HAL
open science

Constrained robust estimation of magnetotelluric impedance functions based on a bounded-influence regression M-estimator and the Hilbert transform

D. Sutarno

To cite this version:

D. Sutarno. Constrained robust estimation of magnetotelluric impedance functions based on a bounded-influence regression M-estimator and the Hilbert transform. Nonlinear Processes in Geophysics, 2008, 15 (2), pp.287-293. hal-00302983

HAL Id: hal-00302983

https://hal.science/hal-00302983v1

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nonlin. Processes Geophys., 15, 287–293, 2008
www.nonlin-processes-geophys.net/15/287/2008/
© Author(s) 2008. This work is licensed
under a Creative Commons License.

Nonlinear Processes in Geophysics

Constrained robust estimation of magnetotelluric impedance functions based on a bounded-influence regression M-estimator and the Hilbert transform

D. Sutarno

Physics of Complex System Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia

Received: 12 May 2006 – Revised: 1 October 2007 – Accepted: 2 February 2008 – Published: 25 March 2008

Abstract. Robust impedance estimation procedures are now in standard use in magnetotelluric (MT) measurements and research. These always yield impedance estimates which are better than the conventional least square (LS) estimation because the 'real' MT data almost never satisfy the statistical assumptions of Gaussian distribution upon which normal spectral analysis is based. The robust estimation procedures are commonly based on M-estimators that have the ability to reduce the influence of unusual data (outliers) in the response (electric field) variables, but are often not sensitive to exceptional predictors (magnetic field) data, which are termed leverage points.

This paper proposes an alternative procedure for making reliably robust estimates of MT impedance functions, which simultaneously provide protection from the influence of outliers in both response and input variables. The means for accomplishing this is based on the bounded-influence regression M-estimation and the Hilbert Transform operating on the causal MT impedance functions. In the resulting regression estimates, outlier contamination is removed and the self consistency between the real and imaginary parts of the impedance estimates is guaranteed. Using synthetic and real MT data, it is shown that the method can produce improved MT impedance functions even under conditions of severe noise contamination.

1 Introduction

In the frequency domain, the equations governing the magnetotelluric (MT) relations between the signal components of electric and magnetic fields at the earth's surface are given by:

$$E_i = Z_{ix} H_x + Z_{iy} H_y, \qquad i = x \text{ or } y \tag{1}$$

From the viewpoint of linear system theory, the tensor impedances Z_{ix} and Z_{iy} are transfer functions of a dual input, single output linear system through which the horizontal magnetic field components H_x and H_y (input) are related deterministically to the horizontal electric field components E_i (output).

A variety of methods have been proposed for the numerical computation of impedance and its associated errors. Most of these are based on classical least squares regression. It is well known that the concept, as well as the computation, of LS estimates is quite simple. Furthermore, with the Gaussian error assumption, LS procedures are statistically optimal in a precise sense. Unfortunately, the presence of outliers (abnormal data) superimposed on a common Gaussian noise background which constitutes the normal ambient noise field, often makes such assumptions about error untenable and results in distortion of the estimates. One appealing approach to dealing with outliers is to make the estimation procedure robust, i.e. to modify it in such manner that it is resistant to the effects of the outliers.

In MT context, robust procedures were introduced by Chave and Thomson (1989), Sutarno and Vozoff (1989, 1991a, b), and Jones et al. (1989). More recently, Chave and Thomson (2003, 2004), and Sutarno (2005) proposed and discussed the application of a robust procedure which can be classified as a bounded-influence M-estimator to the MT data. This paper aims at further improving the performance of the robust procedure. Our object is to obtain a reliable

Correspondence to: D. Sutarno
(sutarno@fi.itb.ac.id)

Published by Copernicus Publications on behalf of the European Geosciences Union and the American Geophysical Union.

Table 1. Estimation results for apparent resistivity on noisy synthetic MT data.

Maximum Noise Level	RMS Error LS Estimator	Hybrid Estimator	Bounded-influence Estimator
10%	3%	3%	3%
20%	7%	4%	4%
50%	34%	13.5%	11%

impedance estimation procedure, which is robust to the overall influence of outlier contamination.

