


Estimation of the Magnetotelluric Response Function: The Path from Robust Estimation to a Stable Maximum Likelihood Estimator

Alan D. Chave¹ 

Received: 28 November 2016 / Accepted: 21 July 2017 / Published online: 24 August 2017
© Springer Science+Business Media B.V. 2017

Abstract The robust statistical model of a Gaussian core contaminated by outlying data in use since the 1980s, and which underlies modern estimation of the magnetotelluric (MT) response function, is re-examined from first principles. The residuals from robust estimators applied to MT data are shown to be systematically long-tailed compared to a distribution based on the Gaussian and hence inconsistent with the robust model. Instead, MT data are pervasively described by the stable distribution family for which the Gaussian is an end member, but whose remaining distributions have algebraic rather than exponential tails. The validity of the stable model is rigorously demonstrated using a permutation test. A maximum likelihood estimator (MLE), including the use of a remote reference, that exploits the stable nature of MT data is formulated, and its two-stage implementation, in which stable parameters are first fit to the residuals, and then the MT responses are solved for, with iteration between them, is described. The MLE is inherently robust, but differs from a conventional robust estimator because it is based on a statistical model derived from the data rather than being *ad hoc*. Finally, the covariance matrices obtained from MT data are pervasively improper as a result of weak non-stationarity, and the Cramér–Rao lower bound for the improper covariance matrix is derived, resulting in reliable second-order statistics for MT responses. The stable MLE was applied to an exemplar broadband data set from northwest Namibia. The stable MLE is shown to be consistent with the statistical model underlying linear regression and hence is unconditionally unbiased, in contrast to the robust model. The MLE is compared to conventional robust remote reference and two-stage estimators, establishing that the standard errors of the former are systematically smaller than for either of the latter, and that the standardized differences between them exhibit excursions that are both too frequent and too large to be described by Gaussian statistics. These excursions are more prevalent when the tail thickness parameter of the stable distribution is small, and are attributed to rising bias in

✉ Alan D. Chave
achave@whoi.edu

¹ Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

the robust estimator that is also consistent with the Berry–Esséen theorem that defines the rate of convergence to the central limit theorem value. An explanation for weak non-stationarity of MT data is proposed, and several extensions to the present work are described.

Keywords Magnetotellurics · Time-series analysis · Probability distributions · Numerical approximations and analysis

1 Introduction

The fundamental datum in magnetotellurics (MT) is a location-specific, frequency-dependent tensor linearly connecting the horizontal electric and magnetic fields measured at Earth's surface (or at the seafloor). Under general conditions (Weidelt and Chave 2012, Section 4.1.2), in the absence of noise and with precise data, the relationship may be written as follows:

$$\mathbf{E} = \vec{\mathbf{Z}} \cdot \mathbf{B} \quad (1)$$

where \mathbf{E} and \mathbf{B} are two-vectors of the horizontal electric and magnetic field components at a specific site and frequency, $\vec{\mathbf{Z}}$ is the second-rank, 2×2 MT response tensor connecting them, and \cdot denotes the inner product. Additional transfer functions exist between the vertical and horizontal magnetic fields, or between the horizontal magnetic fields at different sites, for which the estimation methodologies are similar, and hence only the MT response $\vec{\mathbf{Z}}$ will be considered in this paper.

When \mathbf{E} and \mathbf{B} are actual measurements (1) does not hold exactly due to the finite size of the data sample and the presence of noise, and it becomes necessary to estimate both $\vec{\mathbf{Z}}$ and its uncertainty $\delta\vec{\mathbf{Z}}$ in a statistical manner. The initial approach used for MT was a standard application of least squares principles (e.g., Sims et al. 1971; Vozoff 1972). In the standard linear regression model for the row-by-row solution of (1), the equivalent set of matrix equations is:

$$\mathbf{e} = \vec{\mathbf{b}} \cdot \mathbf{z} + \boldsymbol{\varepsilon} \quad (2)$$

where there are N observations (i.e., N Fourier transforms of N independent data sections at a given frequency), so that \mathbf{e} is the $N \times 1$ response vector, $\vec{\mathbf{b}}$ is the $N \times 2$ predictor matrix, \mathbf{z} is a two-vector corresponding to a row of the MT response tensor, and $\boldsymbol{\varepsilon}$ is an N -vector of unobservable random errors.

However, early results from least squares frequently exhibited bias and erratic variability with frequency, along with unreliable error estimates. The bias problem was addressed through the remote reference method (Goubau et al. 1978; Gamble et al. 1979) that utilized magnetic field measurements at a secondary site to reduce downward bias by replacing auto-powers with cross-powers in the solution of (2). Variability of the responses and their errors eventually led to the introduction of robust estimators (Egbert and Booker 1986; Chave et al. 1987) to eliminate the influence of outliers and non-stationarity, and quickly became the standard approach. Over time, extensions to standard robust estimators have been introduced, such as the robust remote reference implementation with jackknife error estimates of Chave and Thomson (1989), the multi-site principal components

approach of Egbert (1997, 2002), or the bounded influence estimator of Chave and Thomson (2004). The principles and implementation of such robust estimators for MT data will not be covered in this paper, as the topic was thoroughly reviewed by Chave (2012).

This paper primarily covers more recent work on MT response estimation. The major extension that will be described is due to the serendipitous discovery (Chave 2014) that MT data in the frequency domain are pervasively described by well-understood statistical distributions having algebraic rather than exponential tails. This enables the formulation of a statistically optimal maximum likelihood estimator (MLE) whose performance will be compared to the robust methods that have been in standard use since the late 1980s.

The success of any MT response estimator depends on the fidelity of the spectral analysis that produces the frequency domain data that are its inputs. Chave (2012, Section 5.2) provides a review of the important principles that yield low bias spectral estimates that are statistically independent at a given frequency. At a high level, the spectral analysis steps are based on the Welch overlapped section averaging (WOSA) approach (Percival and Walden 1993, Section 6.17):

1. The raw time series are prewhitened, typically by convolution with an autoregressive filter fit to the time series, to reduce the spectral dynamic range. The autoregressive filter must be computed robustly;
2. Starting at the lowest frequency (or longest period) of interest, a section length N that is of order a few over the frequency of interest is selected. N is typically much smaller than the length of the entire time series;
3. The resolution bandwidth W is selected, and each section is tapered with the lowest order Slepian sequence parameterized by N and W that is an optimal data taper, in the sense that it maximizes the energy concentration in the band $(-W, W)$ about any frequency (Slepian 1978). With adequate prewhitening, the best choice for the time-bandwidth product NW is 1, as it ensures independence of adjacent frequencies on the standard discrete Fourier transform grid. Data sections may be overlapped to improve statistical efficiency, as described by Percival and Walden (1993, Section 6.17);
4. Discrete Fourier transforms of the tapered data sections are taken, and prewhitening and the instrument response are corrected in the frequency domain. In the present context, the data are the Fourier transforms of the windowed time series at a given frequency from each section;
5. The section length is then repetitively reduced as higher frequencies are addressed.

2 The statistical problem

In MT, the predictor variables in $\vec{\mathbf{b}}$ are random variables, and the linear regression statistical model underlying (2) is:

$$\begin{aligned} E(\mathbf{e}|\vec{\mathbf{b}}) &= \vec{\mathbf{b}} \bullet \mathbf{z} \\ cov(\mathbf{e}|\vec{\mathbf{b}}) &= \sigma^2 \vec{\mathbf{I}}_N \end{aligned} \quad (3)$$

where E and cov denote the expected value and covariance, $\mathbf{e}|\vec{\mathbf{b}}$ means \mathbf{e} conditional on $\vec{\mathbf{b}}$, and $\vec{\mathbf{I}}_N$ is the $N \times N$ identity matrix. $E(\mathbf{e}|\vec{\mathbf{b}})$ and $cov(\mathbf{e}|\vec{\mathbf{b}})$ are both random variables

(hereafter rvs); a different result will ensue for each instance of $\vec{\mathbf{b}}$. The theory of least squares applies equally well when $\vec{\mathbf{b}}$ contains rvs under very general conditions (Shaffer 1991). It will further be assumed that the statistical model is linear in the parameters, a condition that holds for MT based on physics [see Eq. (1)], and that \mathbf{z} is not affected by linear equalities among the predictor variables, so that $\text{rank}(\vec{\mathbf{b}}) = 2$.

The least squares estimator for the statistical model (3) is:

$$\hat{\mathbf{z}} = (\vec{\mathbf{b}}^H \bullet \vec{\mathbf{b}})^{-1} \bullet (\vec{\mathbf{b}}^H \bullet \mathbf{e}) \tag{4}$$

where the superscript H denotes the Hermitian transpose and the elements of $\vec{\mathbf{b}}^H \bullet \vec{\mathbf{b}}$ and $\vec{\mathbf{b}}^H \bullet \mathbf{e}$ are averaged estimates of the auto- and cross-power spectra based on the available data. Chave (2017, Section 9.2) defines the properties and definitions that apply to the linear regression estimator, and which in turn motivate robust estimation when they are violated. A subset of these issues that will be further evaluated in Sect. 7 are:

1. The estimator is unconditionally unbiased:

$$E(\hat{\mathbf{z}}) = E[E(\hat{\mathbf{z}} | \vec{\mathbf{b}})] = \mathbf{z} \tag{5}$$

when the random errors $\mathbf{\varepsilon}$ have zero mean and a common variance. These conditions are equivalent to requiring that all of the data have equal influence on the result.

2. The N -vector of predicted values for the response variables is:

$$\hat{\mathbf{e}} = \vec{\mathbf{b}} \bullet \hat{\mathbf{z}} = \vec{\mathbf{b}} \bullet \left(\vec{\mathbf{b}}^H \bullet \vec{\mathbf{b}} \right)^{-1} \bullet \vec{\mathbf{b}}^H \bullet \mathbf{e} \equiv \vec{\mathbf{H}}_N \bullet \mathbf{e} \tag{6}$$

where $\vec{\mathbf{H}}_N$ is the $N \times N$ hat matrix that depends only on the predictor variables. Chave and Thomson (2004) or Chave (2012) describe the properties of the hat matrix.

3. The residuals are given by:

$$\hat{\mathbf{r}} = \mathbf{e} - \vec{\mathbf{b}} \bullet \hat{\mathbf{z}} = \left(\vec{\mathbf{I}}_N - \vec{\mathbf{H}}_N \right) \bullet \mathbf{e} \tag{7}$$

4. An unbiased estimate for σ^2 is:

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{r}}^H \bullet \hat{\mathbf{r}}}{N-2} \tag{8}$$

5. The residuals $\hat{\mathbf{r}}$ are uncorrelated with the predicted values $\hat{\mathbf{e}}$.
6. The parameter vector $\hat{\mathbf{z}}$ is the best linear unbiased estimator for \mathbf{z} . This statement is the Gauss–Markov theorem that underlies linear regression.
7. While statements 1–8 do not depend on distributional assumptions, if $\mathbf{\varepsilon} \sim C_N(\mathbf{0}, \sigma^2 \vec{\mathbf{I}}_N)$ (where \sim specifies “is distributed as”), meaning that the random errors are N -variate complex normal with zero mean and common variance σ^2 (implying that they are identically distributed), then \mathbf{z} is also the maximum likelihood estimate, and hence is

asymptotically Gaussian, so that:

$$\begin{aligned}\hat{\mathbf{z}} &\sim C_2 \left[\mathbf{z}, \sigma^2 \left(\overline{\mathbf{b}} \bullet \overline{\mathbf{b}} \right)^{-1} \right] \\ \hat{\mathbf{e}} &\sim C_N \left(\overline{\mathbf{b}} \bullet \mathbf{z}, \sigma^2 \overline{\mathbf{H}}_N \right) \\ \frac{(N-2)\hat{\sigma}^2}{\sigma^2} &\sim \chi_{N-2}^2\end{aligned}\quad (9)$$

where χ_v^2 is the chi square distribution with v degrees of freedom. Further, $\hat{\mathbf{z}}$ is consistent (i.e., it becomes more tightly concentrated around the population value as the number of data increases), and $\hat{\mathbf{z}}$ and $\hat{\sigma}^2$ are independent, serving as sufficient statistics for estimating \mathbf{z} and σ^2 . The term sufficient statistic means that no other statistic for a parameter calculated from a given data sample can provide additional information about the value of the parameter.

