Geophysical Journal International

Geophys. J. Int. (2014) **198,** 622–636
Advance Access publication 2014 May 30
GJI Geomagnetism, rock magnetism and palaeomagnetism

doi: 10.1093/gji/ggu121

Magnetotelluric data, stable distributions and impropriety: an existential combination

Alan D. Chave

Department of Applied Ocean Physics and Engineering, Deep Submergence Laboratory, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. E-mail: achave@whoi.edu

Accepted 2014 March 28. Received 2014 March 26; in original form 2013 November 21

SUMMARY

The robust statistical model of a Gaussian core contaminated by outlying data that underlies robust estimation of the magnetotelluric (MT) response function has been re-examined. The residuals from robust estimators are systematically long tailed compared to a distribution based on the Gaussian, and hence are inconsistent with the robust model. Instead, MT data are pervasively described by the alpha stable distribution family whose variance and sometimes mean are undefined. A maximum likelihood estimator (MLE) that exploits the stable nature of MT data is formulated, and its two-stage implementation in which stable parameters are first fit to the data and then the MT responses are solved for is described. The MLE is shown to be inherently robust, but differs from the conventional robust estimator because it is based on a model derived from the data, while robust estimators are *ad hoc*, being based on the robust model that is inconsistent with actual data. Propriety versus impropriety of the complex MT response was investigated, and a likelihood ratio test for propriety and its null distribution was established. The Cramér-Rao lower bounds for the covariance matrix of proper and improper MT responses were specified.

The MLE was applied to exemplar long period and broad-band data sets from South Africa. Both are shown to be significantly stably distributed using the Kolmogorov–Smirnov goodness of fit and Ansari-Bradley non-parametric dispersion tests. Impropriety of the MT responses at both sites is pervasive, hence the improper Cramér-Rao bound was used to estimate the MLE covariance. The MLE is shown to be nearly unbiased and well described by a Gaussian distribution based on bootstrap simulation. The MLE was compared to a conventional robust estimator, establishing that the standard errors of the former are systematically smaller than for the latter and that the standardized differences between them exhibit excursions that are both too frequent and too large to be described by a Gaussian model. This is ascribed to pervasive bias of the robust estimator that is to some degree obscured by their systematically large confidence bounds. Finally, a series of topics for further investigation is proposed.

Key words: Time series analysis; Numerical approximations and analysis; Fractals and multifractals; Probability distributions; Magnetotellurics.

1 INTRODUCTION

The fundamental datum in magnetotellurics (MT) is a location-specific, frequency-dependent tensor $\overleftrightarrow{\mathbf{Z}}$ linearly connecting the horizontal electric and magnetic fields measured at Earth's surface (or at the seafloor):

$$\mathbf{E} = \overleftrightarrow{\mathbf{Z}} \bullet \mathbf{B}, \tag{1}$$

where $\bullet$ denotes the inner product. When the electric $\mathbf{E}$ and magnetic $\mathbf{B}$ fields are measurements, (1) does not hold exactly due to the finite size of the sample and the presence of noise, and it becomes necessary to estimate the MT response tensor in a statistical manner. The standard approach is application of least squares principles to the row-by-row solution of (1) at a given frequency in the form:

$$\mathbf{e} = \overleftrightarrow{\mathbf{b}} \bullet \mathbf{z} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{e}$ is the electric field response N-vector, $\overleftrightarrow{\mathbf{b}}$ is the $N \times 2$ magnetic field predictor matrix, $\mathbf{z}$ is the MT response function 2-vector and $\boldsymbol{\varepsilon}$ is an N-vector of unobservable random errors, with all of these entities being complex. While (2) was solved using ordinary least squares in the early days (e.g. Sims *et al.* 1971), beginning in the 1980s robust (e.g. Egbert & Booker 1986; Chave *et al.* 1987; Chave

& Thomson 1989) and subsequently bounded influence (Chave & Thomson 2004) estimators were developed to eliminate the pervasive influence of outliers and putative non-stationarity. Chave (2012) provides a recent review of the topic. Such estimators are now in standard use, and are widely recognized as an essential tool for the MT practitioner. In the sequel, these will collectively be called robust estimators.

The statistical model (hereafter the robust model) underlying robust estimation is the existence of a Gaussian core of data contaminated by a fraction of outlying ones that are statistically distinct and a source of serious bias, with the influence of the outliers being removed through processing. However, when the statistical distribution of the weighted estimated random errors (hereafter residuals) $\mathbf{w} \bullet (\mathbf{e} - \overleftrightarrow{\mathbf{b}} \bullet \hat{\mathbf{z}})$ is examined, where $\mathbf{w}$ is a diagonal matrix of robust weights and $\hat{\mathbf{z}}$ is the estimated MT response function, it is readily apparent that the model assumptions are not valid because the residuals are systematically long tailed. Fig. 1 shows the ordered magnitudes of the residuals for two periods at two sites in South Africa against truncated Rayleigh quantiles. The quantiles of a statistical distribution divide the area under the probability density function (pdf) into equal size pieces (hence equal probability increments), so that if the residuals were drawn from that distribution, they would plot against the quantiles as a straight line. Site 172 (22°37′49″S, 29°30′40″E), using Site 145 (26°19′13″S, 26°05′18″E) as a remote reference, is a long period (5 s sample interval) MT data set that was used as an exemplar by Chave (2012). Site 013 (30°44′05″S, 21°53′28″E), using Site 014 (30°36′58″S, 22°00′21″E) as a remote reference, provides 24 Hz data that span the MT dead band at around 1 s period. Both data sets were collected in geomagnetic coordinates, with the *x*-orientation electric field pointing north and the *y*-orientation electric field pointing east. As described in Chave *et al.* (1987), the Rayleigh distribution is appropriate for the magnitude of complex Gaussian data, and the importance of utilizing the truncated form of a pdf that accounts for data that are removed by robust weighting is outlined in Chave & Thomson (2004). The bottom panel in Fig. 1 shows Site 172 residuals at 1280 s, where robust weighting has removed about 11 per cent of the data. The top panel shows Site 013 residuals at 1.3 s, where robust weighting has eliminated about 22 per cent of the data. Both quantile-quantile (q-q) plots are systematically long tailed, and are not in accord with the robust model that would yield a straight line. As a further demonstration (not shown) of the systematic nature of the residual distribution, if the severity of robust weighting is increased, the q-q plots continue to have the same shape, although with fewer data. The residual q-q plots in Fig. 1 are typical of MT data at all periods between 1s and 1 d, as demonstrated by examining hundreds of sites globally.

The observations surrounding Fig. 1 are not new, but have received little attention over the years, perhaps because robust estimators do appear to dramatically improve MT response estimates. It is well known that the least squares estimator is the maximum likelihood estimator (MLE) when the regression residuals are Gaussian, and the robust estimator should approximate the MLE if the robust model is correct. The observation that residuals from robust estimators are systematically long-tailed, and do not have a Gaussian core, means that they lack the optimality properties of the MLE, particularly asymptotic consistency, normality and efficiency. This raises issues regarding the bias properties of robust estimators and especially their second-order statistics. However, resolving them requires information on the actual distribution of MT residuals that has heretofore not been available.

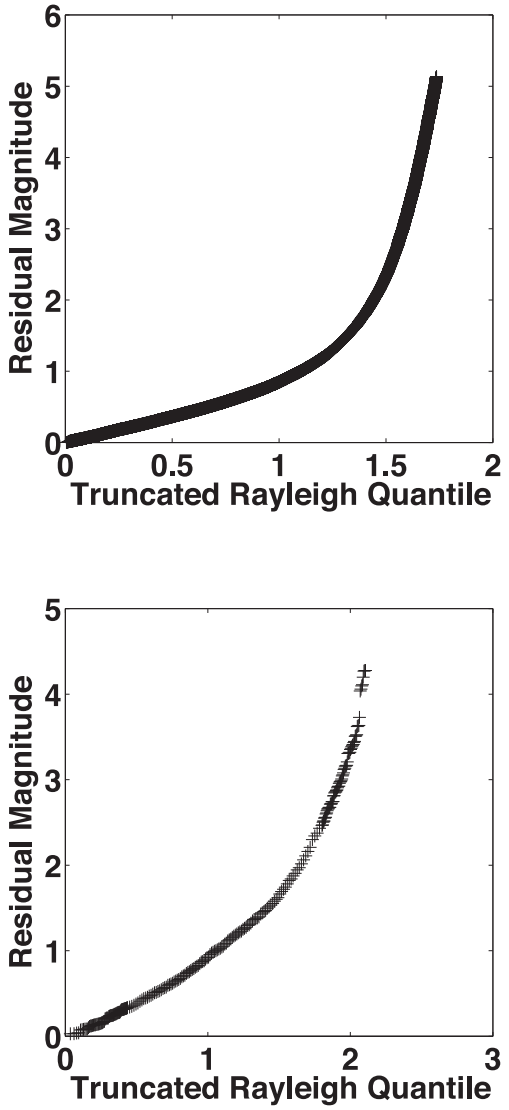

Figure 1. Quantile-quantile plots of the magnitude of the complex residuals from the robust remote reference estimator described by Chave & Thomson (2004) against the truncated Rayleigh distribution quantiles. The bottom panel shows the result at 1280 s for the *y*-orientation of the response tensor at Site 172, while the top panel shows the result at 1.3 s for the *x*-orientation at Site 013, in both cases using geomagnetic coordinates. The ordinate data have been normalized to the variance of the standardized Rayleigh distribution. Both exemplars are systematically long tailed rather than representative of a Gaussian core with outlier contamination.