2 Robust impedance estimation

In the real situation, Eq. (1) does not hold exactly due to presence of noise. Therefore, it is necessary to estimate the impedance function from many noisy data, and the problem becomes statistical. Suppose n independent sets of values at a given harmonic are available to estimate Z_{ij}. Then the following equation, which represents a linear regression problem, can be formed

$$\boldsymbol{x} = \mathbf{U}\boldsymbol{\beta} + \boldsymbol{r} \tag{2}$$

The vector $\boldsymbol{x}$ consists of n observations of the ith horizontal electric field component. Then the $n\times 2$ matrix $\mathbf{U}$ gives n values of the dual horizontal magnetic field components, the elements of $\boldsymbol{\beta}$ are impedance tensor elements (the unknown parameters to be estimated), and the elements of $\boldsymbol{r}$ are unknown errors or residuals.

The M-estimator (or Maximum likelihood type estimator of Huber, 1981) for $\boldsymbol{\beta}$ in Eq. (2) is the value of $\boldsymbol{\beta}$ which minimises

$$\sum_{i=1}^{n} \rho(r_i/s), \tag{3}$$

where $\rho(t)$ is a general function, called the loss function, and s is a scale parameter. Many methods exist to solve this minimisation problem. However, it is easiest to write the corresponding normal equation as a weighted LS problem and iterate to get a linear approximation.

There are a great number of possible forms for the weight function to be used in the iteratively reweighted LS algorithm. The most widely used is a "standard hybrid" (Huber, 1981). However, the Huber weights fall slowly for large residuals and provide inadequate protection again severe residuals. Therefore, more severe types of weight function, such as the one proposed by Thomson (1977), need to be applied for a few iterations after convergence with the Huber weights has been achieved. The scale parameter s must also be estimated robustly, for which the median absolute deviation (MAD) is one suitable choice.

The robust estimation procedures with the Huber and Thomson weights have capability to reduce the influence of unusual data (outliers) in the response (electric field) variables. However, they are often not sensitive to exceptional predictors (magnetic field) data, which are termed leverage points. Recall that for regression problem, the hat or predictor matrix is given by Huber (1981)

$$\mathbf{H} = \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T. \tag{4}$$

The above matrix $\mathbf{H}$ is a symmetric $n\times n$ projection matrix, that is $\mathbf{H}^2=\mathbf{H}$, and has p nonzero eigenvalues (the values equal to 1) and $(n-p)$ zero eigenvalues, where p is the number of columns in $\mathbf{U}$ (usually 2 for MT). The rank of the matrix is equal to p, and its diagonal elements satisfy

$$0 \le h_{ii} \le 1 \tag{5}$$

A large h_{ii} usually indicates that, in the p-dimensional space defined by the vectors whose components are the carriers or input variables, the stand linearly apart from other $\boldsymbol{u}$-vectors. In this sense, the $\boldsymbol{u}_i$ is an outlier in the carrier space. Huber (1981) referred to the points with large as the leverage points. If one makes least squares robust by using Huber weights, for instance, and if point i happens to be a leverage point with a high h_{ii}, then x_i can be grossly aberrant. However, the scaled residuals (r_i/s) will still remain at the constant part of the weight function. This undesirable condition has led to introduction of different robust estimation schemes for linear regression with the basic purpose of bounding the influence of outlying $\boldsymbol{u}_i$. On this basis, the undesirable condition can be corrected by cutting down the overall influence of observation i as well as the scale s, by a factor which depends on h_{ii} (Sutarno, 2005). In effect this divides each residual r_i by the modified scale factor $s\sqrt{1-h_{ii}}$, so that the corresponding weight can reduce larger residuals more rapidly at point of higher leverage.

For the MT problem, in which the data are complex rather than real [the vectors and matrices in Eq. (2) are complex], the determination of the robust scale as well as iteratively reweighting may be performed either in a real or a complex framework. However, the use of complex framework, within which the scale and weights are determined based on magnitude of residuals, is preferable because it is rotationally (i.e. phase) invariant.