The theory outlined in (3)–(9) strictly applies when $\overline{\mathbf{b}}$, while containing random variables, is measured without error. A more realistic model incorporates measurement error into both \mathbf{e} and $\overline{\mathbf{b}}$ and is called an errors-in-variables model. It has been known since at least the time of Adcock (1878) that the presence of errors in all of the measured variables makes the ordinary linear regression estimator (4) biased. Additional analysis steps are required to minimize bias when an errors-in-variables model obtains, and is addressed by the now standard remote reference method in which the horizontal magnetic field at a reference site \mathbf{b}_r^H replaces the local magnetic field \mathbf{b}^H in (4).

With natural source electromagnetic data, the conditions on the least squares estimate are rarely tenable even when $\overline{\mathbf{b}}$ is measured without error for at least three reasons:

- (a) The variance of the residuals $\hat{\mathbf{r}}$ is often dependent on that of the data, especially when energetic intervals coincide with source field complexity, as is the case for many classes of geomagnetic disturbances;
- (b) The finite duration of many geomagnetic or cultural events results in data anomalies that occur in patches, violating the independent residual requirement;
- (c) Extreme residuals are much more common with MT data than would be expected for a Gaussian model, and hence the residual distribution is very long-tailed, typically with an algebraic rather than exponential character. The data corresponding to large residuals are called influential and can result in serious bias to $\hat{\mathbf{z}}$ and even larger bias in parametric estimates of its uncertainty unless properly accounted for.

Any one of these issues can seriously impact the linear regression solution (4) in unpredictable and sometimes insidious ways; in the presence of more than one, difficulty is guaranteed.

These problems have led to the introduction to statistics, and subsequently MT, of procedures that are robust, in the sense that they are relatively insensitive to a moderate amount of unusual data. The principle underlying all robust estimators is that the data are described by a Gaussian core contaminated by outlying values that are longer-tailed than Gaussian, and hence elimination of the extreme data will yield an estimate based only on the Gaussian data that is consistent with (3)–(9). This model will hereafter be called the

robust model. A key purpose of this paper is evaluation of the validity of this model for MT data.

For the purposes of illustration, an exemplar broadband data set from northwest Namibia consisting of 1,000,000 points sampled at 15 Hz will be utilized. The primary site is located at 20°13'S, 18°20'E, and the reference site is located at 20°36'S, 18°22'E. The x- and y-axes are aligned with local geomagnetic north and east, respectively. These data can reasonably be assumed to be free of cultural noise given the remoteness of these sites, and hence the focus will be on the character of the geomagnetic field used to estimate the MT response.

3 Statistical verification

It is important to devise statistical tools to evaluate the performance of an estimator, and this of necessity means a posteriori tests to validate the results of MT data analysis. In this context, the most useful have proven to be quantile–quantile (q–q) and percent–percent (p–p) plots of the regression residuals against an entity dependent on a target distribution.

Both types of plots compare the statistical distribution of an observed quantity with a theoretical entity and hence provide a qualitative means for assessing the statistical outcome of data processing that can be quantified by placing error bounds on the result. The N quantiles $\{q_j\}$ of a probability distribution divide the area under the probability density function (PDF) into $N + 1$ equal area pieces and hence define equal probability intervals. They are easily obtained from:

$$q_i = F^{-1}[(i - 1/2)/N] \quad i = 1, \dots, N \quad (10)$$

where $F^{-1}(x)$ is the target quantile or inverse cumulative distribution function (cdf). A q–q plot is a comparison of the quantiles with the residual-order statistics $\hat{r}_{(i)}$ obtained by sorting them into an ascending sequence. If the postulated distribution fits the residuals, then a q–q plot will approximate a straight line. Quantile–quantile plots emphasize the distribution tails; most of a q–q plot covers only the last few percent of the distribution range. Consequently, they are quite useful to assess whether the robust model pertains. Further detail may be found in Chave (2017, Section 4.8.4).

Suppose a rv \mathbf{X} has a continuous cdf $F(x)$. If a new random variable is defined as $u = F(x)$, then \mathbf{U} has the uniform distribution on $[0, 1]$, which defines the probability integral transform (Chave 2017, Section 4.8.1). A percent–percent plot consists of the uniform distribution quantiles $u_i = (i - 1/2)/N$ for $i = 1, \dots, N$ against the residual-order statistics transformed using the probability integral transform. If the target cdf is correct for the residuals, then the plot will approximate a straight line. The p–p plot is most sensitive at the mode of the distribution and hence is appropriate for evaluating heavy-tailed distributions, but is not very useful for detecting outliers under the robust model.

A useful variant on the p–p plot was introduced by Michael (1983). The standard p–p plot has its highest variance near the distribution mode and its lowest variance in the distribution tails. Michael applied an arcsine transformation to equalize the variance at all points on a stabilized p–p plot. The variance-stabilized p–p plot consists of:

$$x_i = 2 \sin^{-1}(\sqrt{u_i})/\pi \quad (11)$$

plotted against:

$$\hat{s}_i = \frac{2}{\pi} \sin^{-1} \sqrt{F[(\hat{r}_{(i)} - \hat{\mu})/\hat{\sigma}]} \quad (12)$$

where $\hat{\mu}(\hat{\sigma})$ are location (scale) estimates for the distribution parameters. As with the ordinary p–p plot, a stabilized p–p plot will approximate a straight line if the target distribution is appropriate for the residuals.

The data in a MT context are Fourier transforms and therefore complex, and it is convenient to measure residual size using the magnitude since that quantity is rotationally (phase) invariant. The appropriate distribution for the magnitude of a complex Gaussian number is Rayleigh (Chave 2017, Section 3.4.3). For a residual q–q plot, the N quantiles of the Rayleigh distribution are plotted against the residual absolute value order statistics $|\hat{r}_{(i)}|$.

A robust estimator utilizes data-adaptive weighting that censors a fraction of the data, hence requiring that the quantiles be obtained from the truncated form of the original target distribution, or else the result will inevitably appear to be short-tailed. The truncated distribution is easily obtained from the original one, as described in Chave and Thomson (2003, Appendix C) or Chave (2017, Section 4.8.8).

Whether the data have been censored or not, a straight line q–q or p–p plot indicates that the residual elements are drawn from the target distribution. Data that are inconsistent with the target distribution appear as departures from linearity manifest as sharp upward (downward) shifts in the order statistics at the distribution ends for q–q (p–p) plots. The residual q–q or p–p plots from the output of a robust estimator should be approximately linear or slightly short-tailed if the robust model is correct.

Fig. 1 Quantile–quantile residual plots for the y-orientation at the exemplar site at a period of 45.5 s. The ordinary least squares result is shown in the *top panel*, and the robust result is shown in the *bottom panel*. There are 740 data for the ordinary least squares and 724 data for the robust result

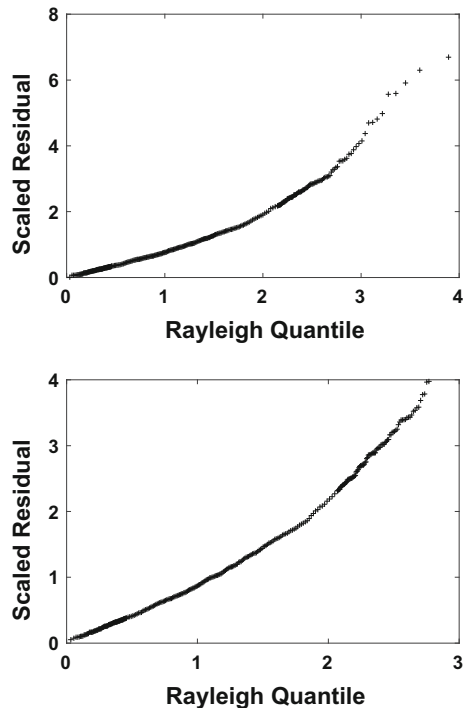
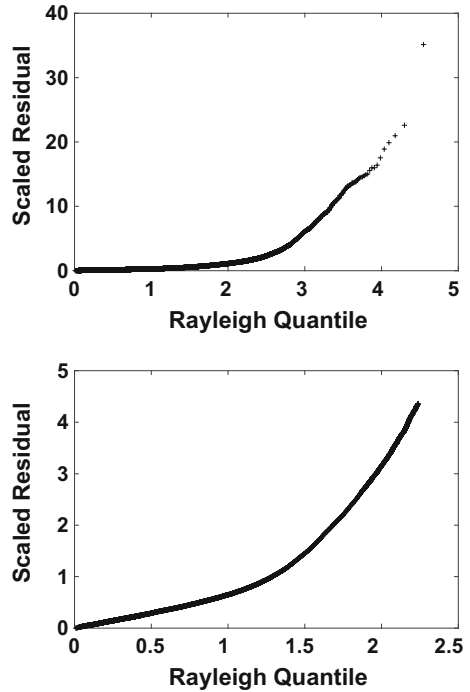


Fig. 2 Quantile–quantile residual plots for the y-orientation at the exemplar site at a period of 2.1 s. The ordinary least squares result is shown in the top panel, and the robust result is shown in the bottom panel. There are 11,866 data for the ordinary least squares and 10,893 data for the robust result



Figures 1 and 2 show q–q plots for the residuals from the y-orientation (i.e., electric field oriented east–west) estimate for the exemplar data set at periods of 45.5 and 2.1 s, respectively. The top panels show the ordinary least squares, and the bottom panels show the robust (scaled by the robust weights) residuals, in both cases using the standard remote reference method to control bias caused by magnetic field noise. The robust q–q plots utilize the truncated form of the Rayleigh distribution to account for data censored by robust weighting. The ordinary least squares q–q plots are long-tailed, but do not suggest the presence of outliers at the top of a Gaussian core. Rather, they suggest a distribution that is systematically longer-tailed than Gaussian. A robust estimator removes the most extreme values (note the different scales for the ordinary and robust cases), but the residual distribution remains systematically long-tailed. This is especially true for Fig. 2 whose period lies within the so-called dead band around 1 s. For both examples, the final residual distributions are not even approximately Rayleigh and are not explained by the robust model of a Gaussian (or Rayleigh) core contaminated by a fraction of outlying data. This effect is present in the q–q plots at all periods for this data set and has also been observed in dozens of additional, globally distributed data sets; see Chave (2014) for additional examples. The observations surrounding Figs. 1 and 2 are not new, but have received little attention over the years, perhaps because robust estimators do appear to dramatically improve MT response estimates over ordinary least squares.

This raises the obvious question “what distribution does describe the regression residuals for MT?” The answer appears to be that they are pervasively described by a distribution family whose tails are algebraic rather than exponential called stable distributions, as reviewed in Chave (2014) or Chave (2017, Section 3.4.2). Power laws such as the stable family approximate the distribution of many natural phenomena, such as the size of earthquakes, volcanic eruptions, solar flares and lightning strikes.

Stable distributions are characterized by a tail thickness parameter α , a skewness parameter β and scale (location) parameters $\gamma(\delta)$ that are analogous to the standard deviation and mean. The tail thickness parameter ranges over $(0, 2]$, with 2 corresponding to a Gaussian distribution, $1 < \alpha < 2$ having infinite variance with increasingly thick tails as α decreases, 1 corresponding to the Cauchy distribution and $0 < \alpha \leq 1$ having infinite mean and variance, along with very thick tails. Except for three special cases, stable distributions cannot be expressed in closed form, and numerical methods are required to compute their properties. Based on the analysis of dozens of data sets by the author, MT data are typically stably distributed with the empirical observation that $0.8 \leq \alpha \leq 1.8$ and $\beta \approx 0$.

In addition to qualitative evaluation using a stabilized p–p plot, it is important to utilize a statistical hypothesis test to assess the fit of a stable model to a given data set. This is most easily achieved using a Kolmogorov–Smirnov test (Chave 2017, Section 7.2.3) to compare the empirical cdf of the residuals to a fitted stable cdf, yielding a test p value. A p value is the random probability of observing a value of the test statistic as large as, or larger than, the one that is observed for a given data set. However, it is well known that the Kolmogorov–Smirnov test p value is biased when the distribution parameters are estimated from the data, as in the present application. The bias can be removed using a permutation method (Chave 2017, Section 8.3) as follows:

1. Fit a stable distribution to the residuals;
2. Obtain a random draw of the same size from the fitted stable distribution;
3. Compute the two-sample Kolmogorov–Smirnov test statistic for the data and random draw;
4. Merge the two data sets;
5. Randomly permute the indices of the merged data set;
6. Compute the two-sample Kolmogorov–Smirnov test statistic as if the first half of the permuted data are the original data and the second half are the random draw;
7. Repeat 5–6 many (in the present application, 10,000) times;
8. Compute a two-tailed p value by comparing the original test statistic to the permutation distribution defined by steps 5–7 in the standard way.