As will be shown, MT residuals are persistently drawn from a family of distributions with algebraic rather than exponential tails, but a complex version of this family has not been investigated. Consequently, an augmented real version of (2) will be used for the remainder of this paper. Let $\mathbf{e}' = \left(\mathbf{e}^r \ \mathbf{e}^i\right)^T$ be a column $2N$-vector of the real parts of the electric field stacked above the imaginary parts and let

$$\overleftrightarrow{\mathbf{b}}' = \begin{bmatrix} \overleftrightarrow{\mathbf{b}}^r & -\overleftrightarrow{\mathbf{b}}^i \\ \overleftrightarrow{\mathbf{b}}^i & \overleftrightarrow{\mathbf{b}}^r \end{bmatrix} \tag{3}$$

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

$$\mathbf{e}' = \overleftrightarrow{\mathbf{b}}' \bullet \mathbf{z}' + \boldsymbol{\varepsilon}' \tag{4}$$

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

Quantile-quantile plots such as Fig. 1 emphasize the distribution tails, hence do not capture the nature of systematically long-tailed distributions very well. However, under the probability integral transformation, applying a hypothesized cumulative distribution function (cdf) to a set of data results in uniformly distributed variates if the data are drawn from that cdf. This motivates percent-percent (p-p) plots of the uniform quantiles against a hypothesized cdf applied to the ranked MT residuals. By contrast to q-q plots, p-p plots emphasize the distribution mode rather than the tail extremes, and approximate a straight line if the hypothesized distribution is correct. A modification to the p-p plot that equalizes the variance of the data using a sine transformation of both the ordinate and abscissa is given by Michael (1983), and is shown in Fig. 2 for the ordinary least squares residuals along with 95 per cent confidence bands derived from the Kolmogorov–Smirnov statistic (Kvam & Vidakovic 2007, section 6.1) at the same sites and periods as in Fig. 1 using a Gaussian distribution as the target. Fig. 2 combines the real and imaginary parts of the residuals, but equivalent results are obtained if they are treated separately. In both cases, the Kolmogorov–Smirnov test for normality strongly rejects, with p-values of $\sim 10^{-4}$ and floating point zero for the bottom and top panels, respectively. It is very clear from these plots that the least squares residuals are not close to being Gaussian near the centre. This is especially apparent for the 1.3 s data, where the discrepancy is enormous and systematic.

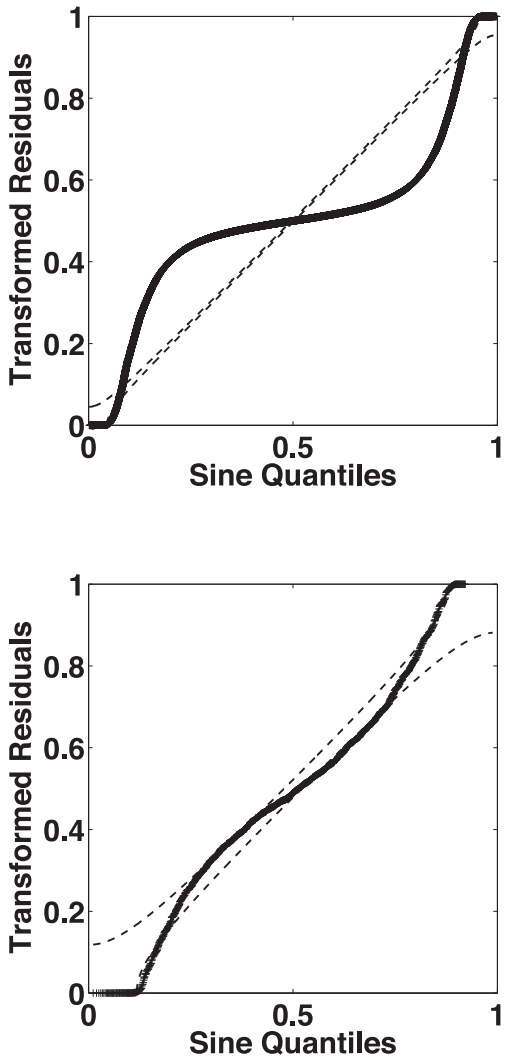

Figure 2. Variance equalized percent-percent plots of the ordinary least squares residuals for the data in Fig. 1, with the ordinate being the sine transformed Gaussian cumulative distribution applied to the standardized real and imaginary residuals. The bottom panel shows the result at 1280 s for the x-orientation of the response tensor at Site 172, while the top panel shows the result at 1.3 s for the y-orientation at Site 013, in both cases using geomagnetic coordinates. There are, respectively, 800 and 36 710 data at the two periods.

The p-value from a Kolmogorov–Smirnov test will be biased if the distribution parameters are estimated from the data under evaluation, as is the case in the present work. However, a Monte Carlo simulation (Davison & Hinkley 1997, section 4.2.1) can be used to obtain an exact p-value as follows. First, obtain the sample Kolmogorov–Smirnov test statistic for the data against the target distribution with the distribution parameters estimated from them using the maximum likelihood method. Secondly, obtain random draws with replacement from the target distribution with the same parameters containing the same number of values as the original data. Third, compute the Kolmogorov–Smirnov test statistic for each of the random draws against the target distribution. Repeat steps 2 and 3 a large (999 in the present case) number of times. The double sided Monte Carlo p-value is given by twice the smaller of the number of times the random draw test statistic is larger or smaller than the sample test statistic divided by the number of random draws. The Monte Carlo approach is computationally challenging, but was carried out for target distribution simulations and for a subset of the data analysed in this paper. The results show that p-value bias is not sufficient to change any of the conclusions about the distribution of MT data in the sequel.

Fig. 3 repeats the results from Fig. 2 but with a target distribution that is alpha stable with MLE distribution parameters estimated using the algorithm of Nolan (2001). Alpha stable distributions are described in Section 2, and consist of a continuous family, one of whose end members is Gaussian, but have infinite variance (and sometimes infinite mean) for all other constituents. The agreement with the alpha stable model shown in Fig. 3 is striking. Further, a Kolmogorov–Smirnov test of the null hypothesis that the residuals are stable provides strong support for the Site 172 data [p-value of 0.88 with 1200 degrees of freedom (dof)], but weak rejection for the Site 013 example (p-value of 0.03 with 57 000 dof). However, reducing the tail probability to 0.025 results in test acceptance. Further, a non-parametric Ansari-Bradley test (Ansari & Bradley 1960) based on ranks for the null hypothesis that the distribution of the residuals and a random sample from a stable distribution with the fitted MLE parameters are the same yields a p-value of 0.91, providing support for the stable model. It is obvious from (2) and (4) that stable residuals imply that the electric and magnetic fields are themselves stably distributed. These exemplar observations have been reproduced for a large number of data sets over a wide period range, and while it is impossible to argue that stably distributed MT data are ubiquitous from a finite sample, they are certainly pervasive.

The remainder of this paper develops and demonstrates a proof-of-concept data analysis technique that exploits the stable nature of MT data. Section 2 summarizes the characteristics of stable distributions. Section 3 introduces a MLE for stable MT data and describes its implementation. Section 4 discusses impropriety of the

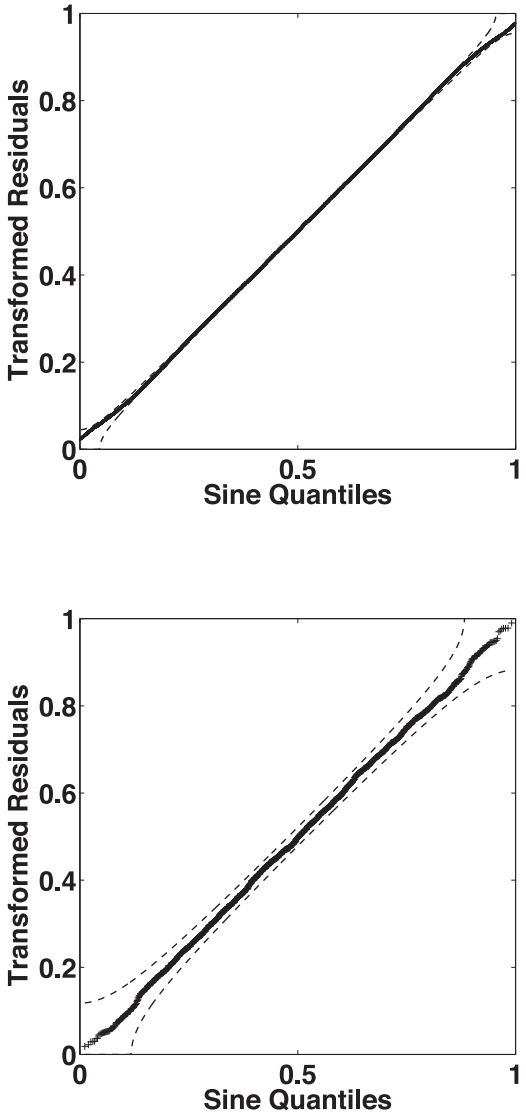

Figure 3. Variance equalized percent-percent plots of the ordinary least squares residuals of Fig. 2, with the ordinate being the sine transformed alpha stable cumulative distribution applied to the standardized residuals. The bottom panel shows the result at 1280 s for the y-orientation of the response tensor at Site 172, while the top panel shows the result at 1.3 s for the x-orientation at Site 013, in both cases using geographic coordinates. The stable distribution parameters α and β are (1.16, 0.07) and (1.02, 0.00), respectively; see Section 2 for their definition.

complex MT response function, which also appears to be pervasive, introduces a likelihood ratio test for it, and summarizes its effect on the asymptotic second-order statistics of the MLE. Section 5 applies these principles to exemplar data sets. Section 6 discusses implications and next steps. Section 7 contains conclusions.

2 STABLE DISTRIBUTIONS

A random variable $\mathbf{X}$ is stable if a linear combination of two independent realizations of it has the same distribution. Stable distributions were first described by Lévy (1925), and may be defined through their characteristic function (the inverse Fourier transform of the pdf):

$$\phi(s) = \mathcal{E}\left[e^{is\mathbf{X}}\right] = \begin{cases} e^{-\gamma^\alpha |s|^\alpha \left[1 + i\beta \tan\left(\frac{\pi\alpha}{2}\right) \mathrm{sgn}(s) \left(\gamma^{1-\alpha}|s|^{1-\alpha} - 1\right)\right] + i\delta s} & \alpha \neq 1 \\ e^{-\gamma|s|\left[1 + i\beta \frac{2}{\pi} \mathrm{sgn}(s) \log(\gamma|s|)\right] + i\delta s} & \alpha = 1, \end{cases} \quad (5)$$

where $\mathcal{E}$ denotes the expected value. Stable distributions are parameterized by tail thickness $\alpha \in (0,\ 2]$, skewness $\beta \in [-1,\ 1]$, scale $\gamma \in (0,\ \infty)$ and location $\delta \in (-\infty,\ \infty)$; the latter two parameters are analogous to the standard deviation and mean. The pdf $S(x|\alpha,\ \beta,\ \gamma,\ \delta)$ obtained as the Fourier transform of (5) cannot be expressed in closed form except for three special cases: when $\alpha = 2$, the distribution is Gaussian with parameters $\left(\delta,\ 2\gamma^2\right)$ and β has no meaning; when $\alpha = 1,\ \beta = 0$, the distribution is Cauchy; and when $\alpha = 1/2,\ \beta = 1$, the distribution is Lévy. There are several alternate expressions to (5) for the stable characteristic function that frequently cause confusion, but the present version [the 0-parameterization of Nolan (1998)] has the advantage of being continuous in the parameters. It is also a location-scale parameterization, in the sense that if $\mathbf{X} \sim S(x|\alpha,\ \beta,\ \gamma,\ \delta)$, then $(\mathbf{X} - \delta)/\gamma \sim S(x|\alpha,\ \beta,\ 1,\ 0)$, where $\sim$ means 'is distributed as'. Stable distributions are unimodal, have an infinitely differentiable pdf, and their support is doubly infinite except when $\beta = \pm 1$, where they are bounded on (totally skewed to) the left (right) or right (left), respectively. Further information may be found in Feller (1971), Samorodnitsky & Taqqu (1994) and Uchaiken & Zolotarev (1999).