A performance test for the bounded-influence estimator was carried out by using synthetic MT data developed by Sutarno and Vozoff (1989). The synthetic data were generated to simulate MT data contaminated by non-Gaussian noise that uniformly distributed. In this simulation the earth is assumed to be homogeneous and isotropic, with a resistivity of $100\,\Omega-m$. Three sets of synthetic MT data with maximum noise percentages of 10%, 20% and 50% were used in the test. The rms error of the apparent resistivities computed by three different analyses (LS, hybrid, and bounded-influence estimator), with respect to the assumed $(100\,\Omega-m)$

value, are shown in Table 1. It can be seen that for data with 10% maximum noise, the conventional LS estimation still can easily deal with this noise level. However, as the noise increases, the conventional estimation starts having some difficulties, and robust estimations evidently provide better results. It is clear that the superiority of the robust procedures over the conventional method increase with increasing percentage error in the data. The test also shows that the best result is given by the bounded-influence robust estimator.

3 Constrained robust impedance estimation

The superiority of the robust impedance estimation procedures over the conventional LS procedure was demonstrated by Chave and Thomson (1989, 2003, 2004), Jones et al. (1989), Sutarno and Vozoff (1989, 1991a, b), and Sutarno (2005). Using real data they showed that the robust procedures always yield impedance estimates which are better than the conventional estimates. However, further improvements are desirable to find more reliably robust estimates for MT impedance functions. The improvement is based on physical constraints imposed on the impedance functions, such as smoothness and consistency between the real and imaginary parts of the impedance. The means for accomplishing this is to incorporate the Hilbert transform (HT) operation on the causal impedance functions, in the regression M-estimation procedure, and to require that the real and imaginary part satisfy HT relationships which must apply to causal systems.

Let $\Re Z(\omega)$ and $\Im Z(\omega)$ be the real and imaginary parts of a complex variable $Z(\omega)$ which represents a causal system that has no poles in the right half of the complex plane. Thus they are related as HT pair at radian frequency ω, as (Boehl et al., 1977):

$$\Re Z(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\Im Z(\omega')}{\omega - \omega'} d\omega' \tag{6a}$$

and

$$\Im Z(\omega) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\Re Z(\omega')}{\omega - \omega'} d\omega'. \tag{6b}$$

The complex variable $Z(\omega)$ can be expressed in amplitude and phase form as

$$Z(\omega) = |(Z(\omega)|e^{i\Phi} \tag{7}$$

By taking the natural logarithms of $Z(\omega)$, then

$$\ln Z(\omega) = \ln |(Z(\omega)| + i\Phi(\omega). \tag{8}$$

If the HT pair is to be applied to the real and imaginary parts of $\ln Z(\omega)$, then this variable must represent a real causal system. Since both the poles and zeros of $Z(\omega)$ become the poles $\ln Z(\omega)$, the requirement of causality for $\ln Z(\omega)$ is equivalent to $Z(\omega)$ not having any poles or zeros in the right half of the complex plane. Therefore, if the HT is to be applied to $\ln Z(\omega)$, the complex variable $Z(\omega)$ must be a minimum phase function (Boehl et al., 1977). In this case, the relation is given by

$$\ln |(Z(\omega)| = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\Phi(\omega')}{\omega - \omega'} d\omega' \tag{9a}$$

and

$$\Phi(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\ln |Z(\omega')|}{\omega - \omega'} d\omega. \tag{9b}$$