If the data are statistically consistent with the random draws, then the original test statistic should lie near the middle of the permutation distribution, yielding a large p value and hence acceptance of the null hypothesis that the data are stably distributed according to the fitted parameters. Conversely, if the original test statistic lies in one of the permutation distribution tails, then the test will reject. Permutation tests are exact (meaning the probability of a false-positive outcome always has the same value) under very general circumstances.

4 Maximum likelihood estimation for stable MT data

Convergence of the least squares estimator to the true value of the MT response function occurs at the rate $N^{-1/2}$, where N is the number of data. Ordinary least squares applied to stable data converges as $N^{(1-\alpha)/\alpha}$ (McCulloch 1998), and hence is increasingly slow as the tail thickness parameter decreases, failing for $\alpha \leq 1$. However, an iteratively re-weighted least squares estimator for stable data that is analogous to a robust estimator is easily devised and works well provided that α is sufficiently large.

In the absence of a theory for complex-valued stable distributions, an augmented real version of (4) will be used in the remainder of this paper. Let $\mathbf{e}' = (\mathbf{e}^r \mathbf{e}^i)^T$ be a column $2N$ -vector of the real parts of the electric field stacked above the imaginary parts, and let

$$\vec{\mathbf{b}}' = \begin{bmatrix} \vec{\mathbf{b}}^r & -\vec{\mathbf{b}}^i \\ \vec{\mathbf{b}}^i & \vec{\mathbf{b}}^r \end{bmatrix} \tag{13}$$

be the $2N \times 4$ real magnetic field predictor matrix. Then, the real equations:

$$\mathbf{e}' = \vec{\mathbf{b}}' \bullet \mathbf{z}' + \boldsymbol{\varepsilon}' \tag{14}$$

are equivalent to (4), where $\mathbf{z}' = (\mathbf{z}^r \mathbf{z}^i)^T$ is the MT response function 4-vector and $\boldsymbol{\varepsilon}'$ is a $2N$ -vector of random errors.

For stable MT data, the PDF of a single residual (real or imaginary part) is $S(\hat{r}_i | \alpha, \beta, \gamma, \delta)$. For independent data, the sampling distribution is:

$$S_N(\hat{\mathbf{r}} | \alpha, \beta, \gamma, \delta) = \prod_{i=1}^{2N} S(\hat{r}_i | \alpha, \beta, \gamma, \delta) \tag{15}$$

The likelihood function is the sampling distribution (15) regarded as a function of the parameters for a given set of residuals. The MLE is obtained by maximizing the likelihood function, or equivalently, its logarithm:

$$\mathcal{L}(\boldsymbol{\varsigma}, \mathbf{z} | \hat{\mathbf{r}}) = \sum_{i=1}^{2N} \log S(\hat{r}_i | \boldsymbol{\varsigma}) \tag{16}$$

where $\boldsymbol{\varsigma} = (\alpha, \beta, \gamma, \delta)$ is a vector of stable parameters. The first-order conditions for the MLE solution follow by setting the derivatives of (16) with respect to the parameters to zero:

$$\begin{aligned} \partial_{\varsigma_j} \mathcal{L}(\boldsymbol{\varsigma}, \mathbf{z} | \hat{\mathbf{r}}) &= \sum_{i=1}^{2N} \frac{\partial_{\varsigma_j} S(\hat{r}_i | \boldsymbol{\varsigma})}{S(\hat{r}_i | \boldsymbol{\varsigma})} = 0 \quad j = 1, \dots, 4 \\ \partial_{z_k} \mathcal{L}(\boldsymbol{\varsigma}, \mathbf{z} | \hat{\mathbf{r}}) &= - \sum_{i=1}^{2N} \frac{\partial_{z_k} S(\hat{r}_i | \boldsymbol{\varsigma})}{S(\hat{r}_i | \boldsymbol{\varsigma})} b_{ik} = 0 \quad k = 1, \dots, 4 \end{aligned} \tag{17}$$

The sufficient condition for the solution of (17) to be a maximum is that the Hessian matrix of the log likelihood function be negative definite. Equations (17) will be solved using an iterative two-stage process that decouples its two types of parameters. In the first stage, the stable distribution parameter vector $\boldsymbol{\varsigma}$ is estimated using the stable MLE algorithm of Nolan (2001). In the second stage, the MLE response function $\hat{\mathbf{z}}$ is computed using these values for the stable distribution parameters. Iteration between the two stages then continues until convergence is achieved.

Let $\lambda(\hat{r}_i) = \log S(\hat{r}_i | \boldsymbol{\varsigma})$ where $\boldsymbol{\varsigma}$ is assumed known. The second equation in (17) can be rewritten as:

$$- \sum_{i=1}^{2N} \lambda'(\hat{r}_i) b_{ik} = - \sum_{i=1}^{2N} \frac{\lambda'(\hat{r}_i)}{\hat{r}_i} \left(e_i b_{ik} - \sum_{j=1}^4 b_{ij} \hat{z}_j b_{ik} \right) = 0 \quad k = 1, \dots, 4 \tag{18}$$

where $\lambda'(x)$ is called the score function. Equation (18) may be recast in matrix form to yield the iteratively re-weighted MLE solution:

$$\hat{\mathbf{z}}_{ML} = \left(\bar{\mathbf{b}}^H \bullet \bar{\mathbf{w}} \bullet \bar{\mathbf{b}} \right)^{-1} \bullet \left(\bar{\mathbf{b}}^H \bullet \bar{\mathbf{w}} \bullet \mathbf{e} \right) \quad (19)$$

where $\bar{\mathbf{w}}$ is a diagonal weight matrix whose i th element is $-\lambda'(\hat{r}_i)/\hat{r}_i$. A remote magnetic field reference is easily implemented by replacing $\bar{\mathbf{b}}$ with $\bar{\mathbf{b}}_r$ in (19), where $\bar{\mathbf{b}}_r$ is the magnetic field at a reference site. Equation (19) is very similar in form to a robust estimator. However, there is one very important distinction: The weights for the stable MLE are based on a statistical model for the residuals that is derived directly from them, and whose statistical significance can be evaluated, while robust estimator weights are *ad hoc*, being based on the robust model that is inconsistent with actual MT data.

As proof that (19) reduces the influence of outlying data, Fig. 3 shows the weight function $\bar{\mathbf{w}}$ for a standardized ($\gamma = 1, \delta = 0$) symmetric ($\beta = 0$) stable distribution for several values of the tail thickness parameter α compared to a robust estimator weight function as used in the last stage of the Chave and Thomson (2004) algorithm (as implemented in the widely used BIRRP code). From (19), the solution is unaffected by scaling the weight function by a constant, so the relative size of the curves in Fig. 3 is irrelevant. For the Gaussian end member ($\alpha = 2$), the weights are independent of the size of the residuals and yield the ordinary least squares solution. As α decreases, the weight function becomes increasingly peaked at the origin and falls off more rapidly with residual size. By comparison, the Chave and Thomson robust estimator utilizes a weight function that is constant between $\pm\xi$, where $|\xi|$ usually lies between 3 and 5 in scaled units and then descends to zero for larger values; it is shown as a boxcar function for $\xi = 4$ in Fig. 3. It is readily apparent that, for data with a tail thickness parameter lying below 2 that is typical for MT, the robust estimator consistently overemphasizes data corresponding to large residuals within the pass band relative to those at the center when compared to the stable MLE estimator and does not utilize any data beyond the cutoff point. In fact, the robust estimator simply truncates the stable distribution without adapting to its form. This

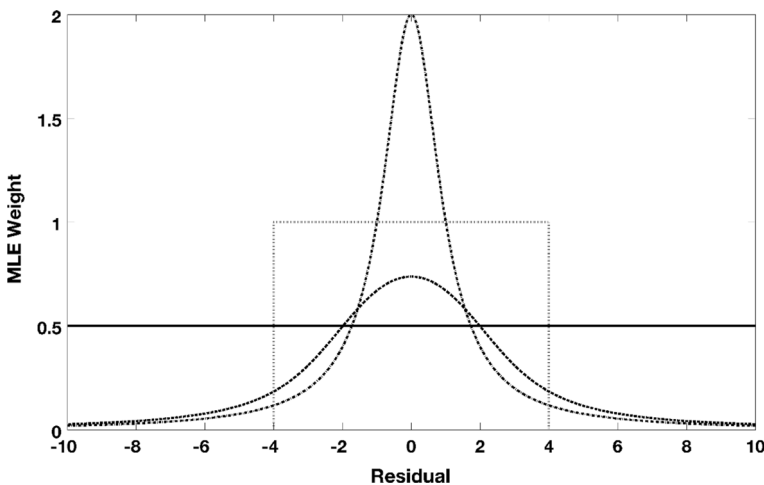


Fig. 3 MLE weight function for a standardized ($\gamma = 1, \delta = 0$), symmetric ($\beta = 0$), stable distribution for tail thickness parameter α values of 2 (solid line), 1.5 (dashed line) and 1.0 (dot-dash line). The dotted boxcar function is a typical robust weight function with a cutoff of 4 on the abscissa. See text for discussion. Taken from Chave (2014)

has implications for the bias and variance properties of the robust estimator that will be elaborated in the sequel.

Figure 4 shows the MLE weights for a skewed stable distribution ($\beta = -0.1$) for the same values of α as in Fig. 3. In this case, the weights are asymmetric around the origin, with large negative (positive) values for positive (negative) residuals close to it. The conventional robust estimator is insensitive to skewness, and so its weight function would appear as in Fig. 3.

An iteratively re-weighted least squares MLE is easily implemented by initiation using the ordinary least squares solution with data corresponding to the 5% most extreme residuals at both distribution ends trimmed to reduce the influence of actual outliers. The estimated residuals are fit with stable parameters using the MLE algorithm of Nolan (2001), standardized as $(\hat{r}_i - \delta)/\gamma$ to ensure scale independence and then applied to computation of the weights in (19). The weighted least squares problem (19) is solved, the residuals are subsequently used to re-compute the stable parameters, and the process is repeated iteratively until the median absolute deviation from the median, or MAD, of the iteration residuals does not change by more than 1%. This typically takes 3–9 iterations, with the number increasing as the tail thickness parameter approaches unity from above due to the efficiency properties of the least squares estimator. The iteratively re-weighted solution has the advantage of being relatively fast, but has the disadvantage of slow convergence for $\alpha \rightarrow 1$ and failure below that value, which is a significant practical limitation.

An alternative MLE solution that does not suffer from the convergence issue, and is similar to the algorithm described by Nolan and Ojeda-Revah (2013), has also been implemented. As with the iteratively re-weighted least squares solution, this utilizes a two-stage approach initialized by the 5% trimmed least squares solution, with the Nolan (2001) MLE used at the first stage to compute the stable parameters, and an unconstrained

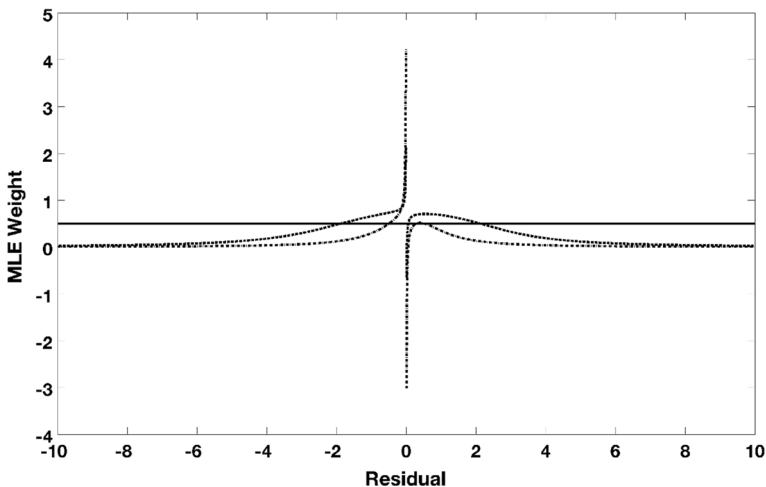


Fig. 4 MLE weight function for a standardized ($\gamma = 1$, $\delta = 0$), skewed ($\beta = -0.1$), stable distribution for tail thickness parameter α values of 2 (solid line), 1.5 (dashed line) and 1.0 (dot-dash line). Taken from Chave (2014)

nonlinear multivariable function minimizer based on a trust region algorithm (Conn et al. 2000) utilized for the second stage. Alternation between the two stages ensures convergence, which typically requires no more than 2–3 iterations. Numerical solutions for the gradient and Hessian matrix are implemented to speed convergence for the second stage. The objective function that is minimized is the negative log likelihood given by (16) preceded by a minus sign, and with the stable parameter vector ζ fixed. Remote referencing is easily implemented by using two-stage regression as described by Chave and Thomson (2004) or Chave (2012), in which the transfer functions between the local magnetic field components and all of the reference ones are first estimated using the nonlinear MLE, and then the magnetic field values predicted via the transfer functions are used to replace the local magnetic field to get the MT response function. This method works well for all values of the tail thickness parameter, but requires substantially more computer time compared to the iteratively re-weighted least squares approach.