Stable distributions possess finite variance only when $\alpha = 2$. For $1 < \alpha < 2$, stable distributions have finite mean but infinite variance, while for $0 < \alpha \leq 1$, both the mean and variance are undefined. The tails of stable distributions are algebraic except for the Gaussian end member, with the tail thickness decreasing with α. For example, the Cauchy distribution has $1/|x|^2$ tails, while the Lévy distribution has a $1/x^{1.5}$ right tail. More generally, $S(x|\alpha,\ \beta,\ \gamma,\ \delta) \to |x|^{-(\alpha+1)}$ as $x \to \pm\infty$ for $0 < \alpha < 2$ as long as $\beta \neq \pm 1$. As a consequence, stable data exhibit more extreme values and much more variability than Gaussian ones, as has long been observed for MT data.

Aside from the empirical observation that stable distributions fit many types of data in fields ranging from physics to finance, as described by Uchaiken & Zolotarev (1999), they also have significance due to the generalized central limit theorem. The classical central limit theorem states that, for independent data with a finite variance σ^2, $\sqrt{N}(\bar{X} - \mu)/\sigma$ converges in distribution to a standardized Gaussian distribution as the number of data $N \to \infty$, where $\bar{X}$ is the sample mean (hence a sum of independent random variables) and μ is the population mean. The generalized central limit theorem extends the classical central limit theorem to cover random variables with infinite variance, stating that the sum of independent and identically distributed random variables with infinite variance converges in distribution to a stable distribution with $0 < \alpha < 2$.

The paucity of closed form expressions for the density/distribution functions and their derivatives means that numerical methods must be used both to compute those entities and to estimate the stable parameters from data, as was done for Fig. 3. Nolan (1997) describes algorithms to compute univariate stable density functions. McCulloch (1986) used a small set of estimated quantiles to retrieve the stable parameters. Koutrouvelis (1980) and Kogon & Williams (1998) computed the stable parameters from the empirical characteristic function. Maximum likelihood estimation of the stable parameters was introduced by DuMouchel (1975) and refined by Nolan (2001). In the present work, the MLE approach

will be used exclusively, as it has all of the standard optimality properties of the MLE (DuMouchel 1973).

3 MAXIMUM LIKELIHOOD ESTIMATION FOR STABLE MT DATA

By the Gauss-Markov Theorem, ordinary least squares yields the best linear unbiased estimator when the common variance of the residuals is finite, and convergence of the least squares estimator to the true value of the MT response function occurs at the rate $N^{-1/2}$. 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 $\alpha \leq 1$. In other words, the least squares estimator has zero asymptotic efficiency due to overemphasis of data with large influence that rises as α falls. However, an iteratively reweighted 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. It is also instructive towards understanding the distinction between robust estimation and the present robust modelling approach to MT response function computation. McCulloch (1998) also describes an iteratively reweighted stable MLE.

The standard approach towards implementing a MLE requires independence of the data. As described by Chave (2012), a properly computed set of Fourier transforms obtained by dividing a time-series into sections whose length is a few times the period of interest will be independent. This entails pre-whitening the entire time-series, weighting each section with a low bias data taper such as a Slepian sequence, and overlapping the sections in accordance with the correlation properties of the data taper. The resulting section Fourier transforms at a given period become the data $\mathbf{e}$ or $\overleftrightarrow{\mathbf{b}}$ used in (2) or (4).

For stable MT data, the pdf of a single residual is $S\left(\hat{\varepsilon}_i|\alpha,\ \beta,\ \gamma,\ \delta\right)$, where $\hat{\varepsilon}_i = \mathrm{e}_i - \mathbf{b}_i \bullet \hat{\mathbf{z}}$ is the estimated residual and $\mathbf{b}_i$ is the i-th row of $\overleftrightarrow{\mathbf{b}}$ in (4) after dropping the prime notation. For independent data, the sampling distribution is

$$S_N\left(\hat{\boldsymbol{\varepsilon}}|\alpha,\ \beta,\ \gamma,\ \delta\right) = \prod_{i=1}^{2N} S\left(\hat{\varepsilon}_i|\alpha,\ \beta,\ \gamma,\ \delta\right). \tag{6}$$

The likelihood function is the sampling distribution (6) 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:

$$\mathcal{L}\left(\boldsymbol{\xi},\ \mathbf{z}|\hat{\boldsymbol{\varepsilon}}\right) = \sum_{i=1}^{2N} \log S\left(\hat{\varepsilon}_i|\boldsymbol{\xi}\right), \tag{7}$$

where $\boldsymbol{\xi} = (\alpha,\ \beta,\ \gamma,\ \delta)$ is a vector of stable parameters. The first-order conditions for the MLE solution follow by setting the derivatives of (7) with respect to the parameters to zero:

$$\partial_{\xi_j}\mathcal{L}\left(\boldsymbol{\xi},\ \mathbf{z}|\hat{\boldsymbol{\varepsilon}}\right) = \sum_{i=1}^{2N} \frac{\partial_{\xi_j} S\left(\hat{\varepsilon}_i|\boldsymbol{\xi}\right)}{S\left(\hat{\varepsilon}_i|\boldsymbol{\xi}\right)} = 0 \quad j = 1, \ldots 4$$

$$\partial_{z_k}\mathcal{L}\left(\boldsymbol{\xi},\ \mathbf{z}|\hat{\boldsymbol{\varepsilon}}\right) = -\sum_{i=1}^{2N} \frac{\partial_{z_k} S\left(\hat{\varepsilon}_i|\boldsymbol{\xi}\right)}{S\left(\hat{\varepsilon}_i|\boldsymbol{\xi}\right)} b_{ik} = 0 \quad k = 1, \ldots 4. \tag{8}$$

The sufficient condition for the solution of (8) to be a maximum is that the Hessian matrix of the log likelihood function be negative definite. Eq. (8) will be solved using a two-stage process that decouples its two types of parameters. In the first stage, the stable distribution parameter vector $\boldsymbol{\xi}$ will be estimated using the stable MLE algorithm of Nolan (2001). In the second stage, the MT response function will be computed using these values for the stable distribution parameters. Let $\lambda\left(\hat{\varepsilon}_i\right) = \log S\left(\hat{\varepsilon}_i|\boldsymbol{\xi}\right)$ where $\boldsymbol{\xi}$ is assumed known. The second equation in (8) can be rewritten:

$$-\sum_{i=1}^{2N} \lambda'\left(\hat{\varepsilon}_i\right) b_{ik} = -\sum_{i=1}^{2N} \frac{\lambda'\left(\hat{\varepsilon}_i\right)}{\hat{\varepsilon}_i}\left(e_i b_{ik} - \sum_{j=1}^{4} b_{ij}\,\hat{z}_j\,b_{ik}\right) = 0$$
$$k = 1, \ldots 4, \tag{9}$$

where $\lambda'(x)$ is the score function. Eq. (9) may be recast in matrix form to yield the iteratively reweighted MLE solution:

$$\hat{\mathbf{z}}_{ML} = \left(\overleftrightarrow{\mathbf{b}}^T \bullet \mathbf{w} \bullet \overleftrightarrow{\mathbf{b}}\right)^{-1} \bullet \left(\overleftrightarrow{\mathbf{b}}^T \bullet \mathbf{w} \bullet \mathbf{e}\right), \tag{10}$$

where $\mathbf{w}$ is a diagonal weight matrix whose i-th element is $-\lambda'\left(\hat{\varepsilon}_i\right)/\hat{\varepsilon}_i$. A remote magnetic field reference is easily implemented by replacing $\overleftrightarrow{\mathbf{b}}^T$ with $\overleftrightarrow{\mathbf{b}}_r^T$ in (10). Except for some implementation details, (10) is very similar to a robust estimator. However, there is one very important distinction: the weights for the stable MLE are based on a model for the residuals that is derived directly from them, while robust estimator weights are *ad hoc*, being based on the robust model that is inconsistent with actual MT data.

As proof that (10) reduces the influence of outlying data, Fig. 4 shows the weight function $\mathbf{w}$ for a standardized ($\gamma = 1,\ \delta = 0$) symmetric ($\beta = 0$) stable distribution for several values of the tail thickness parameter α and compared to a typical robust estimator weight function. From (10), the solution is unaffected by scaling the weight function by a constant, so the relative size of the curves in Fig. 4 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 robust estimator of Chave & Thomson (2004) utilizes a weight function that is constant between $\pm\boldsymbol{\kappa}$, where $|\boldsymbol{\kappa}|$ depends on the number of data but typically lies between 3 and 5, and then descends abruptly to zero for larger values; it is shown for $\boldsymbol{\kappa} = 4$ in Fig. 4. It is readily apparent that for data with a tail thickness parameter lying below 2 that are typical for MT, the robust estimator consistently overemphasizes data corresponding to large residuals within the passband relative to those at the centre when compared to the stable MLE estimator. This has

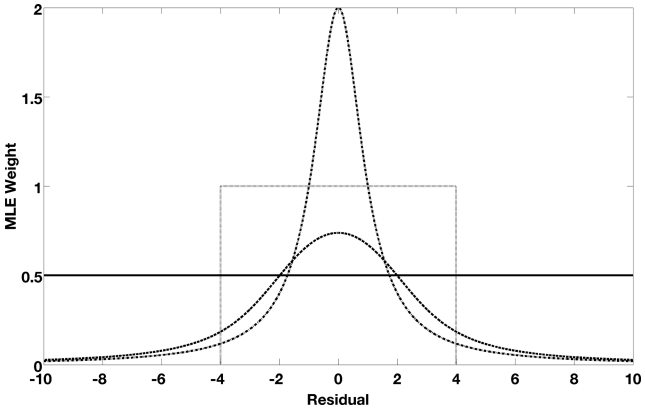

Figure 4. The MLE weight function for a standardized ($\gamma = 1,\ \delta = 0$) symmetric ($\beta = 0$) stable distribution for tail thickness parameter α values of 2 (solid line), 1.5 (dashed line) and 1.0 (dot-dash line). The dotted boxcar function is a typical robust weight function with a cutoff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