Sutarno and Vozoff (1991a, b) proposed a phase-smoothed robust M-estimation procedure that incorporates the above phase-amplitude relationship in the regression M-estimation, for MT data. Recently, this relationship has also been applied in a bounded-influence robust regression procedure for MT impedance estimation (Sutarno, 2005). However, as addressed (Boehl et al., 1977; Jones, 1980), the minimum phase condition is applicable for a one-dimensional (1-D) response or a two-dimensional (2-D) MT response rotated to its principal axes, but it is not valid for more general 2-D and three-dimensional (3-D) earth system. It is well known that in a multi-dimensional or an anisotropic earth model, the electric and magnetic fields may no longer be orthogonal because of the spatial and directional irregularities in resistivity distribution. This implies that the corresponding impedance function can not be associated with a minimum phase system. On the other hand, there is no doubt that the causality condition in Eq. (6) is valid for a physical earth system. This is because the condition is a general property that must be verified by any physically realizable system (Yee and Paulson, 1988). Hence, the MT impedance function can be considered to be a causal transfer function. On that basis, it is then possible to find a reliable impedance estimate by imposing this physically realistic constraint in the regression M-estimation. By assuming that the true response is most likely to lie between the one indicated by the real part and the other by the imaginary part, a compromise between these two extremes can be sought to represent an idealized response. In order to apply the above constraint in the regression M-estimation, let us recall the regression problem given by Eq. (2), but in the presence of $l \leq p$ linearly independent constraints

$$\boldsymbol{v} = \mathbf{C}\boldsymbol{\beta}, \tag{10}$$

where $\boldsymbol{v}$ is l independent restriction on the elements of $\boldsymbol{\beta}$. For simplicity let us consider a real case first, from which its complex analogue can be deduced.

The constrained M-estimate of $\boldsymbol{\beta}$ is the value of $\boldsymbol{\beta}$ which minimises

$$\sum_{i=1}^{n} \rho(r_i/s) \tag{11}$$

subject to the restriction

$$\mathbf{C}\boldsymbol{\beta} - \boldsymbol{v} = 0. \tag{12}$$

Therefore we minimise

$$\varsigma = \sum_{i=1}^{n} \rho(r_i/s) + \sum_{k=1}^{l} \lambda_k \left(\sum_{j=1}^{p} c_{kj} - v_k \right) \tag{13}$$

with respect to β_j and λ_k, where λ_k are Lagrange multipliers. Setting the derivative of ς with respect to β_j equal to zero, give for minimising value $\boldsymbol{\beta}^c$

$$\mathbf{U}^T\mathbf{W}\mathbf{U}\boldsymbol{\beta}^c - \mathbf{U}^T\mathbf{W}\boldsymbol{x} - s\mathbf{C}^T\boldsymbol{\Lambda} = 0, \tag{14}$$

whence

$$\boldsymbol{\beta}^c = (\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{W}\boldsymbol{x} + s(\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^T\boldsymbol{\Lambda}), \tag{15}$$

or

$$\boldsymbol{\beta}^c = \boldsymbol{\beta}^u + s(\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^T\boldsymbol{\Lambda}), \tag{16}$$

where $\boldsymbol{\beta}^u = (\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{U}^T\boldsymbol{x}$ is the unconstrained M-estimator, $\boldsymbol{\Lambda}$ is an $l \times 1$ vector of Lagrange multipliers, and $\mathbf{W}$ is a diagonal weight matrix. Premultiplying Eq. (16) by $\mathbf{C}$ and imposing the restriction $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{v}$, gives

$$\boldsymbol{v} = \mathbf{C}\boldsymbol{\beta}^c = \mathbf{C}\boldsymbol{\beta}^u + s\mathbf{C}(\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^T\boldsymbol{\Lambda}, \tag{17}$$

whence

$$\boldsymbol{\Lambda} = \frac{1}{s}(\mathbf{C}(\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^T)^{-1}(\boldsymbol{v} - \mathbf{C}\boldsymbol{\beta}^u). \tag{18}$$

Inserting this back into Eq. (16) gives

$$\boldsymbol{\beta}^c = \boldsymbol{\beta}^u + (\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^T(\mathbf{C}(\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^T)^{-1}(\boldsymbol{v} - \mathbf{C}\boldsymbol{\beta}^u). \tag{19}$$