The MT processing algorithms described in this section share all of the standard optimality properties of the MLE (Stuart et al. 1999, chap. 17–18):

1. The solution is asymptotically consistent, meaning that it converges in probability to the true solution, or $\Pr(|\hat{\mathbf{z}} - \mathbf{z}| > \tau) \rightarrow 0$ as $N \rightarrow \infty$ for every positive τ . It is asymptotically unbiased, but not necessarily unbiased for finite N . Further, the finite sample bias is $O(1/N)$ (Stuart et al. 1999, sec. 18.14) and hence is typically negligible for the sample sizes used in MT. This statement does not provide any insight into the rate of convergence (i.e., the number of data required to reach the asymptotic limit), which will be further addressed in Sect. 7;
2. The solution is equivariant, meaning that if $\hat{\mathbf{z}}$ is the MLE, then $f(\hat{\mathbf{z}})$ is the MLE for $f(\mathbf{z})$, where f is some function. For example, this means that the apparent resistivity computed from a MLE MT response is also the MLE for that parameter;
3. The solution is asymptotically Gaussian, meaning that the difference between the estimated and true MT response converges in distribution to $N_4\left(\mathbf{0}_4, \bar{\mathcal{I}}^{-1}\right)$ as $N \rightarrow \infty$, where N_4 is the four-variate Gaussian distribution, $\mathbf{0}_4$ is a vector of four zeroes, and $\bar{\mathcal{I}}$ is the Fisher information matrix defined in Sect. 5;
4. The solution is asymptotically efficient, meaning roughly that its sampling distribution is the most tightly coupled around \mathbf{z} of all possible sampling distributions. An equivalent statement is that the covariance of the MLE achieves the lower bound of the Cramér–Rao inequality described in Sect. 5.

It follows directly that a conventional robust estimator may not be in compliance with these properties since it cannot be the MLE unless (9) pertains, which is contradicted by the discussion around Figs. 1 and 2. In particular, violation of the optimality properties (1) and (2) may result in asymptotic (and hence finite sample) bias of the MT response function and derived quantities like apparent resistivity, violation of (3) means that parametric estimates for MT response uncertainty will be larger than those for the MLE, and violation of (4) means that data are being wasted. How important these issues are can only be determined with actual data and is addressed in Sect. 6.

5 Uncertainty estimates on the MLE response

A complex random vector $\mathbf{x} = \mathbf{x}_r + i\mathbf{x}_i$ has 2 column elements, where \mathbf{x}_r and \mathbf{x}_i are, respectively, the real and imaginary parts. The probability distribution of a complex random vector is the joint distribution of the real and imaginary parts and hence is at least bivariate. Let a complex random vector possess N data rows. Its second-order statistics are described by the covariance and pseudo-covariance matrices (e.g., van den Bos 1995; Picinbono 1996):

$$\begin{aligned} \vec{\Gamma} &= E[(\mathbf{x} - \boldsymbol{\mu}_x)^H(\mathbf{x} - \boldsymbol{\mu}_x)] = \vec{\Gamma}_{x_r x_r} + \vec{\Gamma}_{x_i x_i} + i\left(\vec{\Gamma}_{x_r x_i}^T - \vec{\Gamma}_{x_r x_i}\right) \\ \overleftarrow{\Gamma} &= E[(\mathbf{x} - \boldsymbol{\mu}_x)^T(\mathbf{x} - \boldsymbol{\mu}_x)] = \vec{\Gamma}_{x_r x_r} - \vec{\Gamma}_{x_i x_i} + i\left(\vec{\Gamma}_{x_r x_i}^T + \vec{\Gamma}_{x_r x_i}\right) \end{aligned} \tag{20}$$

The covariance matrix $\vec{\Gamma}$ is complex and Hermitian and will be assumed positive definite, while the pseudo-covariance matrix $\overleftarrow{\Gamma}$ is complex and symmetric. The terms on the right-hand side of (20) are the real covariance matrices for the indexed components.

A complex random vector \mathbf{x} is proper if the pseudo-covariance $\overleftarrow{\Gamma}$ is identically zero, and otherwise it is improper. The complex covariance matrix for proper data is then given by

the first equation in (20). However, in the improper case, both $\vec{\Gamma}$ and $\overleftarrow{\Gamma}$ are required for a complete description of the second-order statistics. The existence of impropriety is a consequence of non-stationarity of the underlying physical processes; see Haykin et al. (2009) or Schreier and Scharf (2010, Chapter 9) for details.

Extending statistical definitions from the real to the complex case utilizes the concept of the so-called augmented variables and covariance matrices. The augmented complex random variable for \mathbf{x} is given by $\underline{\mathbf{x}} = (\mathbf{x} \ \mathbf{x}^*)$, where the $*$ denotes the complex conjugate, and is obtained by adding the columns of \mathbf{x}^* to those of \mathbf{x} . Its elements \mathbf{x} and \mathbf{x}^* are clearly not independent, but augmented variables simplify statistical algebra. The augmented covariance matrix is given by:

$$\underline{\underline{\Gamma}} = \begin{bmatrix} \vec{\Gamma} & \overleftarrow{\Gamma} \\ \overleftarrow{\Gamma}^* & \vec{\Gamma} \end{bmatrix} \tag{21}$$

$\underline{\underline{\Gamma}}$ is block-structured and Hermitian and will be assumed positive definite.

The standard approach to estimation of the variance of an estimator is the use of the Fisher score to compute the Cramér–Rao lower bound. The principles are covered in Chapter 17 of Stuart et al. (1999), and its extension to complex random variables is reviewed in Schreier and Scharf (2010, Chapter 6). Define the complex Fisher score given by the gradient of the log likelihood:

$$\boldsymbol{\lambda}(\boldsymbol{\Psi}|\boldsymbol{\varepsilon}) = \hat{\partial}_{\boldsymbol{\Psi}} \mathcal{L}(\boldsymbol{\Psi}|\hat{\mathbf{r}}) \tag{22}$$

where $\boldsymbol{\Psi} = (\boldsymbol{\zeta}, \mathbf{z})$ is the parameter vector. It is straightforward to show that the Fisher score is a random variable with zero mean. The 8×8 Fisher information matrix for the stable MLE is given by:

$$\vec{\mathcal{I}}(\boldsymbol{\Psi}) = E \left\{ [\partial_{\boldsymbol{\Psi}} \mathcal{L}(\boldsymbol{\Psi} | \hat{\mathbf{r}})]^H [\partial_{\boldsymbol{\Psi}} \mathcal{L}(\boldsymbol{\Psi} | \hat{\mathbf{r}})] \right\} = -E \left\{ \partial_{\boldsymbol{\Psi}} [\partial_{\boldsymbol{\Psi}} \mathcal{L}(\boldsymbol{\Psi} | \boldsymbol{\varepsilon})]^H \right\} \quad (23)$$

For an unbiased estimator of $\boldsymbol{\Psi}$, the covariance matrix is bounded by:

$$\text{cov}(\boldsymbol{\Psi}) \geq \vec{\mathcal{I}}^{-1}(\boldsymbol{\Psi}) \quad (24)$$

Equation (24) is the Cramér–Rao inequality when the MT responses are proper complex random variables.

In the presence of impropriety, the augmented Fisher information matrix (21) replaces the ordinary one and is given by:

$$\vec{\mathcal{I}}(\boldsymbol{\Psi}) = \begin{bmatrix} \vec{\mathcal{I}}(\boldsymbol{\Psi}) & \vec{\mathcal{I}}(\boldsymbol{\Psi}) \\ \vec{\mathcal{I}}^*(\boldsymbol{\Psi}) & \vec{\mathcal{I}}^*(\boldsymbol{\Psi}) \end{bmatrix} \quad (25)$$

where the Fisher pseudo-information matrix is:

$$\vec{\mathcal{I}}(\boldsymbol{\Psi}) = E \left\{ [\partial_{\boldsymbol{\Psi}} \mathcal{L}(\boldsymbol{\Psi} | \boldsymbol{\varepsilon})]^T [\partial_{\boldsymbol{\Psi}} \mathcal{L}(\boldsymbol{\Psi} | \boldsymbol{\varepsilon})] \right\} = -E \left\{ \partial_{\boldsymbol{\Psi}} [\partial_{\boldsymbol{\Psi}} \mathcal{L}(\boldsymbol{\Psi} | \boldsymbol{\varepsilon})]^T \right\} \quad (26)$$

The Cramér–Rao inequality for improper complex random variables becomes:

$$\text{cov}(\boldsymbol{\Psi}) \geq \vec{\mathcal{I}}^{-1}(\boldsymbol{\Psi}) \quad (27)$$

where \geq means the right-hand side is more positive definite than the left-hand side. The block matrix inversion lemma reduces (27) to:

$$\text{cov}(\boldsymbol{\Psi}) \geq \left\{ \vec{\mathcal{I}}(\boldsymbol{\Psi}) - \vec{\mathcal{I}}(\boldsymbol{\Psi}) \left[\vec{\mathcal{I}}^*(\boldsymbol{\Psi}) \right]^{-1} \vec{\mathcal{I}}^*(\boldsymbol{\Psi}) \right\}^{-1} \quad (28)$$

The right-hand side of (28) is larger (i.e., more positive definite) than $\vec{\mathcal{I}}^{-1}(\boldsymbol{\Psi})$; hence, not accounting for impropriety will result in underestimation of the covariance. Equation (28) yields a covariance matrix with a block structure, with the upper left block comprising the covariance matrix of the stable parameters, the lower right block containing the covariance matrix of the MT response functions, the upper right block yielding the covariance of the stable parameters and MT responses, and the lower left block being the Hermitian transpose of the upper right block. Under the two-stage approach used in this paper, the stable and MT parameters are assumed to be uncoupled, and hence the proper Cramér–Rao bound or (28) may be regarded as block diagonal, with the upper left block defined in Nolan (2001). As a consequence, the focus will be on the lower right block. All of the equations in this section will henceforth be regarded as pertaining only to the lower right block. Chave (2014) defines a likelihood ratio test for impropriety of the MT tensor elements, but since (28) reduces to (24) for proper complex variables, it is simpler to routinely use the improper result in practice. In any case, impropriety is pervasive for MT data, and so its neglect carries with it the risk of systematic underestimation of the covariance matrix.

