однако нормальная модель никогда не бывает абсолютно корректной, поскольку практически во всех реальных ситуациях присутствуют выделяющиеся наблюдения с большими ошибками, отклоняющимися от модели. При этом катастрофически падает эффективность МНК [9].

Анализ пороговых свойств оценки МНК показывает, что достаточно одного резко выделяющегося наблюдения, чтобы существенно изменить результат. При магнитотеллурических зондированиях такими наблюдениями являются, например, мощные импульсные помехи, имеющие отличную от магнитотеллурической передаточную функцию.

Таким образом, желательно использовать процедуры обработки магнитотеллурических данных, более устойчивые к присутствию различного рода помех. Использование робастных методов позволяет получить желаемый результат, поскольку они остаются эффективными и устойчивыми не на точной модели, а в некоторой ее окрестности.

В данной работе используется робастная оценка с максимальной пороговой точкой  $\varepsilon^* = 50\%$ . Эта оценка была впервые предложена Зигелем [10] и вычисляется посредством алгоритма повторяющихся медиан. Для любых  $p$  наблюдений

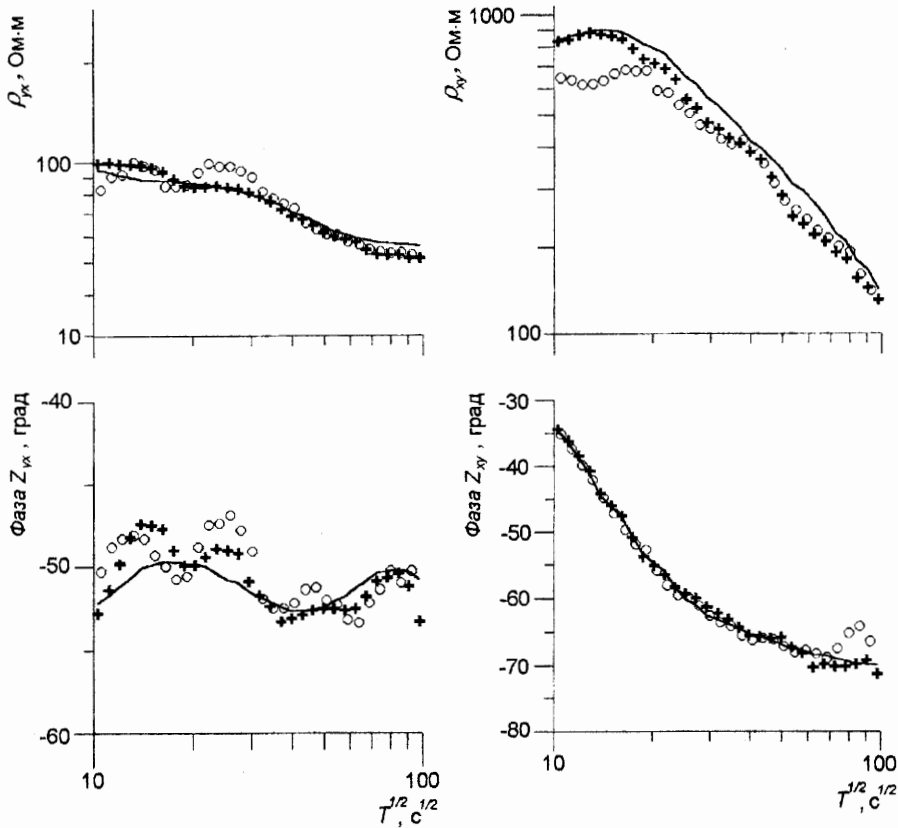


Рис. 1. Результат обработки синтетических данных:

крестики — робастная оценка; кружки — оценка по МНК; жирная линия — модель.

$(x_{i_1}, y_{i_1}), \dots, (x_{i_p}, y_{i_p})$ , заданных в общем виде, обозначим через  $\Theta(i_1, \dots, i_p)$  однозначно определяемый ими векторный параметр. Оценка Зигеля имеет вид

$$T_n^{(j)} = \text{med}_{i_1} \{ \dots \{ \text{med}_{i_{p-1}} \{ \text{med}_{i_p} \{ \Theta^{(j)}(i_1, \dots, i_p) \} \} \} \dots \},$$

где  $T_n^{(j)}$  есть  $j$ -я компонента  $T_n$ . В случае простой регрессии  $y = kx$ , угол наклона  $k$  оценивается по формуле

$$k = \text{med}_i \text{med}_{j \neq i} \frac{y_i - y_j}{x_i - x_j},$$

т. е. вычисляется как медиана всевозможных значений  $k$ .