It can be seen that the constrained M-estimator $\boldsymbol{\beta}^c$ differs from the unconstrained $\boldsymbol{\beta}^u$ by a nonlinear function of the quantity $(\boldsymbol{v} - \mathbf{C}\boldsymbol{\beta}^u)$, by which the unconstrained M-estimator fails to satisfy the constraints. The complex analogue of Eq. (19) given by

$$\boldsymbol{\beta}^c = \boldsymbol{\beta}^u + (\mathbf{U}^*\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^*(\mathbf{C}(\mathbf{U}^*\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^*)^{-1}(\boldsymbol{v} - \mathbf{C}\boldsymbol{\beta}^u). \tag{20}$$

The new robust method proposed here incorporates the HT operation on the causal impedance function, in the regression M-estimation procedure. The procedure is essentially an iteratively reweighted least-squares approach which is similar to the one proposed by Sutarno and Vozoff (1991a, b). Likewise, the procedure uses a complex framework. The main steps of the procedure are as follows:

(a) Use the LS estimator for the preliminary estimator $\beta^{(m)}(m{=}0$ and compute the diagonal elements of hat matrix h_{ii}.

(b) Compute the predicted outputs

$$\hat{x}_i^{(m)} = u_{ij}\beta_j^{(m)}, \qquad i = 1, \ldots, n$$

and residuals

$$r_i^{(m)} = x_i - \hat{x}_i^{(m)}, \qquad i = 1, \ldots, n.$$

(c) Use the magnitude of residuals to obtain the modified scale estimator

$$\hat{s}^{(m)} = \frac{med\left\{\left||r_i^{(m)}| - med(|r_i^{(m)}|)\right|\right\}}{0.44845}\sqrt{1 - h_{ii}}.$$

(d) The Huber weights are computed using the scaled magnitude of residuals and applied to matrix $\mathbf{W}$ of the weighted normal equation given by

$$\mathbf{U}^*\mathbf{W}\mathbf{U}\boldsymbol{\beta} = \mathbf{U}^*\mathbf{W}\boldsymbol{x}. \tag{21}$$

(e) Equation (21) is solved and the entire process is repeated until convergence is achieved. The iteration formula is given by

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{U}^*\mathbf{W}^{(m)}\mathbf{U})^{-1}\mathbf{U}^*\mathbf{W}^{(m)}\boldsymbol{x} \tag{22}$$

(f) Use the Thomson weights and a fixed scale estimate derived from the above final iteration, again solving Eq. (21) and terminating when convergence is achieved.

(g) Apply the HT operation to derive the real part of the estimate from the imaginary part and vice versa.

(h) Compute the mean of the two estimates (the original estimate obtained from (f) and the transformed estimate from (g)) at each frequency to obtain a compromise values of the two estimates, whilst the error estimate is calculated from their difference.

(i) Use the compromise values of the most consistent estimate to represent the idealized estimate, and again use the HT technique to find its pair.

(j) Repeat step (b) to (f), but using the idealized estimate $\breve{\beta}$ as a constraint. More explicitly:

$$\boldsymbol{v} = (\breve{\beta})$$

and

$$\mathbf{C} = \left\{ \begin{array}{l} (\mathbf{1}\ \mathbf{0}) \text{ if } \breve{\beta} = \beta_1; \\ (\mathbf{0}\ \mathbf{1}) \text{ if } \breve{\beta} = \beta_2. \end{array} \right\}$$

The iteration formula is now given by

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} = {} & (\mathbf{U}^*\mathbf{W}^{(m)}\mathbf{U})^{-1}\mathbf{U}^*\mathbf{W}^{(m)}\boldsymbol{x} \\ & + (\mathbf{U}^*\mathbf{W}^{(m)}\mathbf{U})^{-1}\mathbf{C}^*(\mathbf{C}(\mathbf{U}^*\mathbf{W}^{(m)}\mathbf{U})^{-1}\mathbf{C}^*)^{-1} \\ & \cdot\{\boldsymbol{v} - \mathbf{C}(\mathbf{U}^*\mathbf{W}^{(m)}\mathbf{U})^{-1}\mathbf{U}^*\mathbf{W}^{(m)}\boldsymbol{x}\} \end{aligned} \tag{23}$$

(k) Repeat steps (g) to (i), but only for the element that is not used as the constraint in (j).