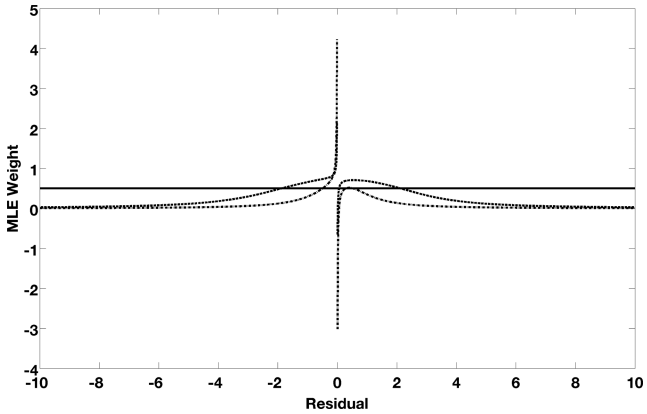

Figure 5. The MLE weight function for a standardized ($\gamma = 1$, $\delta = 0$) skewed ($\beta = -0.1$) stable distribution for tail thickness parameter α values of 2 (solid line), 1.5 (dashed line) and 1.0 (dot-dash line).

implications for the bias properties of the robust estimator and its variance that will be elaborated in the next section.

Fig. 5 shows the MLE weights for a skewed stable distribution ($\beta = -0.1$) for the same values of α as in Fig. 4. 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. 4.

An iteratively reweighted least squares MLE is easily implemented by initiation using the ordinary least squares solution with data corresponding to the 5 per cent most extreme residuals at both distribution ends trimmed to reduce the influence of real outliers caused, for example, by lightning strikes or instrument malfunctions. The estimated residuals are fit with stable parameters using the MLE algorithm of Nolan (2001), standardized as $\left(\hat{\varepsilon}_i - \delta\right)/\gamma$ to ensure scale independence and then applied to computation of the weights in (10). The weighted least squares problem (10) is solved, the residuals are subsequently used to recompute the stable parameters and the process is repeated iteratively until the median absolute deviation from the median of the iteration residuals does not change by more than 1 per cent. This typically takes 3–9 iterations, with the number increasing as the tail thickness parameter approaches unity due to the efficiency properties of the least squares estimator. The significance of the stable fit at the final iteration may be assessed using a Kolmogorov–Smirnov or Ansari-Bradley test on the residuals. The iteratively reweighted solution has the advantage of being relatively fast, but has the disadvantage of slow convergence for $\alpha \to 1$ and failure below that value. As will be shown in Section 5, MT data sometimes exhibit tail thickness parameters that are below 1, so this 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 & Ojeda-Revah (2013) has also been implemented. As with the iteratively reweighted least squares solution, this utilizes a two stage approach initialized by the 5 per cent trimmed least squares solution, with the Nolan (2001) MLE used at the first stage to compute the stable parameters and an unconstrained non-linear multivariable function minimizer based on a trust region algorithm (Conn *et al.* 2000) utilized for the second stage. Numerical solutions for the gradient and Hessian matrix have been implemented to speed convergence. The objective function to be minimized is the negative log likelihood given by (7) preceded by a minus sign and with the stable parameter vector $\boldsymbol{\xi}$ fixed. Remote referencing is easily implemented by using the adaptation of two stage least squares introduced by Chave & Thomson (2004) and elaborated by Chave (2012), in which the transfer functions between the local magnetic field components and all of the reference ones are first estimated using the non-linear 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 one to two orders of magnitude more computer time compared to the iteratively reweighted least squares approach.

The MT processing algorithms described in this section share all of the standard optimality properties of the MLE [DuMouchel (1973); Stuart *et al.* (1999), ch. 17–18]:

(1) The solution is asymptotically consistent, meaning that it converges in probability to the true solution, or $\Pr\left(|\hat{\mathbf{z}} - \mathbf{z}| > \tau\right) \to 0$ as $N \to \infty$ for every positive τ. It is asymptotically unbiased, but not necessarily unbiased for finite N.

(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 standardized difference between the estimated and true MT response converges in distribution to $\mathcal{N}(\mathbf{0_4}, \mathcal{I}^{-1})$ as $N \to \infty$, where $\mathcal{N}$ is the multivariate Gaussian distribution, $\mathbf{0_4}$ is a vector of four zeroes and $\mathcal{I}$ is the Fisher information matrix described in the next section. This is equivalent to specifying that $\left(\hat{z}_j - z_j\right)/\hat{\sigma}_j$ converges in distribution to the univariate standardized normal distribution, where $\hat{\sigma}_j^2$ is the j-th diagonal element of $\mathcal{I}^{-1}$.

(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 variance of $\hat{\mathbf{z}}$ achieves the lower bound of the Cramér-Rao inequality, as described in the next section.

It follows directly that a conventional robust estimator cannot be in compliance with these properties since it cannot also be the MLE. In particular, violation of (1) and (2) will result in asymptotic (and probably 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 Section 5.

4 IMPROPRIETY OF MT DATA

It has become common practice within the MT community to use resampling methods such as those described by Thomson & Chave (1991) rather than parametric estimators to determine the uncertainty of the robust MT response function, in large part because parametric results are strongly influenced by the pervasive non-normality of the robust residuals. Resampling methods rapidly become computationally prohibitive with a non-linear MLE algorithm, but they are no longer needed given the excellent agreement of a stable model with MT data. However, the use of an asymptotic complex Gaussian model for the MLE first requires consideration of which type of complex Gaussian distribution to implement. It has long been a standard practice, usually implicitly, to assume that MT responses are proper, meaning that they are uncorrelated with their complex conjugates. As will be demonstrated, MT responses

are pervasively improper, and hence a more general approach to estimation of the covariance matrix is required.

A complex random vector $\mathbf{x} = \mathbf{x}_r + i\,\mathbf{x}_i$ has p 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 the complex random vector possess N data rows. Its second-order statistics are described by the covariance and pseudo-covariance (sometimes called the complementary covariance) matrices (e.g. van den Bos 1995; Picinbono 1996):

$$\begin{aligned}\mathbf{\Gamma} &= \mathcal{E}\left[(\mathbf{x}-\boldsymbol{\mu}_x)^H \cdot (\mathbf{x}-\boldsymbol{\mu}_x)\right] = \mathbf{\Gamma}_{x_r x_r} + \mathbf{\Gamma}_{x_i x_i} + \boldsymbol{i}\left(\mathbf{\Gamma}^T_{x_r x_i} - \mathbf{\Gamma}_{x_r x_i}\right)\\ \breve{\mathbf{\Gamma}} &= \mathcal{E}\left[(\mathbf{x}-\boldsymbol{\mu}_x)^T \cdot (\mathbf{x}-\mu_x)\right] = \mathbf{\Gamma}_{x_r x_r} - \mathbf{\Gamma}_{x_i x_i} + \boldsymbol{i}\left(\mathbf{\Gamma}^T_{x_r x_i} + \mathbf{\Gamma}_{x_r x_i}\right).\end{aligned} \tag{11}$$

The covariance matrix $\mathbf{\Gamma}$ is complex, Hermitian and will be assumed positive definite, while the pseudo-covariance matrix $\breve{\mathbf{\Gamma}}$ is complex and symmetric.

A complex random vector $\mathbf{x}$ is proper if $\breve{\mathbf{\Gamma}}$ is identically zero, and otherwise is improper. The conditions for propriety on the covariance matrices of the real and imaginary parts reduce to:

$$\begin{aligned}\mathbf{\Gamma}_{x_r x_r} &= \mathbf{\Gamma}_{x_i x_i},\\ \mathbf{\Gamma}_{x_r x_i} &= -\mathbf{\Gamma}^T_{x_r x_i},\end{aligned} \tag{12}$$

where the second equation in (12) requires that the diagonal elements of $\mathbf{\Gamma}_{x_r x_i}$ vanish. The complex covariance matrix for proper data is then given equivalently by the sum or difference of the two equations in (11). However, in the improper case, both $\mathbf{\Gamma}$ and $\breve{\mathbf{\Gamma}}$ are required for a complete description of the second-order statistics of a complex random vector.

Extending statistical definitions from the real to the complex case is a topic that is not without controversy. In the signal processing literature where there has been extensive recent work, the guiding approach has been that definitions and principles should be the same in the real and complex domains. Schreier & Scharf (2010) provide a comprehensive recent survey. This has led to the concept of so-called augmented variables and covariance matrices. The augmented complex random variable for $\mathbf{x}$ is given by $= (\mathbf{x} \quad \mathbf{x}^*)$, and is obtained by adding the p 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

$$\underset{\frown}{\mathbf{\Gamma}} = \begin{bmatrix} \mathbf{\Gamma} & \breve{\mathbf{\Gamma}} \\ \breve{\mathbf{\Gamma}}^* & \mathbf{\Gamma}^* \end{bmatrix} \tag{13}$$

$\underset{\frown}{\mathbf{\Gamma}}$ is block structured and Hermitian, and will be assumed positive definite.