Оценка Зигеля  $[Z]_S(\omega)$  компонент тензора импеданса вычисляется на основе прошедших отбраковку выборочных несглаженных спектральных плотностей с использованием уравнений (2). Можно использовать также и адмитансную оценку, однако это существенно увеличивает объем вычислений, практически не улучшая оценку.

Оценка компонент тензора импеданса однозначно определяется по двум реализациям. Тогда последовательность вычислений оценки Зигеля следующая:

1. Вычисляется набор оценок тензора импеданса  $[Z]_{ij}(\omega)$  из системы уравнений (2), при этом спектральные плотности в них выражаются следующим образом:

$$S_{XY}(\omega) = (S_{X^*Y_1}^n(\omega) + S_{X^*Y_2}^n(\omega))/2,$$

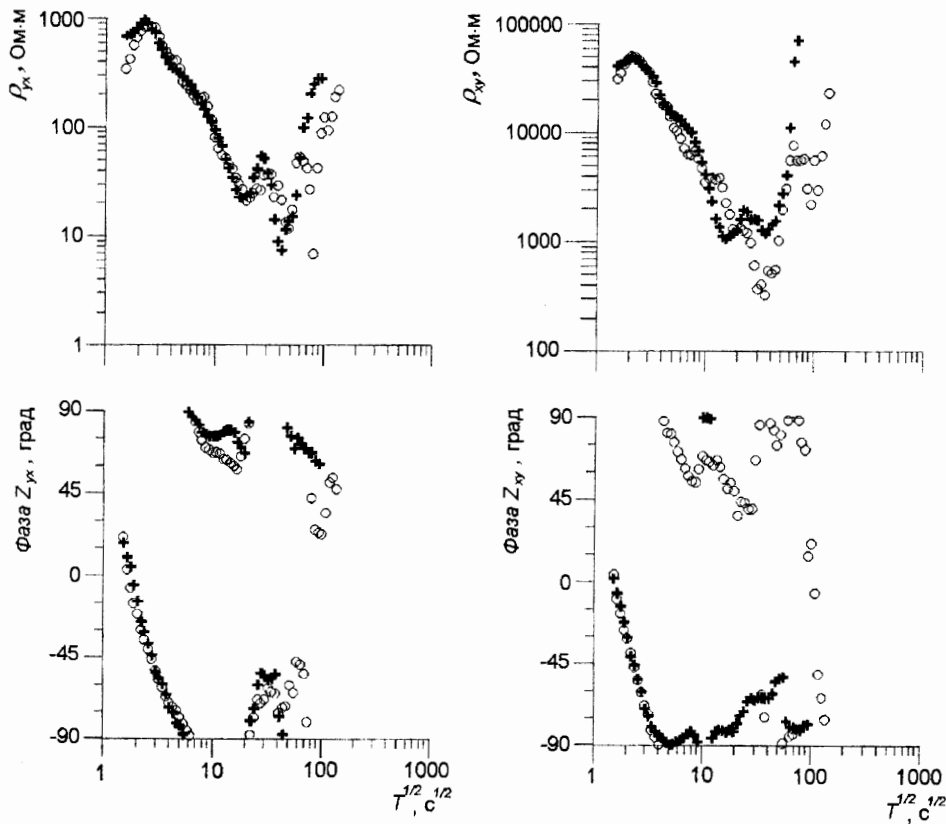


Рис. 2. Результат обработки реальных МТ-данных:  
крестики — робастная оценка; кружки — оценка по МНК.

где  $i = 1, \dots, M$ ;  $j = 1, \dots, M$ ;  $X, Y = E_x, E_y, H_x, H_y, H_z$ .

Для каждого индекса  $i$  перебираются все индексы  $j$ . Чтобы сократить вычислительные затраты, берутся индексы  $j > i$ .

2. Удаляются искаженные оценки. Отбраковка  $[Z]_{ij}(\omega)$  осуществляется по критерию фазы. Фаза главных компонент тензора импеданса должна лежать в пределах [8]:

$$\operatorname{Re} Z_{xy} > 0, \operatorname{Im} Z_{xy} < 0, \operatorname{Re} Z_{yz} < 0, \operatorname{Im} Z_{yz} > 0.$$

Как показала практика, этот критерий является наиболее мощным. Очень важно при использовании этого критерия правильно определять знаки компонент поля.