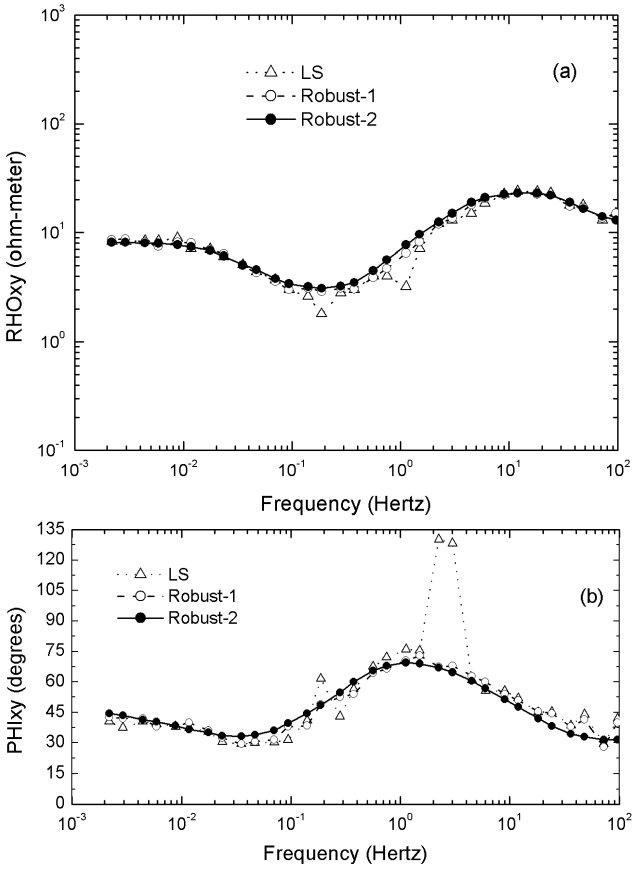

Fig. 1. (a) A comparison of results for apparent resistivity ρ_{xy} on noisy MT data from the Columbia River Plateau, by three different analyses: LS, unconstrained robust (Robust-1), and constrained robust (Robust-2). **(b)** The corresponding phases for panel (a). The new robust analysis provides very smooth results.

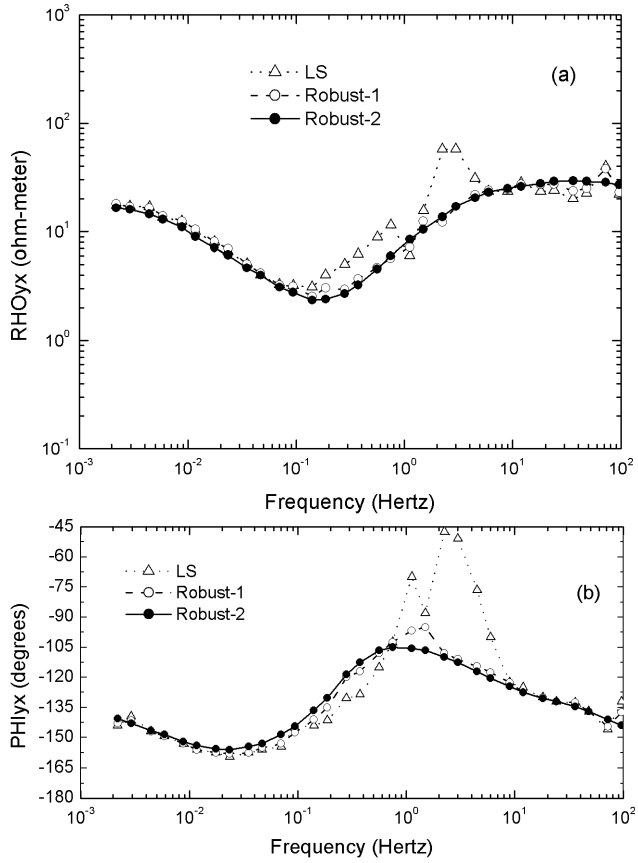

Fig. 2. (a) As Fig. 1a, but for ρ_{yx}. **(b)** The corresponding phases for panel (a). Again, the new robust results are much smoother than those of the unconstrained one.