The standard approach to estimation of the variance of an estimator is use of the Fisher score to compute the Cramér-Rao bound. The principles are covered in chapter 17 of Stuart *et al.* (1999), and its extension to complex random variables is reviewed by Schreier & Scharf (2010). Define the complex Fisher score:

$$\mathbf{s}(\boldsymbol{\psi}|\boldsymbol{\varepsilon}) = \partial_\psi \mathcal{L}(\boldsymbol{\psi}|\boldsymbol{\varepsilon}), \tag{14}$$

where $\mathcal{L}$ is the log likelihood function (7) and $\boldsymbol{\psi} = (\boldsymbol{\xi},\ \mathbf{z})$ is the parameter vector. It is straightforward to show that the Fisher score is a random variable with zero mean. The 8×8 complex Fisher information matrix is given by

$$\begin{aligned}\mathcal{I}(\boldsymbol{\psi}) &= \mathcal{E}\left\{\left[\partial_\psi \mathcal{L}(\boldsymbol{\psi}|\boldsymbol{\varepsilon})\right]^H \left[\partial_\psi \mathcal{L}(\boldsymbol{\psi}|\boldsymbol{\varepsilon})\right]\right\}\\ &= -\mathcal{E}\left\{\partial_\psi \left[\partial_\psi \mathcal{L}(\boldsymbol{\psi}|\boldsymbol{\varepsilon})\right]^H\right\}.\end{aligned} \tag{15}$$

Let $\mathbf{\Upsilon}(\mathbf{x})$ denote an estimator for the vector of parameters $\boldsymbol{\psi}$ and let its expected value vector be $\mathcal{E}[\mathbf{\Upsilon}(\mathbf{x})] = \mathbf{\Phi}(\boldsymbol{\psi})$. The covariance matrix of $\mathbf{\Upsilon}(\mathbf{x})$ is bounded by

$$cov\,[\boldsymbol{\psi}] \geq J \bullet \mathcal{I}^{-1}(\boldsymbol{\psi}) \cdot J^H, \tag{16}$$

where the first term on the right-hand side is the Jacobian matrix $\overleftrightarrow{J}$ of the estimators and parameters. Eq. (16) is the Cramér-Rao inequality when the MT responses are proper complex random variables. The Jacobian matrix may be expressed as:

$$J = \overleftrightarrow{\mathbf{I}}_8 + \partial_\psi \mathbf{b}(\mathbf{\Phi}), \tag{17}$$

where $\mathbf{b}(\mathbf{\Phi})$ is the estimator bias and $\overleftrightarrow{\mathbf{I}}_8$ is the 8×8 identity matrix. Consequently, if the estimator is unbiased so that $\partial_\psi \mathbf{b}(\mathbf{\Phi}) = \mathbf{0}$, as holds asymptotically for the MLE, then the Jacobian matrix becomes the identity matrix, and the covariance matrix for proper complex numbers is bounded by the inverse of the Fisher information matrix.

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

$$\underset{\frown}{\mathcal{I}}(\boldsymbol{\psi}) = \begin{bmatrix} \mathcal{I}(\boldsymbol{\psi}) & \breve{\mathcal{I}}(\boldsymbol{\psi}) \\ \breve{\mathcal{I}}^*(\boldsymbol{\psi}) & \mathcal{I}^*(\boldsymbol{\psi}) \end{bmatrix}. \tag{18}$$

where the Fisher pseudo-information matrix is:

$$\begin{aligned}\breve{\mathcal{I}}(\boldsymbol{\psi}) &= \mathcal{E}\left\{\left[\partial_\psi \mathcal{L}(\boldsymbol{\psi}|\boldsymbol{\varepsilon})\right]^T \left[\partial_\psi \mathcal{L}(\boldsymbol{\psi}|\boldsymbol{\varepsilon})\right]\right\}\\ &= -\mathcal{E}\left\{\partial_\psi \left[\partial_\psi \mathcal{L}(\boldsymbol{\psi}|\boldsymbol{\varepsilon})\right]^T\right\}.\end{aligned} \tag{19}$$

The augmented Jacobian matrix is

$$\underset{\frown}{J} = \begin{bmatrix} \overleftrightarrow{\mathbf{I}}_8 + \partial_\psi \mathbf{b}(\mathbf{\Phi}) \\ \partial_{\psi *} \mathbf{b}(\mathbf{\Phi}) \end{bmatrix}. \tag{20}$$

The Cramér-Rao inequality when the MT responses are improper complex random variables becomes:

$$cov\,[\mathbf{\Upsilon}(\mathbf{x})] \geq \underset{\frown}{J} \bullet \underset{\frown}{\mathcal{I}}^{-1}(\boldsymbol{\psi}) \cdot J^H. \tag{21}$$

For an unbiased estimator, $\underset{\frown}{J} = (\overleftrightarrow{\mathbf{I}}_8 \quad \overleftrightarrow{\mathbf{0}}_8)^T$ where $\overleftrightarrow{\mathbf{0}}_8$ is an 8×8 matrix of zeros. The block matrix inversion lemma yields

$$\underset{\frown}{\mathcal{I}}^{-1} = \begin{Bmatrix} \left[\mathcal{I} - \breve{\mathcal{I}}(\mathcal{I}^*)^{-1}\breve{\mathcal{I}}^*\right]^{-1} & -\mathcal{I}^{-1}\breve{\mathcal{I}}\left(\mathcal{I}^* - \breve{\mathcal{I}}^*\mathcal{I}^{-1}\breve{\mathcal{I}}\right)^{-1} \\ -\left(\mathcal{I}^* - \breve{\mathcal{I}}^*\mathcal{I}^{-1}\breve{\mathcal{I}}\right)^{-1}\breve{\mathcal{I}}^*\mathcal{I}^{-1} & \left[\mathcal{I}^* - \breve{\mathcal{I}}^*\mathcal{I}^{-1}\breve{\mathcal{I}}\right]^{-1} \end{Bmatrix}. \tag{22}$$

so that (21) reduces to:

$$cov\,(\boldsymbol{\psi}) \geq \left\{\mathcal{I}(\boldsymbol{\psi}) - \breve{\mathcal{I}}(\boldsymbol{\psi})\left[\mathcal{I}^*(\boldsymbol{\psi})\right]^{-1}\breve{\mathcal{I}}^*(\boldsymbol{\psi})\right\}^{-1}. \tag{23}$$

The right-hand side of (23) is larger (i.e. more positive definite) than $\mathcal{I}^{-1}(\boldsymbol{\psi})$, hence not accounting for impropriety will result in

underestimation of the covariance. Eq. (23) 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 independent, and hence the proper Cramér-Rao bound and (23) may be regarded as block diagonal, with the upper left block defined by 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.

Given that there are two possible approaches to estimating the covariance for MT data, it is important to be capable of statistically testing the importance of impropriety. A generalized likelihood ratio test for impropriety was proposed by Schreier *et al.* (2006) and elaborated by Walden & Rubin-Delanchy (2009). However, their impropriety hypothesis test is identical to the textbook test of independence for multivariate data (e.g. Anderson 2003, section 9.1) applied to the augmented covariance matrix (13). The null hypothesis holds that the sample version of (13) is block diagonal versus the alternate hypothesis that it is not, or equivalently, the null hypothesis holds that $\breve{\mathbf{\Gamma}} = \mathbf{0}$ versus the alternate hypothesis that it is not. The generalized likelihood ratio test statistic for independence is

$$\hat{\mathbf{\Lambda}} = \frac{\det \breve{\mathbf{\Gamma}}\,(\boldsymbol{\psi})}{\left[\det \hat{\mathbf{\Gamma}}\,(\boldsymbol{\psi})\right]^2} \tag{24}$$

and is identical to the generalized likelihood ratio statistic given by either Schreier *et al.* (2006) or Walden & Rubin-Delanchy (2009). Eq. (24) is distributed as Wilks' lambda distribution $\Lambda_{p,p,N-p-1}$ if the null hypothesis is true, where $p = 2$ for the complex MT response. Wilks' lambda distribution (Anderson 2003, section 9.3) is a multivariate generalization of the F distribution, and the test rejects if the test statistic (24) is smaller than the appropriate critical value. The Wilks' lambda null distribution for the impropriety test statistic was apparently not recognized by Schreier *et al.* (2006), who do not explicitly discuss the null distribution, or by Walden & Rubin-Delanchy (2009), although they did thoroughly investigate the null distribution through simulation and analysis.

Rao (1951) provides an F-distribution approximation to the Wilks' Λ statistic that is more accurate than the standard chi-square approximation for likelihood ratio tests. It is also exact when $p = 2$, as for MT. The likelihood ratio statistic (24) is changed to an F-test statistic through the transformation

$$\hat{F} = \frac{M(1 - \hat{\Lambda}^s)}{p^2 \hat{\Lambda}^s}, \tag{25}$$

where $s = \sqrt{(2p^2 - 5)/(p^4 - 4)} = 1/2$, and $M = (N - p - 3/2)/\mathrm{s} - (p^2 - 2)/2 = 2N - 8$. $\hat{F}$ is assessed in the usual way against the $F_{4,\,M}$ distribution.

Implementation of the Cramér-Rao bound for the MT MLE covariance matrix is straightforward. The observed Fisher information matrix $\mathcal{I}_R$ for the MT responses in (4) is given by the Hessian matrix:

$$\mathcal{I}_R = -\overleftrightarrow{\mathbf{b}}^T \bullet \mathbf{v} \bullet \overleftrightarrow{\mathbf{b}}, \tag{26}$$

where $\mathbf{v}$ is a diagonal matrix whose i-th element is $\phi''\left(\hat{\varepsilon}_i\right)$. $\mathcal{I}_R$ can be transformed into the complex Fisher information matrices $\mathcal{I}$ and $\breve{\mathcal{I}}$ using an analogous expression to (11). The proper and improper variance bounds $\mathcal{I}^{-1}$ and (23) then follow directly, and can be used to compute the generalized likelihood ratio statistic (24). Since the MLE is asymptotically efficient, it achieves the Cramér-Rao bound in the limit of large numbers of data. All other estimators, including robust estimators, must fail to reach the Cramér-Rao lower bound.

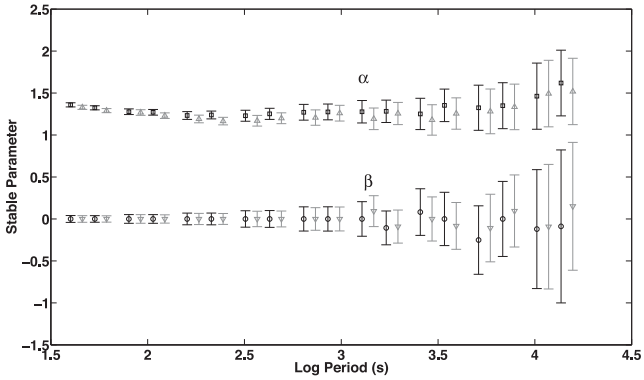

Figure 6. The tail thickness parameter α (open squares and triangles) and the skewness β (open circles and inverted triangles) for the Site 172 data, with the x-orientation error bars in black and the y-orientation error bars in grey. The latter have been offset to the right for clarity.

5 EXAMPLES

Fig. 6 shows the MLE tail thickness and skewness parameters for the Site 172 (5 s) data along with double-sided 95 per cent confidence limits after apportioning the tail probability among all four stable parameters, so that the Gaussian quantile used to define the confidence bound is 2.50 rather than 1.96. The tail thickness parameter hovers around 1.25, rising slightly at the longest and shortest periods, and is well away from the Gaussian end member except at the longest period. It is also statistically identical for both MT response orientations at each period. The skewness parameter is not different from zero within the confidence band at any period, hence leaving it free results in a nuisance parameter. Consequently, it was fixed to zero and the analysis was repeated.