Implementation of the Cramér–Rao bound for the MT MLE covariance matrix is straightforward. The Fisher information matrix $\vec{\mathcal{I}}_R$ for the MT responses in (14) is given by the Hessian matrix:

$$\vec{\vec{\mathcal{I}}}_R = -\mathbf{b} \bullet \vec{\vec{\mathbf{v}}} \bullet \mathbf{b} \tag{29}$$

where $\vec{\vec{\mathbf{v}}}$ is a diagonal matrix whose i th element is the second derivative of the log likelihood $\lambda''(\hat{r}_i)$. $\vec{\vec{\mathcal{I}}}_R$ is real and block-structured, containing the Fisher information for the real and imaginary parts in the upper left and lower right blocks, and that between the real and imaginary components in the two off-diagonal blocks. Its elements can be combined to yield the complex Fisher information matrices $\vec{\vec{\mathcal{I}}}$ and $\vec{\vec{\mathcal{I}}}$ through:

$$\begin{aligned} \vec{\vec{\mathcal{I}}} &= \vec{\vec{\mathcal{I}}}_{x_r, x_r} + \vec{\vec{\mathcal{I}}}_{x_i, x_i} + i \left(\vec{\vec{\mathcal{I}}}_{x_r, x_i}^T - \vec{\vec{\mathcal{I}}}_{x_r, x_i} \right) \\ \vec{\vec{\mathcal{I}}} &= \vec{\vec{\mathcal{I}}}_{x_r, x_r} - \vec{\vec{\mathcal{I}}}_{x_i, x_i} + i \left(\vec{\vec{\mathcal{I}}}_{x_r, x_i}^T + \vec{\vec{\mathcal{I}}}_{x_r, x_i} \right) \end{aligned} \tag{30}$$

The improper covariance bound (28) then follows directly.

6 Example

Unconstrained estimation of the four stable parameters for the residuals using the exemplar data set shows that the skewness β is not statistically distinguishable from zero, and so that parameter was fixed at this value for further analysis. Figure 5 shows the MLE tail thickness parameter α as a function of the base 10 logarithm of the period in seconds, along with double-sided 95% confidence limits after apportioning a 0.05 tail probability among the three stable parameters that were estimated. For the x-orientation, the four longest period estimates are indistinguishable from the Gaussian end member, and then the tail thickness declines with decreasing period to about 1.35 in the dead band below 10 s. The y-orientation displays tail thickness parameters that are somewhat smaller than for the

Fig. 5 Tail thickness parameters for the exemplar data computed after forcing the skewness to be zero. The black error bars are for the x-orientation, and the gray error bars are for the y-orientation; the latter is offset 5% to the right in log space for clarity. The double-sided confidence intervals use the Gaussian quantile 2.39 corresponding to dividing the tail probability of 0.05 among the three parameters that were estimated

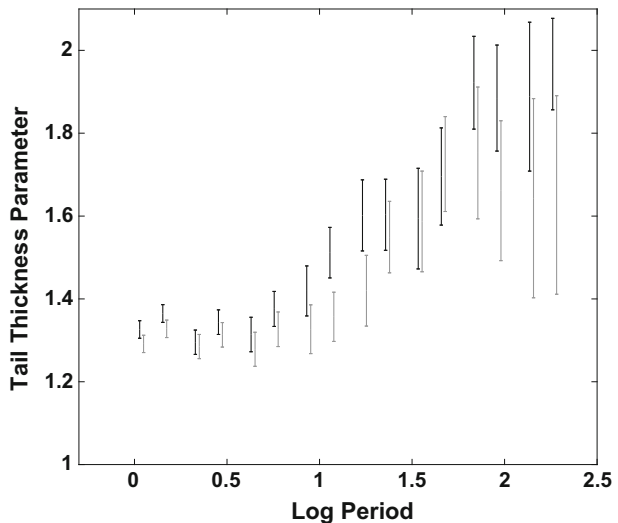


Fig. 6 Variance-stabilized stable MLE residual p–p plots for the x-orientation at 68.3 s (*top*) and 1.4 s (*bottom*). The *dashed lines* are the 95% (*top*) and 99% (*bottom*) confidence bounds on the p–p plot, as described in Michael (1983). The corresponding bias-corrected p values are given in Table 1

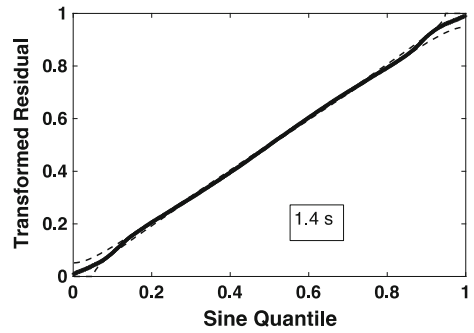
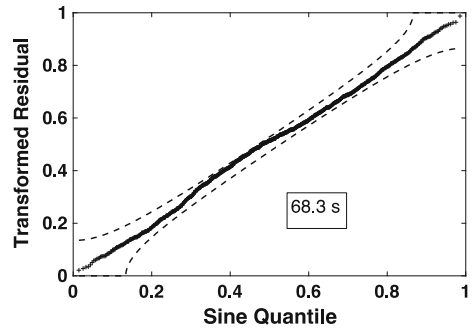


Fig. 7 Variance-stabilized stable MLE residual p–p plots for the y-orientation at 45.5 s (*top*) and 2.1 s (*bottom*). The *dashed lines* are the 95% (*top*) and 99% (*bottom*) confidence bounds on the p–p plot, as described in Michael (1983). The corresponding bias-corrected p values are given in Table 1

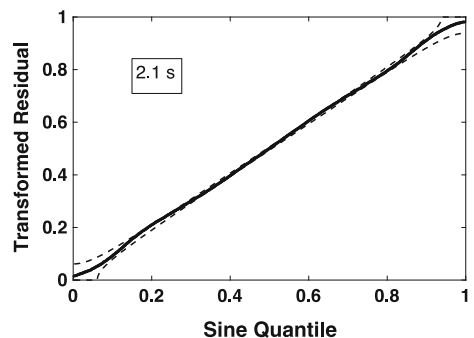
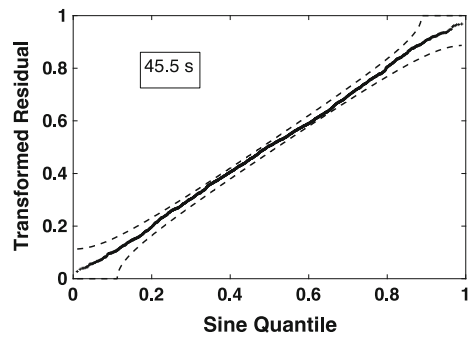


Table 1 Bias-corrected p values for Kolmogorov–Smirnov test for stable distribution

| Period (s) | p_x | p_y |
|------------|--------|-------|
| 182.0 | 0.0032 | 0.213 |
| 136.5 | 0.774 | 0.580 |
| 91.0 | 0.154 | 0.991 |
| 68.3 | 0.195 | 0.501 |
| 45.5 | 0.590 | 0.411 |
| 34.1 | 0.522 | 0.767 |
| 22.8 | 0.063 | 0.014 |
| 11.4 | 0.621 | 0.832 |
| 8.5 | 0.709 | 0.430 |
| 7.1 | 0.851 | 0.902 |
| 5.7 | 0.023 | 0.753 |
| 4.3 | 0.816 | 0.856 |
| 2.8 | 0.358 | 0.293 |
| 2.1 | 0.135 | 0.127 |
| 1.4 | 0.154 | 0.255 |
| 1.1 | 0.095 | 0.104 |

x-orientation, also decreasing into the dead band. Similar behavior was observed for a different broadband data set in Chave (2014), although in that case α went as low as 0.8.

Figures 6 and 7 show variance-stabilized p–p plots for the x- and y-orientations at selected periods outside and inside the dead band, along with double-sided confidence

Fig. 8 Kernel density estimator permutation distributions for the Kolmogorov–Smirnov statistic (solid lines) for the x-orientation at 68.3 s (top) and 1.4 s (bottom) computed using the algorithm of Sect. 3. The vertical dashed line is the Kolmogorov–Smirnov statistic computed from the stable MLE residuals and a random draw with MLE parameters prior to permuting the indices. The two-sided p value is twice the smaller of the area to the left or right of the dashed line and is specified in each panel

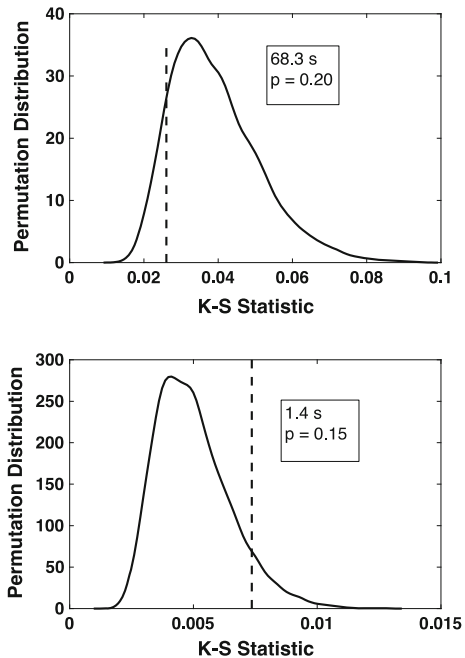


Fig. 9 Kernel density estimator permutation distributions for the Kolmogorov–Smirnov statistic (*solid lines*) for the y -orientation at 45.5 s (*top*) and 2.1 s (*bottom*) computed using the algorithm of Sect. 3. The *vertical dashed line* is the Kolmogorov–Smirnov statistic computed from the stable MLE residuals and a random draw with MLE parameters prior to permuting the indices. The two-sided p value is twice the smaller of the area to the left or right of the *dashed line* and is specified in each panel

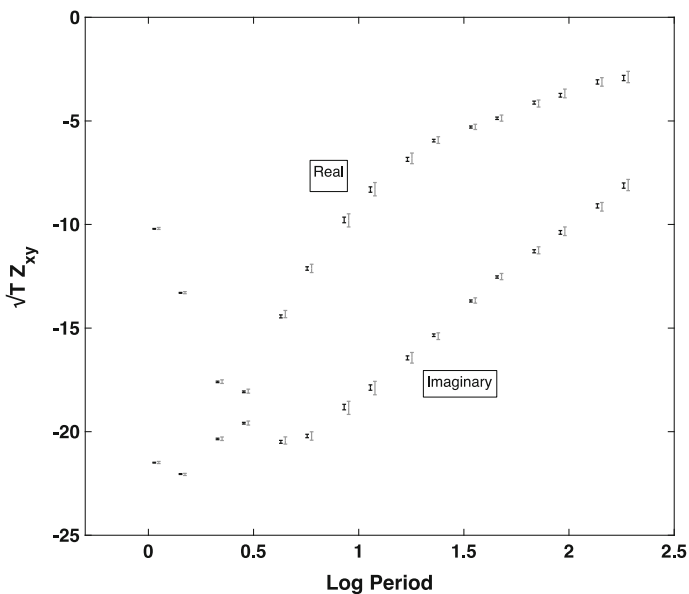
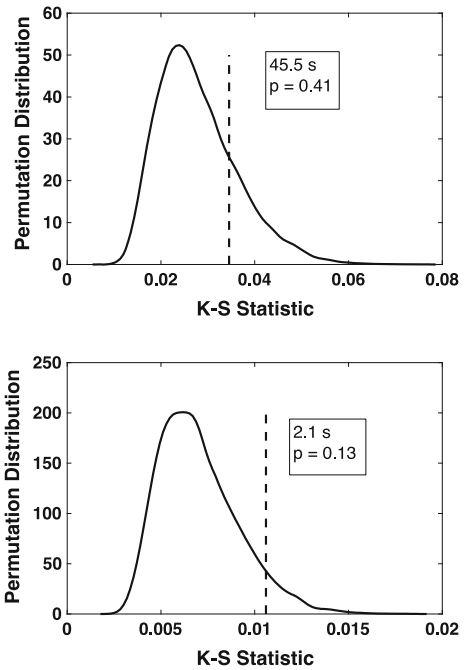


Fig. 10 Real and imaginary parts of the Z_{xy} component of the MT response tensor scaled by the square root of period computed using (*black*) the stable MLE and (*gray*) the robust remote reference estimator described in Chave (2012). The latter are offset to the right by 5% in log space for clarity

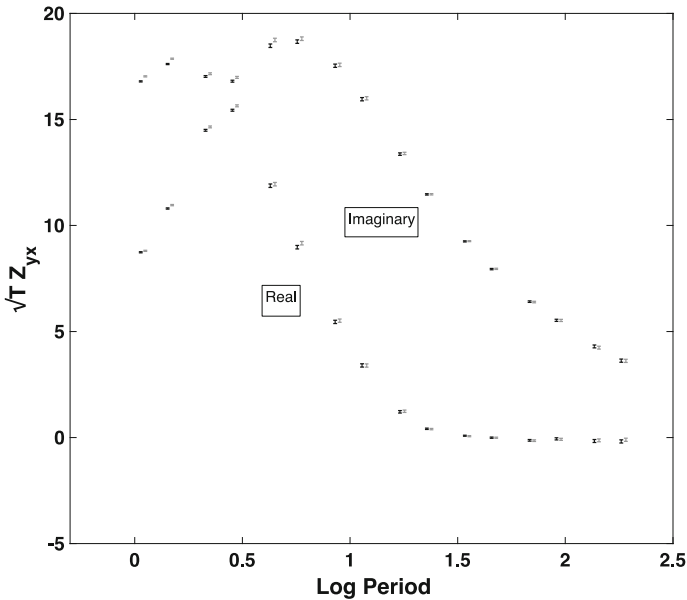


Fig. 11 Real and imaginary parts of the Z_{yx} component of the MT response tensor scaled by the square root of period computed using (*black*) the stable MLE and (*gray*) the robust remote reference estimator described in Chave (2012). The latter are offset to the right by 5% in log space for clarity

bounds obtained using the critical value of the Kolmogorov–Smirnov statistic, as described in Chave (2017, Section 7.2.3). The p–p plots are nearly straight lines that lie within the confidence bounds. There is some slight variability near the distribution center at 68 s for the x-orientation and at the ends of the distribution for both orientations. These figures provide qualitative support for the stable model that will be further explored quantitatively. Similar results are obtained at the remaining periods for the exemplar data set.