3. Вычисляется покомпонентно, т. е. для мнимых и вещественных частей каждой компоненты тензора импеданса отдельно, оценка Зигеля

$$[Z]_S(\omega) = \operatorname{med}_i \operatorname{med}_j \{ [Z]_{ij}(\omega) \}.$$

Аналогичные вычисления повторяется для каждой частоты.

4. Чтобы получить более устойчивую и гладкую оценку, результат сглаживается по соседним периодам в частотном окне от  $(T - \frac{1}{4}T)$  до  $(T + \frac{1}{4}T)$  с использованием одномерной сниженной  $M$ -оценки Хьюбера [11]. Сниженными называются оценки

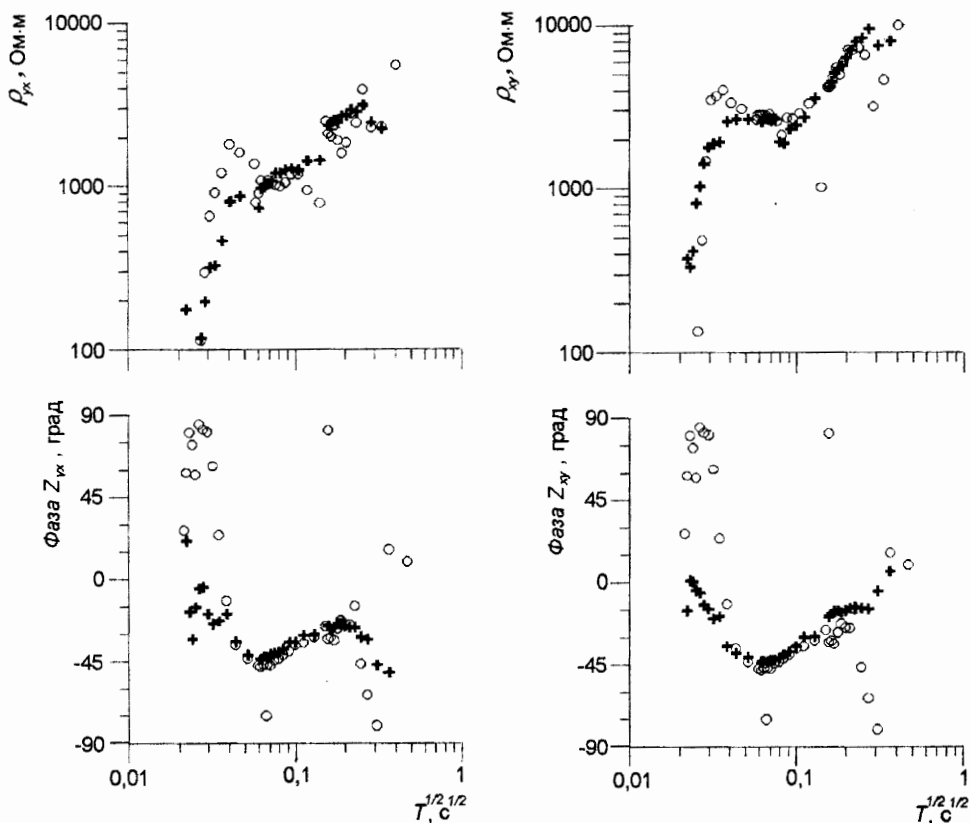


Рис. 3. Результат обработки реальных АМТ-данных:  
крестики — робастная оценка; кружки — оценка по МНК.

по усеченному набору данных. Для вычисления оценки наборы  $\{[Z]_{ij}(\omega)\}$  усекаются относительно оценки  $[Z]_S(\omega)$  соответственно на каждой частоте, т. е. выбираются  $p$  оценок, ближайших к оценке  $[Z]_S(\omega_l)$ , и создается набор  $[Z]_i^*(\omega)$ , состоящий из  $n = pl$  значений, где  $l$  — число периодов в окне сглаживания. В качестве окончательной оценки берется величина, вычисляемая покомпонентно для мнимых и вещественных частей отдельно:

$$\overline{Z_{ij}(\omega)} = \sum_{k=1}^n Z_{ijk}^* w_k / \sum_{k=1}^n w_i,$$

где  $i, j = x, y$ ,

$$w_k = \begin{cases} 1, & |r_k| \leq c, \\ c/|r_k|, & |r_k| \geq c, \end{cases} \quad r_k = \frac{Z_{ijk}^* - Z_{ijs}}{S_{\text{mad},ij}},$$