The Kolmogorov–Smirnov test for goodness of fit to a stable distribution was applied to the final MLE residuals at each period. The resulting p-values typically lie between 0.8 and 0.9, and at no period for either MT orientation did any drop below 0.13. The p-value is the probability of observing data at least as far from the null value given that the null hypothesis is true, and hence these results strongly oppose rejection of the stable model. In fact, a stronger statement can be made by assuming that the p-values p_i at each period are independent and then combining them into a single composite p-value using Fisher's inverse chi-square method (Kvam & Vidakovic 2007, section 6.6). Under the null hypothesis, each p-value is uniformly distributed on [0, 1], so that $-2 \log p_i \sim \chi^2_2$, and hence by the additivity property of the chi-square distribution, $-2 \sum_{i=1}^{m} \log p_i \sim \chi^2_{2m}$, where m is the number of periods. Consequently, the combined p-value is given by $1 - F_\chi\left(-2 \sum_{i=1}^{m} \log p_i,\ 2m\right)$, where F_χ is the chi-square cdf. Using this approach yields combined p-values of 1.0000 for the two MT response components, further supporting the stable model for these MT data.

It is important to assess the significance of impropriety before considering the MT MLE response and its uncertainty. The likelihood ratio statistic for impropriety (24) and double-sided p-values computed using the F transformation (25) at each period are

Table 1. Likelihood ratio test statistic for impropriety and *p*-values for Site 172 data.

Period (s)	$\hat{\Lambda}_x$	*p*-value	$\hat{\Lambda}_y$	*p*-value
13 653	0.9962	0.0301	0.9661	0.9897
10 240	0.9942	0.0664	0.9218	0.1943
6827	0.8141	0.0000	0.9050	0.0012
5120	0.9036	0.0010	0.8608	0.0000
3413	0.7267	0.0000	0.7558	0.0000
2560	0.9120	0.0000	0.8258	0.0000
1707	0.6715	0.0000	0.6027	0.0000
1280	0.8837	0.0000	0.7190	0.0000
853	0.8993	0.0000	0.8597	0.0000
640	0.9060	0.0000	0.7823	0.0000
427	0.7762	0.0000	0.7442	0.0000
320	0.5582	0.0000	0.3463	0.0000
213	0.4485	0.0000	0.4971	0.0000
160	0.7641	0.0000	0.2223	0.0000
107	0.8484	0.0000	0.7247	0.0000
80	0.7439	0.0000	0.8769	0.0000
53	0.8878	0.0000	0.9706	0.0000
40	0.9948	0.0000	0.9269	0.0000

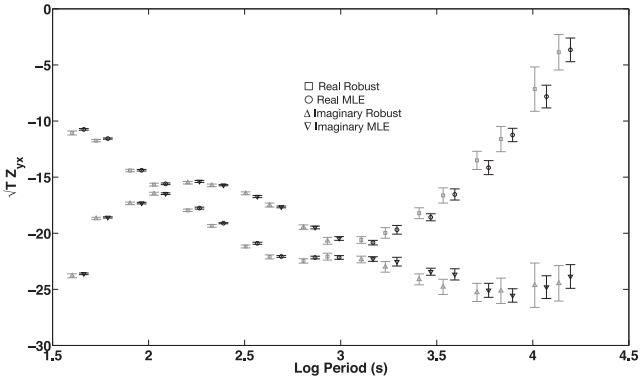

Figure 7. The real and imaginary parts of the Z_{yx} component of the MT response scaled by the square root of the period at Site 172 using Site 145 as a remote reference. The robust (stable MLE) estimates for the real parts are depicted as squares (circles) and the imaginary parts are shown as triangles (inverted triangles). The confidence limits for the robust (grey) and stable MLE (black) results are obtained using the jackknife and diagonal elements of the improper Fisher information matrix, respectively, assuming that the 0.05 tail probability is apportioned equally among the eight real and imaginary elements of the MT response tensor. The stable MLE estimates have been offset to the right for clarity.

presented in Table 1. For most periods in both orientations of the MT response, the *p*-value is nearly zero and hence the null hypothesis that the MT responses are proper is strongly rejected. The only exceptions are at the longest periods. Consequently, impropriety is pervasive, and must be accounted for when estimating the covariance of the MT response. As a result, it is necessary to use the diagonal elements of the improper Cramér-Rao variance bound (23) rather than the proper one $\mathcal{I}^{-1}$ to estimate the confidence limits on the MLE MT response.

Fig. 7 compares the Z_{yx} component (scaled by the square root of period for clarity) of robust estimates for the MT response using the algorithm of Chave & Thomson (2004) and stable MLE estimates using the non-linear optimization algorithm described in Section 3. The number of data ranges between 100 and 24 380 from the longest to the shortest period. Uncertainty estimates were obtained using the jackknife for the robust response and the diagonal elements of the improper Cramér-Rao bound (23) for the MLE response,

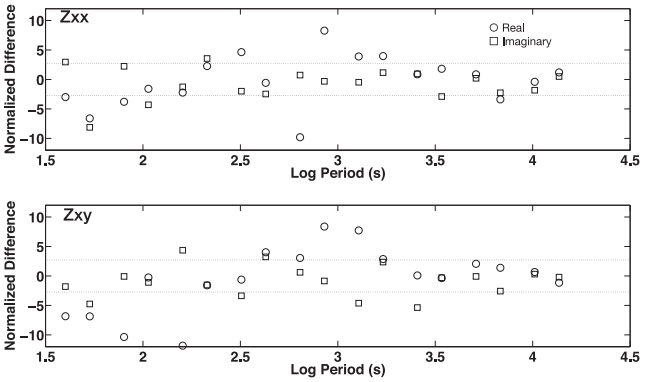

Figure 8. The normalized differences between the stable MLE and robust estimates normalized by the stable MLE standard deviation at Site 172 as a function of period for the *x*-orientation. The top panel shows the Z_{xx}component and the bottom panel shows the Z_{xy} component, with the real and imaginary parts depicted by squares and circles, respectively. The horizontal dashed lines represent ±2.73 standard deviations within which 95 per cent of the data should lie if they are Gaussian.

respectively, and utilize the Gaussian quantile 2.73 corresponding to apportioning the 0.05 tail probability among all eight real and imaginary MT response components at each period. The difference between the proper and improper confidence bounds on the MLE amounts to a few percent for this example. It is obvious that the stable MLE confidence bounds are systematically smaller than the robust ones, as is expected because the robust estimator cannot achieve the Cramér-Rao lower bound due to its lack of MLE optimality. However, there also are frequent differences between the two types of estimates that represent bias of one of the estimators, especially at periods below 1000 s. As will be elaborated in Section 6, the bias should be attributed to the robust estimator.

The differences seen in Fig. 7 may appear to be subtle, but they are not. Figs 8 and 9 present the difference between the stable MLE and robust estimates normalized by the stable MLE standard deviation for all four elements of the *x*- and *y*-orientations of the MT response tensor, respectively. Consequently, the ordinate units are in MLE standard deviations; the dashed horizontal lines represent ±2.73 standard deviations within which 95 per cent of the data (or

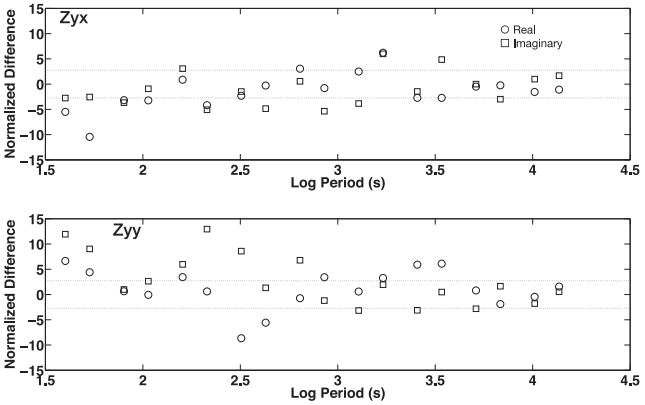

Figure 9. The normalized differences between the stable MLE and robust estimates normalized by the stable MLE standard deviation at Site 172 as a function of period for the *y*-orientation. The top panel shows the Z_{yx} component and the bottom panel shows the Z_{yy} component, with the real and imaginary parts depicted by squares and circles, respectively. The horizontal dashed lines represent ±2.73 standard deviations within which 95 per cent of the data should lie if they are Gaussian.

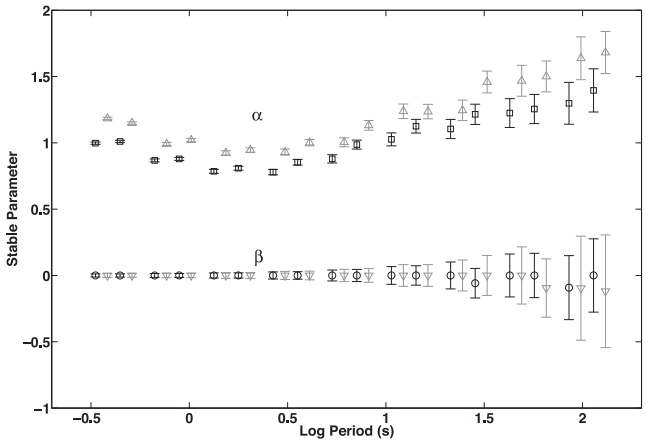

Figure 10. The tail thickness parameter α (open squares and triangles) and the skewness β (open circles and inverted triangles) for the Site 013 data, with the x-orientation error bars in black and the y-orientation error bars in grey. The latter have been offset to the right for clarity.

34 values per panel) should be situated for Gaussian variates. There are 36 real and imaginary parts in each panel of Figs 8 and 9, but 14, 14, 17 and 15 for the Z_{xx} through Z_{yy} elements actually lie outside the dashed horizontal lines. Further, the differences are as large as 12 and 14 standard deviations in Figs 8 and 9, respectively. Such large differences are extremely unlikely if all of the entities used in the figures are Gaussian. This conclusion is unchanged if the pooled (rms of the robust and MLE) standard deviation is used, although the size of discrepancy is reduced because the robust standard deviation is larger than the MLE one.