A Kolmogorov–Smirnov test for goodness of fit to a stable distribution with MLE parameters was applied to the final residuals at each period and bias-corrected as described in Sect. 3. Table 1 lists the p values by period. Figures 8 and 9 show exemplar kernel density estimator (Chave 2017, Section 4.8.3) permutation distributions along with the original test statistic that is used to compute a bias-corrected p value at the same periods and for the same orientations as in Figs. 6 and 7. In each case, the permutation p value is well above a 0.1 threshold, indicating no support for the alternate hypothesis that the residuals are not stably distributed. Table 1 shows that the null hypothesis that the residuals are stable with parameters given by its MLE is strongly rejected ($p < 0.01$) only for the x-orientation at the longest period, and weakly rejected ($0.01 < p < 0.05$) at one additional period for each orientation. Taken in aggregate, Figs. 6, 7, 8 and 9 and Table 1 provide strong affirmation for a stable model for the residuals with these MT data.

Figures 10 and 11 show the comparison between the Z_{xy} and Z_{yx} components (scaled by the square root of period for clarity) of the stable MLE and robust remote reference estimates at the exemplar site. Uncertainty estimates were obtained using the diagonal elements of the improper Cramér–Rao bound (28) for the MLE response and the jackknife for the robust response, respectively, in each case utilizing the Gaussian quantile 2.73 corresponding to apportioning the 0.05 tail probability among all 8 real and imaginary MT

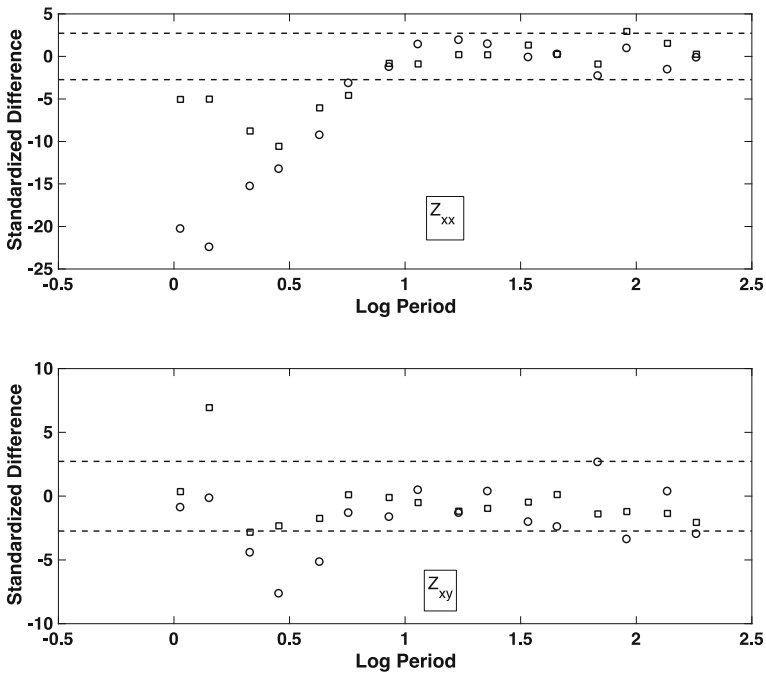


Fig. 12 Difference between the stable MLE and robust remote reference estimates divided by the standard error for the former computed as the square root of the diagonal elements of the improper covariance matrix defined in Sect. 5 for Z_{xx} (top) and Z_{xy} (bottom) plotted against log period. The real parts are shown as circles, and the imaginary parts as squares. The horizontal dashed lines are the Bonferroni 95% confidence bounds after apportioning the 0.05 tail probability among all 8 of the MT tensor elements and define the region within which 95% of the estimates should lie if Gaussian statistics pertain

response components at each period using the Bonferroni method (Chave 2017, Section 5.5). It is obvious that the stable MLE confidence bounds are systematically smaller than the robust ones, especially in Fig. 10. There also are frequent differences between the two types of estimates that represent bias of one of the estimators. This is especially apparent at periods below 10 s for the y-orientation in Fig. 11.

The differences shown in Figs. 10 and 11 are not subtle, but are obscured by the small size of the confidence intervals and the wide range of the scaled MT response. Figures 12 and 13 present the difference between the stable MLE and robust remote reference estimates normalized by the stable MLE standard deviation for all 4 elements of the x- and y-orientations of the MT response tensor. The ordinate units are in MLE standard deviations; the dashed horizontal lines represent ± 2.73 standard deviations within which 95% of the data (or ~ 30 values per panel) should be situated if a Gaussian model pertains. There are 32 real and imaginary parts in each panel of Figs. 12 and 13, but 13, 6, 11 and 11 for the Z_{xx} through Z_{yy} elements actually lie outside the dashed horizontal lines. Further, the differences are as large as 32 standard deviations. Such frequent and large differences are extremely unlikely if all of the entities used in the figures are Gaussian. In addition, the differences are systematic rather than random, being concentrated at short periods, with increasing size as the period decreases.

Figures 14 and 15 show the comparison of the same stable MLE estimates to a two-stage robust estimate, in which the transfer functions between the local and remote

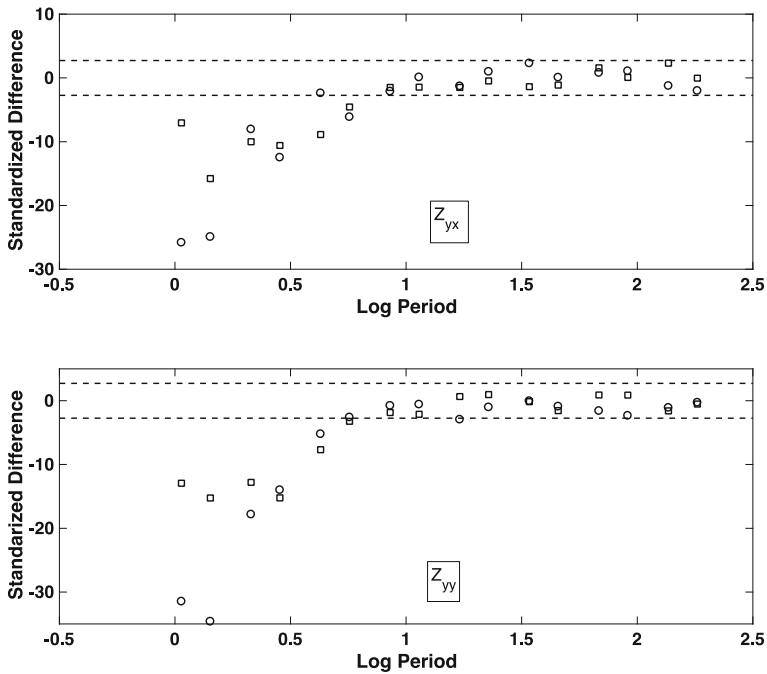


Fig. 13 Difference between the stable MLE and robust remote reference estimates divided by the standard error for the former computed as the square root of the diagonal elements of the improper covariance matrix defined in Sect. 5 for Z_{yx} (top) and Z_{yy} (bottom) plotted against log period. The real parts are shown as circles, and the imaginary parts as squares. The horizontal dashed lines are the Bonferroni 95% confidence bounds after apportioning the 0.05 tail probability among all 8 of the MT tensor elements and define the region within which 95% of the estimates should lie if Gaussian statistics pertain

magnetic fields are first computed robustly and then the projection of the remote magnetic field is used in place of the local values to compute the MT response tensor; see Chave and Thomson (2004) or Chave (2012) for further details. While differences between the two estimators are apparent, they are substantially smaller and less frequent than in Figs. 12 and 13. There are 3, 2, 9 and 7 real or imaginary parts for Z_{xx} through Z_{yy} lying outside the dashed horizontal lines in the figures, and the differences are up to six standard deviations. In fact, the x-orientation result in Fig. 12 is almost consistent with a Gaussian model, although the peak differences are large considering the number of data.

7 Discussion

The differences between the stable MLE and robust remote reference estimates in Figs. 12 and 13 are both too frequent and too large to be consistent with Gaussian behavior of all of the statistical entities they incorporate. Further, the differences are systematic rather than random, with the deflections increasing with decreasing period into the dead band. Over the same period range, the tail thickness parameter is also systematically decreasing (Fig. 5). A similar observation for a different data set with much heavier residual tails was reported in Chave (2014). However, these effects are partially ameliorated in the present case by using a two-stage robust method (Figs. 14, 15), suggesting that noise in the local

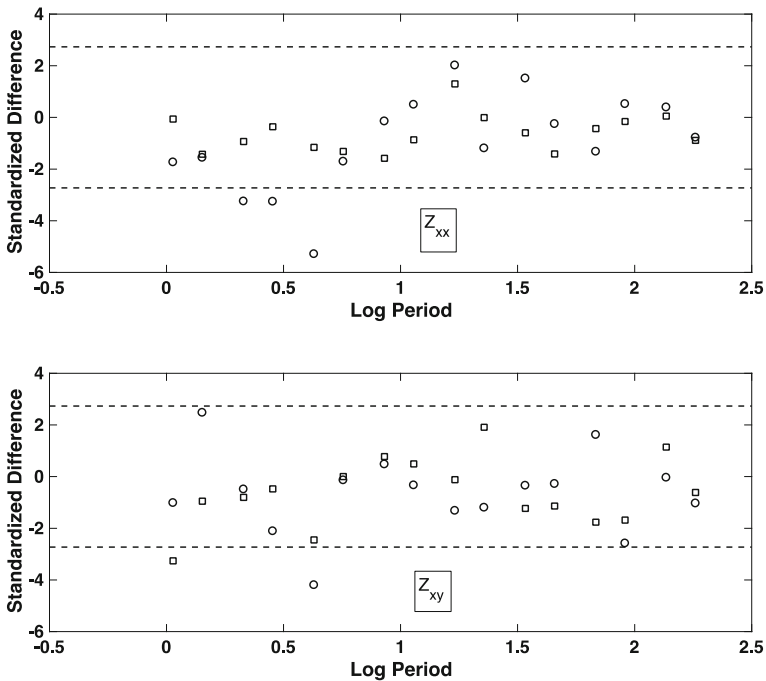


Fig. 14 Difference between the stable MLE and two-stage robust estimates divided by the standard error for the former computed as the square root of the diagonal elements of the improper covariance matrix defined in Sect. 5 for Z_{xx} (top) and Z_{xy} (bottom) plotted against log period. The real parts are shown as circles, and the imaginary parts as squares. The horizontal dashed lines are the Bonferroni 95% confidence bounds after apportioning the 0.05 tail probability among all 8 of the MT tensor elements and define the region within which 95% of the estimates should lie if Gaussian statistics pertain

magnetic field is removed by the two-stage approach, but not by the simpler remote reference method. Nevertheless, both the frequency and size of the differences remain larger than would be expected for Gaussian entities, especially for the y-orientation. However, the two-stage approach is not a panacea, as it fails for the broadband example in Chave (2014).