константа  $c$  выбирается в диапазоне от 1 до 2,  $S_{\text{mad}}$  — параметр масштаба, определяемый как медиана абсолютных отклонений:

$$S_{\text{mad},ij} = 1,483 \text{ med}_k \{|Z_{ijk}^* - Z_{ijs}\}.$$

5. Оценка параметра масштаба используется для вычисления 95 % доверительных интервалов  $\Delta_{ij}$  для компонент тензора импеданса:  $\Delta_{ij} = 1,96 S_{\text{mad},ij}$ .

## Результаты обработки синтетических и реальных данных

Предложенный алгоритм был тестирован на синтетических магнитотеллурических реализациях, любезно предоставленных И. М. Варенцовым и Е. Ю. Соколовой. Данные имели длину в 30000 точек с интервалом дискретизации 20 с и были осложнены различного рода шумами [12].

Результаты обработки по предложенной программе представлены на рис. 1, там же приведен результат обработки по МНК. Робастная оценка имеет меньше выбросов относительно модельной кривой. Заметно лучший результат предложенная оценка дает для амплитудных кривых. На фазовых кривых наблюдается лишь незначительное улучшение результата. Следует отметить, что МНК дает неплохую оценку фазы  $Z_{xy}$  для этих данных.

На рис. 2, 3 показаны соответственно результаты обработки МТ- и АМТ-данных, полученных в одной из точек профиля SVEKALAPKO на Балтийском щите с использованием аппаратуры, разработанной на кафедре физики Земли НИИФ СПбГУ. МТ-данные имели длительность 1 сут. Видно, что робастная оценка дает более устойчивые оценки как для кажущегося сопротивления, так и для фазы импеданса. В диапазоне периодов примерно от 4 до 40 с МНК- и робастная оценки совпадают. Это говорит о высоком качестве данных в этом диапазоне периодов, т. е. магнитотеллурическое поле существенно превосходило помехи. В этом случае проявляются асимптотические свойства робастной оценки, когда при идеальных данных (ошибка распределена нормально) она совпадает с оценкой МНК.

Таким образом, предложенный алгоритм позволяет повысить качество обработки магнитотеллурических данных по сравнению со стандартным подходом, основанном на методе наименьших квадратов.

Автор выражает признательность своему научному руководителю Аиде Андреевне Ковтун за постоянные обсуждения работы и ценные советы.

## Указатель литературы

1. Egbert G. D., Booker J. R. Robust estimation of geomagnetic transfer functions // *Geophys. J. Royal Astron. Soc.* 1986. Vol. 87. P. 173–194.
2. Chave A. D., Thomson D. J., Ander M. E. On the robust estimation of power spectra, coherences and transfer functions // *J. Geophys. Res.* 1987. Vol. 92. P. 633–648.
3. Chave A. D., Thomson D. J. Some comments on magnetotelluric response function estimation // *J. Geophys. Res.* 1989. Vol. 94. N B10. P. 14215–14225.
4. Jones A. G., Chave A. D., Egbert G. D. e. a. A comparison of techniques for magnetotelluric response function estimation // *J. Geophys. Res.* 1989. Vol. 94. N B10. P. 14201–14213.
5. Egbert G. D., Livelybrooks D. W. Single station magnetotelluric impedance estimation: coherence weighting and regression M-estimation // *Geophys.* 1996. Vol. 61. P. 964–970.
6. Larsen J. C., Mackie R. L., Manzella A. e. a. Robust smooth magnetotelluric transfer functions // *Geophys. J. Int.* 1996. Vol. 124. P. 801–819.
7. Семенов В. Ю. Обработка данных магнитотеллурического зондирования. М.: Недра, 1985.
8. Бердичевский М. Н. Электрическая разведка методом магнитотеллурического профилирования. М.: Недра, 1968.
9. Хампель Л., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. М.: Мир, 1989.
10. Siegel A. F. Robust regression using repeated medians // *Biometrika.* 1982. Vol. 69. P. 242–244.
11. Хьюбер П. Робастность в статистике. М.: Мир, 1984.
12. Варенцов И. М., Соколова Е. Ю. Генерация синтетических серий МТ-данных // *Физика Земли.* 1994. № 6. С. 80–88.