Note that this robust regression procedure can be used with the standard (4-channel) or the remote reference (6-channel) data. To employ remote reference (RR) data, elements of matrix $\mathbf{U}^*$ in Eqs. (21), (22) and (23) should be replaced by the remote reference field values. A preliminary performance test for the new robust was carried out by using the same synthetic MT data sets discussed in previous section. The rms error of the apparent resistivities computed by this analysis is, respectively, 2.1%, 2.3% and 5.7% for noise percentages of 10%, 20% and 50%.

4 Application to real MT data

The real MT data set was taken in an extremely noisy area in the Columbia River Plateau, by Phoenix Geophysics Inc., using remote reference MT system. The data are very highly contaminated with common noise at the two locations. It was indicated that a hydroelectric dam and aluminum refinery which exist in this area were responsible for the noise (Sutarno and Vozoff, 1989). This preliminary study of the error behaviour of the data showed that Gaussian error assumption is violated. The error distribution has a typical long-tailed behavior caused by outliers. The impedance function example was computed for site 101 using site 102 horizontal magnetic fields as reference. The distance between the two sites is about 3 miles. Preliminary processing of the data was conducted using the standard MT analysis. The mean and linear trend were subtracted from each segment of 32 points. The segments were multiplied by a Hanning window and the 6th and 8th discrete Fourier transform (DFT) coefficients were calculated for each segment. To calculate the Fourier spectra for the output frequencies of the low bands in the low range, the cascade decimation procedure of Wight and Bostick (1986) was used. The impedance and thus the apparent resistivity and phase values were then computed using the new robust method as well as by the conventional LS and unconstrained robust method. Figures 1a and 2a show the apparent resistivity computed by the new method along with the results from the conventional LS and unconstrained robust,

whereas the corresponding phases are shown in Fig. 1b and Fig. 2b. Considering smoothness, slope and compatibility of the resulting apparent resistivity ρ_a and the phase Φ, it can be seen from the figures that the conventional LS analysis yields impedance estimates that are very poor in these qualities. The unconstrained robust substantially improves the impedance estimates. However, the best improvement is markedly given by the new constrained robust method. As can be seen from the figures, in comparison to the unconstrained robust, the new robust analysis provides much smoother impedance estimates.

Comparisons between the above results may also be statistically evaluated either self-consistently or comparatively. In the first instance, by using the HT operation, it is possible to measure a misfit error between the real part of the resulting estimates and the real part predicted from the imaginary part, or vice versa, at the end of each processing. On this basis, it can be expected that the proposed robust procedure will provide the best fit. This is because the procedure imposes the causality requirement on its estimates. The comparative evaluation can be done based on statistical properties of the resulting estimates, such as confidence limits. However, for this purpose, it is preferable not to use the parametric approach to describe this statistical property. This is because the approach depends on the hypothesis about the probability distribution. As is known, in the field of statistical inference the parametric approach has been displaced by newer nonparametric types such as the jackknife, the bootstrap and cross validation, which have better performance in complicated situations. The most frequently applied nonparametric method is the jackknife, due to its simplicity and reliability with real data. The jackknife estimate of the regression covariance matrix is given by Chave and Thomson (1989)

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \frac{1}{n(n-p)} \sum_{i=1}^{n} (\bar{\boldsymbol{\beta}} - \boldsymbol{p}_i)(\bar{\boldsymbol{\beta}} - \boldsymbol{p}_i)^T, \tag{24}$$

where

$$\boldsymbol{p}_i = \{n - (1 - h_{ii}) + 1\}\hat{\boldsymbol{\beta}} - n(1 - h_{ii})\hat{\boldsymbol{\beta}}_{-i}, \tag{25}$$

are the weighted pseudovalues, and $\bar{\boldsymbol{\beta}}$ is the arithmetic average of the values

$$\bar{\boldsymbol{\beta}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{p}_i. \tag{26}$$