In Sect. 2, the statistical model underlying linear regression (3)-(4) was presented. Insight into the performance of the robust and stable MLE estimators ensues from evaluation of the bias condition on the model given by (5). This specifies that an estimator is unconditionally unbiased provided that the random errors have zero mean and a common variance. The random errors are unobservable, but the residuals (7) serve as a realization to them. Zero-mean residuals are assured for both the robust and stable MLE estimators by the standard practice of removing the data mean prior to carrying out a spectral analysis. However, the robust estimator does not produce residuals with a common variance, as seen through q-q plots that are systematically long-tailed without a constant slope (Figs. 1, 2). By contrast, the stable MLE produces residual p-p plots with a constant slope (Figs. 6, 7), meaning that a common scale parameter describes their statistical distributions. Consequently, standard statistical theory argues that the stable MLE is unconditionally unbiased (see also the consistency property of the MLE given by bullet point 1 in Sect. 4), whereas the robust estimate is not because the data used to compute it do not have equal influence

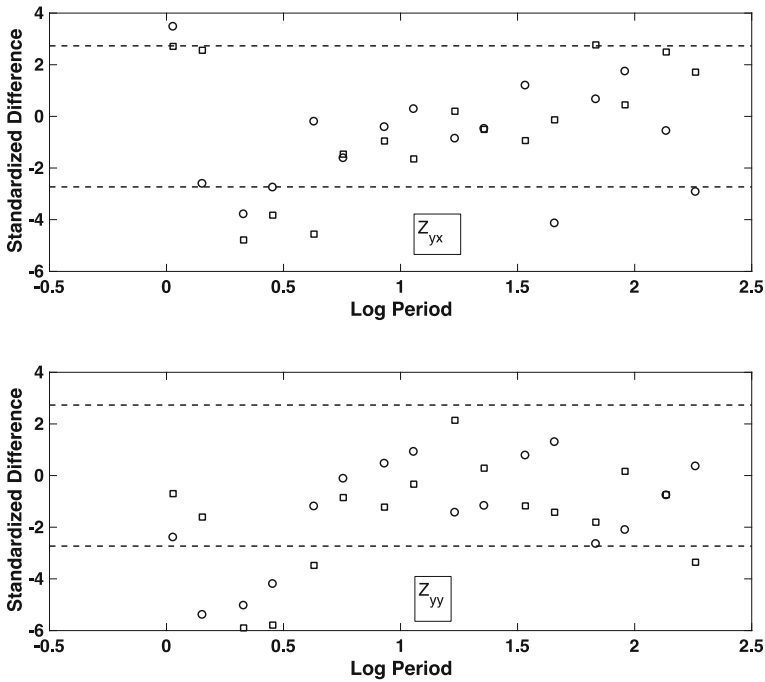


Fig. 15 Difference between the stable MLE and two-stage robust estimates divided by the standard error for the former computed as the square root of the diagonal elements of the improper covariance matrix defined in Sect. 5 for Z_{yx} (top) and Z_{yy} (bottom) plotted against log period. The real parts are shown as circles, and the imaginary parts as squares. The horizontal dashed lines are the Bonferroni 95% confidence bounds after apportioning the 0.05 tail probability among all 8 of the MT tensor elements and define the region within which 95% of the estimates should lie if Gaussian statistics pertain

on the outcome. Instead, data influence the robust estimator in proportion to the size of the random errors, as the underlying robust model is invalid.

A further condition on the linear regression model is no correlation between the residuals (7) and the predicted values (6). Figure 16 contains residual-predicted value plots for the y-orientation at the same periods depicted in Figs. 1 and 2 for the robust estimator. At 45.5 s (top panel), correlation of the real and imaginary parts is weak (-0.05 and 0.04, respectively). A test of the null hypothesis that the correlation is zero against the alternate that it is not is given in Chave (2017, Section 6.3.6) and returns p values of 0.09 and 0.18 for the real and imaginary parts, meaning that there is no correlation. The lower panel of Fig. 16 shows a residual-predicted value plot at a period of 2.1 s. The correlation of the real and imaginary parts is, respectively, -0.22 and -0.23, and the p values for the test of zero correlation are 5×10^{-163} and 8×10^{-188} , strongly rejecting the null hypothesis. In fact, of the 32 real and imaginary parts at all periods, 20 reject the null hypothesis, and the rejection is strong at all periods shorter than 12 s. A similar observation pertains to the x-orientation for the exemplar site.

For comparison purposes, Fig. 17 shows residual-predicted value plots for the stable MLE at the same periods as in Fig. 16. At 45.5 s (top panel), the correlation of the real and imaginary parts is, respectively, -0.03 and 0.07. p values for the test of zero correlation are 0.29 and 0.043, accepting the null hypothesis for the real part and weakly

Fig. 16 Residuals \hat{r} given by (7) plotted against the predicted electric field \hat{e} given by (6) for the y-orientation robust estimator applied at 45.5 s (*top panel*) and 2.1 s (*bottom panel*). The real parts are depicted by pluses, and the imaginary parts by circles

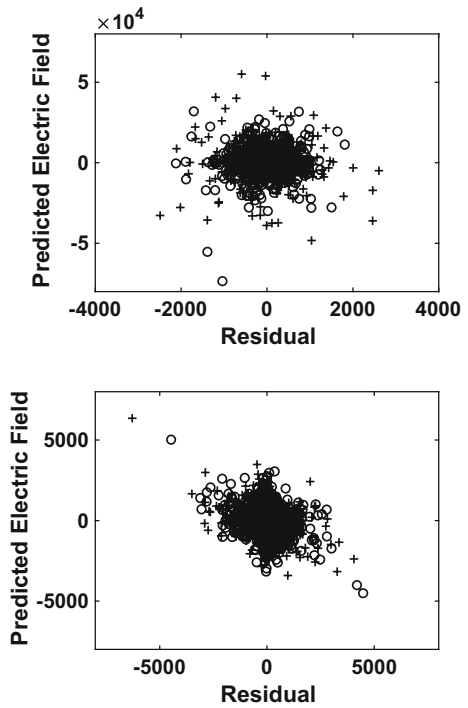
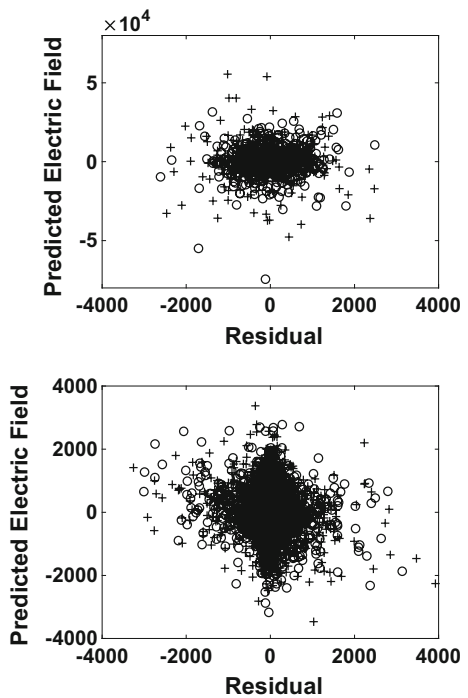


Fig. 17 Residuals \hat{r} given by (7) plotted against the predicted electric field \hat{e} given by (6) for the y-orientation stable MLE estimator applied at 45.5 s (*top panel*) and 2.1 s (*bottom panel*). The real parts are depicted by pluses, and the imaginary parts by circles



rejecting it for the imaginary part. At 2.1 s (bottom panel), the correlation is -0.09 and -0.11 for the real and imaginary parts, and the null hypothesis of zero correlation is strongly rejected, with p values of 4×10^{-48} and 5×10^{-43} , respectively. However, the performance of the stable MLE based on the residual-predicted value criterion is somewhat better than the robust estimator when viewed at all periods; the null hypothesis of zero correlation is rejected at 15 of 32 real and imaginary parts, with these being concentrated at periods under 6 s, or the shortest 5 periods. Similar observations apply to the x-orientation.

The stable MLE is based on a residual model that is independent and identically distributed (iid) that has been proven to be statistically consistent with the data (see Table 1), where the robust model is also iid, but is not consistent with the data statistics. Further, the stable iid model holds even in the presence of the geomagnetic characteristics a–b (where c is the justification for the use of the stable MLE) listed at the end of Sect. 2, but none of these pertain to the robust model. Consequently, the stable MLE represents a significant improvement on the robust model that is statistically consistent with MT data.

It has been argued that since a truncated set of stable rvs has finite variance, robust estimators will obey the classic central limit theorem and hence display Gaussian behavior. Like most central limit theorem arguments, this assertion is naïve and insufficient because it provides no insight into the rate of convergence to a Gaussian limit. The Berry–Esséen theorem (Feller 1971, p. 542) addresses this point:

$$\sup_x |\hat{F}_N(x) - \Phi(x)| \leq \frac{C \hat{\mu}_3}{\hat{\sigma}^3 \sqrt{N}} \tag{31}$$

where the supremum is the least upper bound, \hat{F}_N and Φ are the empirical and Gaussian cdfs, C is a constant of $O(1)$, $\hat{\mu}_3$ is an estimate of the third central moment, $\hat{\sigma}$ is the sample standard deviation, and N is the number of data. Figure 18 shows contours of the least upper bound on the probability difference (31) computed using random draws from

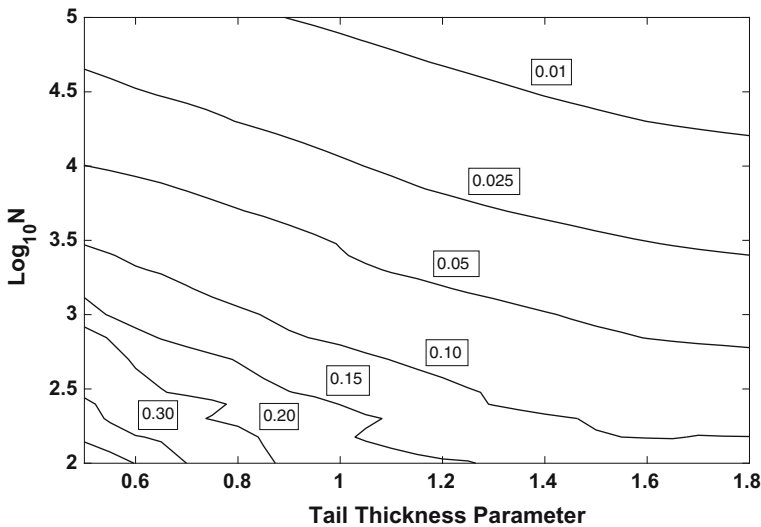


Fig. 18 Contours of the least upper bound on the difference in probability between truncated stable and Gaussian distributions given by Eq. (31) as a function of sample size and tail thickness parameter. See text for discussion

symmetric standard stable distributions having varying tail thickness parameters given by the abscissa, with the most extreme 5% truncated, against the logarithm of the number of data. A truncation level of 5% is an average value for the exemplar robust estimates, with the x - and y -orientations being somewhat more severely and weakly down-weighted, respectively. It is readily apparent that, as the tail thickness parameter decreases, the number of data required to make a truncated stable distribution close to Gaussian must increase, with a range of about an order of magnitude over the specified variation in α . Figure 18 suggests that, all other things being equal, non-Gaussian behavior for a robust estimator is more likely when the tail thickness parameter is small unless the number of data is quite large (and substantially larger than available in the exemplar data set) for any reasonable choice of the probability difference. This is the general tendency seen by comparing Fig. 5 with Figs. 10, 11, 12 and 13 and is also observed for the examples in Chave (2014).

Consequently, the stable MLE estimator is nearly unbiased and has approximate Gaussian second-order statistics. As a result, it is most probable that the differences between the robust estimator and the stable MLE shown in Figs. 12, 13, 14 and 15 reflect issues with the former and hence imply bias in the robust estimates. Further, the systematic dependence of the differences on decreasing tail thickness reinforces this assertion since robust estimators are insensitive to the actual residual distribution. Instead, they will incorporate a larger fraction of long-tailed data into the estimator as tail thickness decreases since the weight function passband (Fig. 3) extends further into the residual distribution tails. Robust estimators merely remove data corresponding to large residuals, leaving behind a population that remains stable but reflecting the truncation. The outcome is MT responses which are sometimes biased and exhibit increasing non-Gaussian behavior as the tail thickness parameter decreases, and whose variances computed using standard methods are systematically larger than the true value, which masks the bias to some degree.