Fig. 10 shows the stable parameters α and β for the Site 013 (24 Hz) data; note that the abscissa is different than in earlier figures. In contrast to the results in Fig. 6, the tail thickness parameter is systematically lower for the x- as compared to the y-orientation, and in addition drops below unity for intermediate periods, reaching as low as ∼0.8 in the x-orientation at short periods. As for the Site 172 data, the skewness is statistically indistinguishable from zero at all periods, and was consequently set to zero for further analysis.

The goodness of fit of the final residuals after minimization of the objective function was assessed using the Kolmogorov–Smirnov test against the final stable distribution. For the x-orientation, the tests reject the null hypothesis that the data are stably distributed at 2/3 of the periods, while for the y-orientation, rejection occurs about 1/3 of the time. However, a non-parametric Ansari-Bradley test of the dispersion of the residuals against random draws from the final fitted stable distribution is consistent with the null hypothesis at all periods for the x-orientation and all but one for the y-orientation. Fig. 11 shows variance equalized p-p plots for the x-orientation at periods of 0.89 and 3.6 s, where the Kolmogorov–Smirnov statistic resulted in rejection of the null hypothesis. At 3.6 s (top panel), the misfit is due to the residual distribution being slightly long tailed compared to the stable distribution fit to it, and at 0.89 s, this behaviour is attenuated but still present. In addition, the very large number of data available at site 013 (between 580 and 140 360 from the longest to the shortest period) means that very subtle inconsistencies of the residuals with the best fit stable distribution may result in Kolmorogov-Smirnov goodness of fit test rejection. The Ansari-Bradley test is less sensitive to such weak departures. It is clear from Fig. 11 that the data are fairly consistent with the stable model, as reinforced by the Ansari-Bradley test results.

The likelihood ratio statistics for impropriety (24) and double-sided p-values computed using the F transformation (25) at each

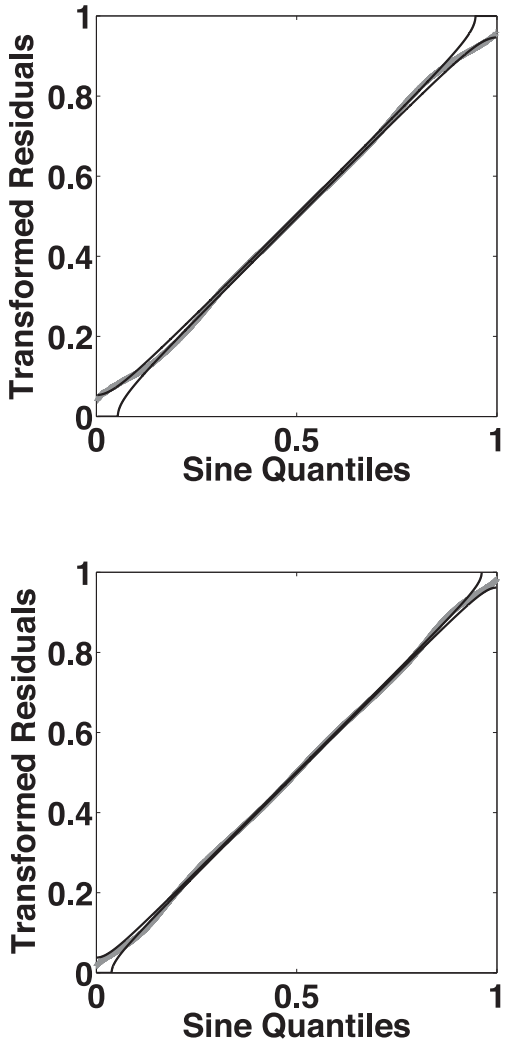

Figure 11. Variance equalized percent-percent plots for the Site 013 data at 0.89 (bottom) and 3.6 (top) s for the x-orientation, with the ordinate being the sine-transformed alpha stable cumulative distribution applied to the standardized final residuals. The stable distribution parameters α and β are (0.88, 0.00) and (0.85, 0.00), respectively; see Section 2 for discussion.

period were computed for the site 013 data. For the x-orientation, the null hypothesis that the result is proper is only accepted at the longest period (114 s), while for the y-orientation, impropriety is required at all periods. Consequently, further analysis incorporated the improper Cramér-Rao bound (23). The differences between the improper and proper confidence bounds amount to a few percent at most periods.

Fig. 12 compares the robust and non-linear MLE Z_{xy} components scaled by the square root of period for site 013. Uncertainty estimates were obtained as for the Site 172 data. As with the site 172 data, and as expected from the optimality properties of the MLE, the stable MLE confidence bands are systematically smaller than the robust ones. In addition, large systematic bias is evident, particularly between 3 and 10 s for the imaginary part.

Figs 13 and 14 present the differences between the stable MLE and robust estimates normalized by the stable MLE standard deviation for all four real elements of the x- and y-orientations of the Site 013 MT response tensor under identical conditions to those used for Figs 8 and 9. There are 36 real and imaginary parts in each panel, but 22, 20, 20 and 23 for the Z_{xx} through Z_{yy} elements are outside

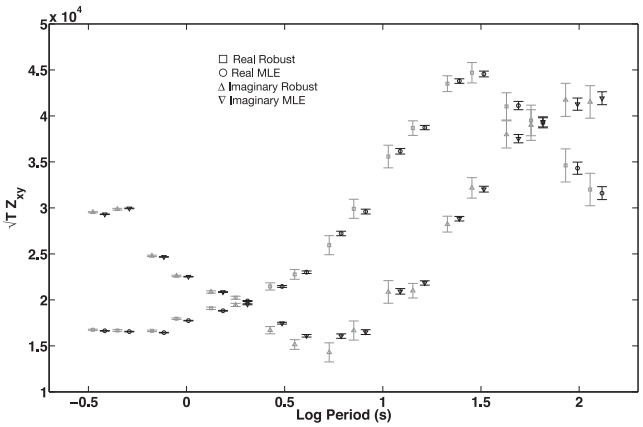

Figure 12. The real and imaginary parts of the Z_{xy} component of the MT response scaled by the square root of the period at Site 013 using Site 014 as a remote reference. The robust (stable MLE) estimates for the real parts are depicted as squares (circles) and the imaginary parts are shown as triangles (inverted triangles). The confidence limits for the robust (grey) and stable MLE (black) results are obtained using the jackknife and diagonal elements of the improper Fisher information matrix, respectively, assuming that the 0.05 tail probability is apportioned equally among the eight real and imaginary elements of the MT response tensor. The stable MLE estimates have been offset to the right for clarity.

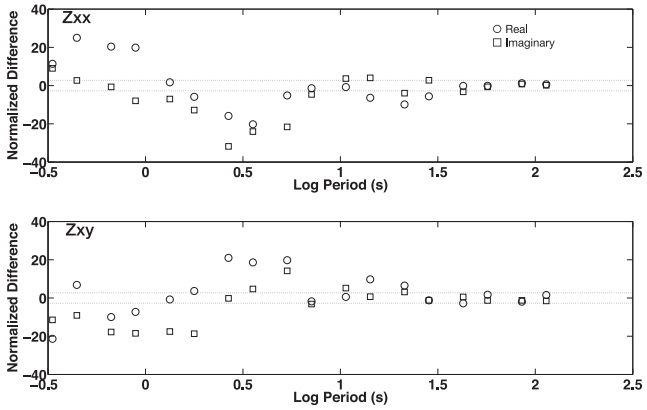

Figure 13. The normalized differences between the stable MLE and robust estimates normalized by the stable MLE standard deviation at Site 013 as a function of period for the *x*-orientation. The top panel shows the Z_{xx} component and the bottom panel shows the Z_{xy} component, with the real and imaginary parts depicted by squares and circles, respectively. The horizontal dashed lines represent ± 2.73 standard deviations within which 95 per cent of the data should lie if they are Gaussian.

the dashed horizontal lines that represent ± 2.73 standard deviations within which 95 per cent of the data should lie for Gaussian variates. Further, the differences are as large as 35 standard deviations. The results are more extreme than for the Site 172 data (see Figs 8 and 9), and are indicative of strongly non-Gaussian behaviour for one or more of the statistical entities.

6 DISCUSSION

It is a standard result of statistical theory that the MLE is asymptotically unbiased, but may be biased for a finite data set. The finite sample bias is $O(1/N)$ (Stuart *et al.* 1999, section 18.14), and hence is typically negligible for the sample sizes used in MT, with the possible exception of the longest periods in cases where the data collection interval is short. The MLE is also asymptotically

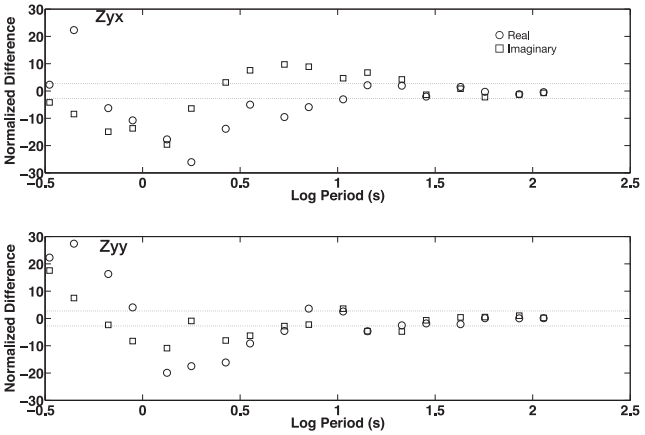

Figure 14. The normalized differences between the stable MLE and robust estimates normalized by the stable MLE standard deviation at Site 013 as a function of period for the *y*-orientation. The top panel shows the Z_{yx} component and the bottom panel shows the Z_{yy} component, with the real and imaginary parts depicted by squares and circles, respectively. The horizontal dashed lines represent ± 2.73 standard deviations within which 95 per cent of the data should lie if they are Gaussian.

Gaussian, with the usual caveat that the number of data required to reach the asymptotic limit is difficult to quantify. A qualitative demonstration of consistency with the Gaussian limit can be achieved by examining bootstrap distributions for the MT response functions. An approximate bootstrap was implemented for the examples of the previous section by sampling 10 000 times with replacement from the data and MLE weights at each period, and then computing (10). Fig. 15 shows a typical result, and is nearly indistinguishable from the best fitting Gaussian distribution, giving confidence that the use of an asymptotic Gaussian model is appropriate. The results of Fig. 15 are typical for both Sites 172 and 013 except at the two longest periods at the former. Further, the bias can be estimated from the bootstrap samples, and is typically under 1 per cent. Consequently, the MLE estimator of this paper is nearly unbiased and has approximate Gaussian second-order statistics for typical MT sample sizes. Bootstrap distributions for the Chave & Thomson (2004) robust estimator whose weight is shown in Fig. 4 are not well fit by a Gaussian distribution and have a greater spread. This reflects overemphasis of data at the edges of the robust weight passband in Fig. 4 whose stable form results in too much variability, and hence incorrect confidence bounds. This effect increases as α decreases because the stable weight function falls more rapidly with distance from the distribution centre, and hence the robust estimator incorporates a greater fraction of extreme data.