In Eq. (25) $\hat{\boldsymbol{\beta}}_{-i}$ represents the estimate of $\boldsymbol{\beta}$ based on the ith subset, where the ith row of and has been removed. Thus, estimation of the jackknife covariance matrix of $\boldsymbol{\beta}$ requires the delete-one estimates of $\boldsymbol{\beta}$. These can be derived for either conventional or robust regression by deleting a row from $\boldsymbol{x}$ and $\mathbf{U}$ in turn and solving the problem. The diagonal terms of Eq. (24) give the jackknife regression variances. Estimates of standard errors on the regression coefficients $\boldsymbol{\beta}$ are given by the square root of the variances, and may then converted to confidence limits. An important property of jackknife variance is robustness in the presence of inhomogeneity of error variance, in contrast to parametric estimators.

5 Conclusions

Bounded-influence regression M-estimation which is robust to outlier contamination has been applied in the processing of extremely noisy MT data to estimate the impedance tensors. The bounded-influence estimator simultaneously cuts down the overall influence of both outliers and leverage points, and yields results that are comparable to the hybrid estimator, and are even better. Since the impedance function is a causal system, incorporating the HT operation in the regression M-estimation guarantees the self consistency between the real and imaginary parts and results in robust, reliable impedance estimates.

Acknowledgements. I thank B. Anderson, Phoenix Geophysics Inc. for provision of the data and permission to release it.

Edited by: O. Talagrand
Reviewed by: three anonymous referees

References

Boehl, J. E., Bostick, F. X., and Smith, H. W.: An application of the Hilbert transform to the magnetotelluric method, Tech. Rep. Electr. Geophys. Res. Lab., University of Texas, Austin, 1977.

Chave, A. D. and Thomson, D. J.: Some comments on magnetotelluric response function estimation, J. Geophys. Res., 94, 14 215–14 225, 1989.

Chave, A. D. and Thomson, D. J.: A bounded influence regression M-estimator based on the statistic of the hat matrix, J. Roy. Stat. Soc., Ser. C, 52, 307–322, 2003.

Chave, A. D. and Thomson, D. J.: Bounded influence magnetotelluric response function estimation, Geophys. J. Int., 157, 988–1006, 2004.

Huber, P. J.: Robust statistics, John Wiley and Sons, Inc, 1981.

Jones, A. G.: Geomagnetic induction studies in Scandanavia 1, J. Geophys. Res., 48, 181–194, 1980.

Jones, A. G., Chave, A. D., Egbert, G. D., Auld, D., and Bahr, K.: A comparison of techniques for magnetotelluric response estimation, J. Geophys. Res., 94, 14 201–14 123, 1989.

Sutarno, D. and Vozoff, K.: Robust M-estimation of magnetotelluric impedance tensors, Expl. Geophys., 20, 283–398, 1989.

Sutarno, D. and Vozoff, K.: Phase-smoothed robust M-estimation of magnetotelluric impedance functions, Geophysics, 56, 1999–2007, 1991a.

Sutarno, D. and Vozoff, K.: An application of regression M-estimation with the Hilbert transform to magnetotelluric data processing, Explor. Geophys. 22, 383–390, 1991b.

Sutarno, D.: Development of Robust Magnetotelluric Impedance Estimation: A Review, Indonesian J. Phys., 16, 79–89, 2005.

Thomson, D. J.: Spectrum estimation techniques for characterization and development of WT4 waveguide I, Bell Syst. Tech. J., 56, 1769–1815, 1977.

Wight, D. E. and Bostick, F. X.: Cascade decimation-A technique for real time estimation of power spectra, in: Magnetotelluric methods, edited by: Vozoff, K., Geophysics Reprint Series no. 5, Soc. Expl. Geophys., 215–218, 1986.

Yee, E. and Paulson, K. V.: Concerning dispersion relations for the magnetotelluric impedance tensor, Geophys. J., 95, 549–559, 1988.