The stable MLE described in this paper is similar to that in Chave (2014), but two significant extensions have been added: the use of a permutation test to establish goodness of fit of a stable model to the residuals, and establishing the numerical accuracy of the solution. The permutation test results are shown in Figs. 8 and 9, and Table 1, and yield high confidence in the stable MLE model because the p values fail to reject the null hypothesis that the residuals are stable at nearly all periods for both the x - and y -orientations. This is a statistically stronger statement than was given in Chave (2014). Re-analysis of the data sets in that paper using the permutation method yields comparable results to those given here.

The unconstrained multivariable optimization algorithm used in the stable MLE is inherently nonlinear, but provides a variety of optimality and step size tolerance parameters. Decreasing these by six orders of magnitude from their defaults does not change the response function estimates beyond the 0.01% level, although considerably more computer time is utilized. This yields a high degree of confidence in the outcome, as numerical inaccuracy is not biasing the stable MLE result.

Pervasive impropriety in the second-order statistics of the MT response function indicates the presence of some degree of non-stationarity in the physical processes underlying them. It is well known that confounding influences such as outliers or weak non-stationarity have substantially more influence on the second-order statistics (i.e., the covariance) than on the first-order statistics (i.e., the MT response itself). Consequently, the approach used here of obtaining the stable MLE to compute the MT response combined with the improper Cramér–Rao bound to obtain its covariance is consistent unless it can be demonstrated that non-stationarity is large. This may occur for extreme geomagnetic

events (e.g., very high energy aurora at high latitudes). In that instance, the Loève (1978) spectral representation that requires correlated spectral increments replaces the ordinary Fourier representation with independent spectral increments. To the author's knowledge, this has not been investigated, although it may prove to be a fruitful avenue for future work.

Thomson et al. (2007) provide convincing evidence that terrestrial geomagnetic field variations display the signature of solar normal modes at periods from a minute to over a day, with harmonics and modulation products that extend to shorter periods. These phenomena begin as normal modes in the Sun, but are affected by a variety of nonlinear processes when traversing interplanetary space and interacting with Earth's geomagnetic field, and in addition exhibit variability due to the nature of the solar surface (Buttighoffer et al. 1999). The outcome is that detection of a given mode on Earth will occur about 1/6 of the time, and perhaps less in the presence of local noise. The Q of the solar modes implies persistence for of order a month, and hence their non-stationarity is weak. In addition, they are typically modulated at multiples of 1 cpd due to the rotation of Earth, and at 0.5 cpd by a nonlinear oscillation of the atmosphere and ionosphere. Consequently, their presentation in terrestrial data is as ephemeral high- Q features that fade in and out over time. This results in persistent non-stationarity in spectra obtained from geomagnetic data; see Thomson and Vernon (2016) for further results. While a link between solar mode phenomena and the stable nature of MT data has not been definitively established, it would be a remarkable coincidence if they were independent given that they occur over the same frequency range. Consequently, it is likely that the persistent impropriety of MT responses reflects the weak non-stationarity induced by the presence of solar normal modes.

In addition, the presence of a quasi-deterministic modal component at a given frequency introduces non-centrality into the MT statistical model. Failure to accommodate non-centrality may result in additional bias to $\hat{\mathbf{z}}$ and parametric estimates of its uncertainty. This issue requires further investigation.

8 Conclusions

Several variants on robust estimators were introduced to MT data analysis beginning in the late 1980s and have since become an essential tool for the analysis of MT time series. The robust model of a Gaussian core contaminated by longer-tailed outlying data underlies all robust estimators and is critically evaluated.

The statistical model for linear regression that underlies MT response function estimation is defined, along with the conditions that various statistical entities, including the regression residuals, must meet to yield an unbiased or Gaussian impedance. Several statistical tools that are useful in evaluating the distribution of the regression residuals and their consistency with a given statistical model are introduced. These include quantile–quantile and percent–percent plots that provide a qualitative picture of the residual distribution, and a Kolmogorov–Smirnov test for goodness of fit that is corrected for bias due to estimation of distribution parameters from the data under test using an exact permutation method.

Using an exemplar data set, the regression residuals from both an ordinary least squares and robust estimator are shown to be systematically long-tailed rather than comprising a Gaussian core contaminated by outliers. This observation is not new and appears to be pervasive. As a result, the robust model that underlies robust estimation is invalid. Instead,

the residuals are pervasively described by the stable distribution family for which the Gaussian is an end member, but for which the remaining distributions possess algebraic rather than exponential tails. Stable distributions cannot usually be expressed in closed form and are parameterized by tail thickness, skewness, scale and location. Bias-corrected Kolmogorov–Smirnov tests demonstrate that residuals from the exemplar data are well described by a stable model, with nearly zero skewness and a tail thickness that decreases with period.

Because the residual distribution for MT data can now be characterized statistically, an optimal maximum likelihood estimator (MLE) for MT response function estimation can be devised. A two-stage approach in which the stable distribution parameters are first estimated from an initial set of residuals, and then the MT response and its residuals are computed with the stable parameters fixed, is defined utilizing a published MLE algorithm and a nonlinear multivariable function minimization algorithm for the first and second stages, and validated using the exemplar data.

The standard approach to computing the covariance for an MLE is through the Cramér–Rao lower bound. For complex data such as the MT response, the second-order statistics are described by covariance and pseudo-covariance matrices. When the latter is zero, the complex data are proper and are otherwise improper. MT responses are shown to be pervasively improper as a consequence of weak non-stationarity of the underlying physical processes. The Fisher information matrix for improper complex data is derived, from which the Cramér–Rao inequality obtains, and whose lower bound describes the covariance of the MT response.

Comparison of the stable MLE and robust estimates for the exemplar data set shows differences that increase as the tail thickness parameter decreases, and that are both too frequent and too large to be consistent with Gaussian behavior of all of their statistical entities. This can be partially ameliorated using a two-stage robust estimator, although this has not worked on other data sets. The condition for unconditional unbiasedness of the linear regression model is zero-mean random errors (for which the residuals are a realization) with a common variance. This condition holds for the stable MLE, but not for the robust estimator, implying that the latter is biased and explaining the disparity between the two types of estimates. The stable MLE is also shown to yield lower correlation of the residuals with the predicted electric field as compared to the robust result. Finally, the rate of convergence of truncated stable data to the Gaussian limit via the central limit theorem is bounded using the Berry–Esséen theorem, and it is shown that the number of data needed to reach that limit increases as the tail thickness parameter decreases. As a result, non-Gaussian behavior of a robust estimator is more likely when the tail thickness parameter is small unless the number of data is very large, which in part explains the observed differences between the robust and stable MLE estimates.

Both the stable MLE and the robust estimator utilize a residual model that is iid, but only the stable MLE can be shown to be consistent with such a model. Further, the iid stable model holds even when the variance of the residuals depends on that of the data and when data anomalies occur in patches, both of which are characteristic of geomagnetic data. As a result, the stable MLE is a significant improvement on the robust model that is itself not consistent with MT data.

The pervasive impropriety of the second-order statistics of the MT response function implies a degree of non-stationarity in the underlying physical processes. When the non-stationarity is weak, the use of the stable MLE together with the improper Cramér–Rao bound to characterize the first and second-order statistics of MT data is appropriate. Weak

non-stationarity is consistent with the observation of high-Q solar normal modes in geomagnetic data in the MT band, with a persistence time of order a month.

Acknowledgements This work was partially supported by the Walter A. and Hope Noyes Smith Chair for Excellence in Oceanography at the Woods Hole Oceanographic Institution.

References

- Adcock RJ (1878) A problem in least squares. *Analyst* 5:53–54
- Buttighoffer A, Lanzerotti LJ, Thomson DJ, MacLennan CG, Forsyth RJ (1999) Spectral analysis of the magnetic field inside particle propagation channels detected by ULYSSES. *Astron Astrophys* 351:385–392
- Chave AD (2012) Chapter 5: Estimation of the magnetotelluric response function. In: Chave AD, Jones AG (eds) *The magnetotelluric method: theory and practice*. Cambridge University Press, Cambridge, pp 165–218
- Chave AD (2014) Magnetotelluric data, stable distributions and impropriety: an existential combination. *Geophys J Int* 198:622–636
- Chave AD (2017) *Computational statistics for the earth sciences*. Cambridge University Press, Cambridge
- Chave AD, Thomson DJ (1989) Some comments on magnetotelluric response function estimation. *J Geophys Res* 94:14215–14225
- Chave AD, Thomson DJ (2003) A bounded influence regression estimator based on the statistics of the hat matrix. *J R Stat Soc Ser. C* 52:307–322
- Chave AD, Thomson DJ (2004) Bounded influence estimation of magnetotelluric response functions. *Geophys J Int* 157:988–1006
- Chave AD, Thomson DJ, Ander ME (1987) On the robust estimation of power spectra, coherences, and transfer functions. *J Geophys Res* 92:633–648
- Conn NR, Gould NIM, Toint PL (2000) *Trust region methods*. SIAM & MPS, Philadelphia, PA, p 959
- Egbert GD (1997) Robust multiple-station magnetotelluric data processing. *Geophys J Int* 130:475–496
- Egbert GD (2002) Processing and interpretation of electromagnetic induction array data. *Surv Geophys* 23:207–249
- Egbert GD, Booker JR (1986) Robust estimation of geomagnetic transfer functions. *Geophys J R Astr Soc* 87:173–194
- Feller W (1971) *An introduction to probability theory and its applications*, vol 2. Wiley, New York
- Gamble TD, Goubau WM, Clarke J (1979) Magnetotellurics with a remote reference. *Geophysics* 44:53–68
- Goubau WM, Gamble TD, Clarke J (1978) Magnetotelluric data analysis: removal of bias. *Geophysics* 43:1157–1162
- Loève M (1978) *Probability theory II*, 4th edn. Springer, Berlin
- McCulloch JH (1998) Linear regression with stable disturbances. In: Adler RJ, Feldman RE, Taqqu MS (eds) *A practical guide to heavy tails: statistical techniques and applications*. Birkhäuser, Basel, pp 359–378
- Michael JR (1983) The stabilized probability plot. *Biometrika* 70:11–17
- Nolan JP (2001) Maximum likelihood estimation and diagnostics for stable distributions. In: Barndorff-Nielsen OE, Mikosch T, Resnick SI (eds) *Lévy processes: theory and applications*. Birkhäuser, Basel, pp 379–400
- Nolan JP, Ojeda-Revah D (2013) Linear and nonlinear regression with stable errors. *J Econom* 172:186–194
- Percival D, Walden A (1993) *Spectral analysis for physical applications*. Cambridge University Press, Cambridge
- Picinbono B (1996) Second order complex random vectors and normal distributions. *IEEE Trans Signal Process* 44:2637–2640
- Schreier PJ, Scharf LL (2010) *Statistical signal processing of complex-valued data*. Cambridge University Press, Cambridge, p 309
- Shaffer JP (1991) The Gauss–Markov theorem and random regressors. *Am Stat* 45:269–273
- Sims WE, Bostick FX, Smith HW (1971) The estimation of magnetotelluric impedance tensor elements from measured data. *Geophysics* 36:938–942
- Slepian D (1978) Prolate spheroidal wave functions, Fourier analysis, and uncertainty-V: the discrete case. *Bell Syst Tech J* 57:1371–1430

- Stuart A, Ord JK, Arnold S (1999) Kendall's advanced theory of statistics, vol 2A: classical inference and the linear model. Edward Arnold, New York
- Thomson DJ, Vernon FL (2016) Some comments on the analysis of 'big' scientific time series. *Proc IEEE* 104:2220–2248
- Thomson DJ, Lanzerotti LJ, Vernon FL, Lessard MR, Smith LTP (2007) Solar modal structure of the engineering environment. *Proc IEEE* 95:1085–1132
- Van den Bos A (1995) A multivariate complex normal distribution—a generalization. *IEEE Trans Inf Theory* 41:537–539
- Vozoff K (1972) The magnetotelluric method in the exploration of sedimentary basins. *Geophysics* 37:98–141
- Weidelt P, Chave AD (2012) Chapter 4: The magnetotelluric response function. In: Chave AD, Jones AG (eds) *The magnetotelluric method: theory and practice*. Cambridge University Press, Cambridge, pp 122–164