Figs 8–9 and 13–14 show that the normalized deviations between the MLE and robust estimators exhibit far more frequent and substantially larger differences than would be expected for a Gaussian model. Given the demonstrated consistency of the MLE solution with a Gaussian statistical model, this non-Gaussian behaviour must be attributed to the robust estimator. Fig. 1 demonstrates that the residuals from a MT robust estimator are systematically long-tailed compared to Gaussian expectations, and the real and imaginary parts of their residuals can consistently be fit by a truncated stable distribution. This suggests that 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 frequently biased and exhibit systematic non-Gaussian behaviour, but whose apparent variances computed using standard methods are systematically larger than the true value,

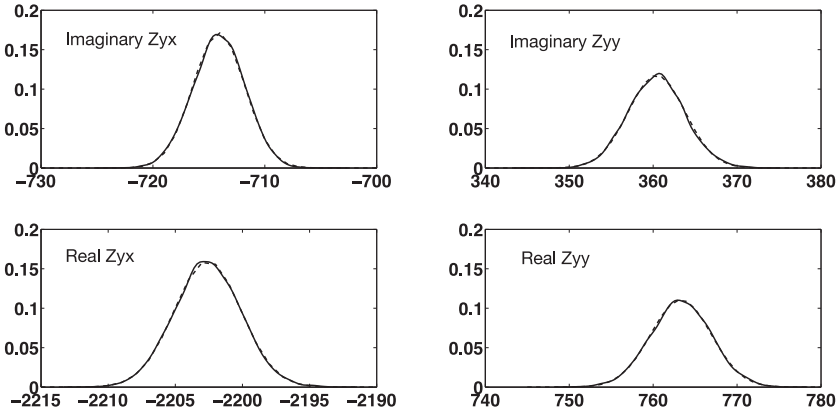

Figure 15. Kernel density estimates (solid lines) for the bootstrap distributions of the MT response functions for the *y*-orientation at Site 013 at a period of 14 s resulting from 10 000 random draws with replacement from the data and MLE weights. From the lower left proceeding clockwise, the panels show the real and imaginary parts of Z_{yx} and the imaginary and real parts of Z_{yy}. The dotted lines that are nearly indistinguishable from the kernel density estimates are Gaussian distributions with means and standard deviations estimated from the bootstrap replicates.

which masks the bias to some degree. The bias will rise as α decreases because the robust weights include a larger fraction of stable data in the pass-band. This is readily apparent in Figs 12 and 13, where the long period data correspond to the largest tail thickness parameter and display the smallest differences between the robust and MLE estimates. By contrast, the differences at periods shorter than 10 s, where the tail thickness parameters are small, display large discrepancies.

It is common practice in MT inversion to place an error floor (either absolute or as a fraction of the response) on the response functions, in effect replacing the observed statistical errors with a typically larger value on the presumption that this allows for the influence of electromagnetic distortion and unresolvable small-scale electrical structures. For example, Khoza *et al.* (2013) used final 5 and 10 per cent error floors for phase and apparent resistivity, respectively, for 2-D inversion and 10 and 15 per cent error floors for the diagonal and anti-diagonal elements of the response tensor, respectively, for 3-D inversion of data from South Africa. Tietze & Ritter (2013) used smaller error floors of 5 and 3 per cent for the diagonal and anti-diagonal elements for 3-D data from the San Andreas Fault in central California. However, the demonstration of frequent bias for robust MT response estimates implies that, at least in part, increasing the size of the error estimates compensates for bias that was previously unrecognized. How important this is can only be evaluated through reanalysis of data using the MLE and subsequent re-inversion. Further, whether the lower bias of the MLE results with their statistically defensible albeit smaller error estimates will improve the ability of MT to image earth structure remains to be established.

A comparison of the improper to the proper variances for the exemplar data in this paper shows that the improper variance is systematically larger, as expected. The differences in standard deviation typically amount to only a few percent, so for these examples the influence of impropriety on the MT response confidence limits is insignificant. However, for other data sets, variance differences of as much as 25 per cent have been observed, so impropriety cannot be dismissed as an issue. It must be emphasized that the test for impropriety (24) applies to the structure of the full augmented covariance matrix rather than just to its diagonal elements, so rejection of propriety does not automatically entail a substantial change to the variance. In any case, including impropriety through (23) rather than just $\mathcal{I}^{-1}$ represents a trivial addition to MT response computation, and hence it is recommended that it be done routinely. In addition, the improper covariance bound reduces to the proper one when $\breve{\mathcal{I}} = \mathbf{0}$, so no error will ensue from using it in the event that the MT response is proper.

The presence of stably distributed data is intimately intertwined with governing physics that contains fractional rather than integer derivatives. Spatial and temporal fractional derivatives reflect the existence of long range ordering in the relevant domain. Meerschaert (2012) summarizes the arguments. The pre-Maxwell equations that govern MT have been known for about 150 yr, and there is an extensive base of theoretical and empirical evidence against the existence of fractional derivatives in them. This does not necessarily apply to the constitutive relations in some media, although there is again a strong base of empirical evidence in support of a linear tensor form for Ohm's Law for earth materials. However, some of the non-linear and non-equilibrium processes that occur in the ionosphere and magnetosphere where the MT source fields originate do display evidence for fractional derivative and multifractal behaviour along with self-organized criticality (e.g. Consolini 2002; Consolini *et al.* 2005; Anh *et al.* 2007; Zaslavsky *et al.* 2007; Yu *et al.* 2009, 2010), and very likely account for the pervasive stable probability structure of MT data documented in this paper.

For many years, MT practitioners (including the present author) have maintained that MT data are inherently non-stationary. There are known relationships for the skewness, scale and location parameters of ensembles of stable random variables with different values for these parameters (e.g. Samorodnitsky & Taqqu 1994), but not for random variables with varying tail thickness parameters. It can be demonstrated through simulation that mixtures of stable random variables with different tail thickness parameters remain stably distributed with a new value for α. As an example, 10 000 random draws from the uniform distribution over the range [0.6, 2] will specify the tail thickness parameters for individual random draws from a standardized symmetric stable distribution. The resulting ensemble has a tail thickness parameter of 1.16 with the remaining stable parameters unchanged. Fig. 16 shows a variance equalized p-p plot of the result; the Kolmogorov–Smirnov *p*-value is 0.56, so the null hypothesis that the ensemble is stably distributed is strongly accepted. Similar results are obtained with non-symmetric stable random variables. MT data in the frequency domain can be viewed as mixtures of stable random variables with varying parameters; as data are added or removed, the ensemble stable parameters will

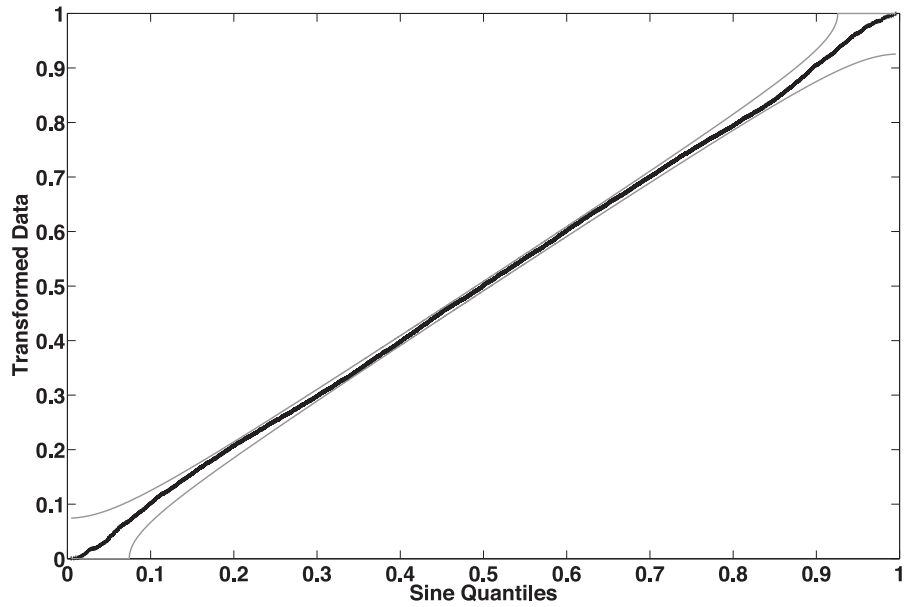

Figure 16. Variance equalized percent-percent plot of 10 000 random draws from a standardized symmetric stable distribution whose tail thickness parameter is a random draw from the uniform distribution with a range of [0.6, 2.0]. The grey lines are the 95 per cent confidence band computed using the critical value of the Kolmogorov–Smirnov statistic.

change, but the data remain approximately stable. Consequently, the high variability and frequent large excursions from the mean of MT data can be explained by the characteristics of stable data, and there is no need to invoke non-stationarity to explain their behaviour.

The methods developed in this paper represent a proof of concept, and certainly there is considerable potential for improvements in solving the MLE equations for the MT responses. First, the two-stage solution that was utilized should be replaced by simultaneous estimation of the stable parameters and MT responses. It must be recognized that the two-stage approach implicitly neglects any covariance between the stable and MT response parameters by treating the stable and MT Fisher information matrices separately. DuMouchel (1975) and Nolan (2001) note that the Fisher information matrix for the stable parameters is not block diagonal in general, as every parameter is correlated with every other parameter. Whether this characteristic extends to correlation between the stable and MT response parameters remains to be determined, but can only be evaluated through a combined MLE.

Secondly, the trust region algorithm used to minimize the negative log likelihood is one of several possible approaches, but is the only one that has been tested. It is possible that more rapid convergence or greater accuracy can be achieved with an alternate method. A key issue is determining the smoothness of the MT MLE objective function in the vicinity of its minimum to help in selecting the best algorithm. An alternate approach is numerical solution of the non-linear set of MLE eq. (8) rather than minimization of the log likelihood.

Thirdly, while the stable MLE solution is inherently robust, it remains unclear whether it can handle extensive true outliers such as those caused by instrument problems or lightning strikes. A naïve extension of the robust model to one in which the core of the data is stable rather than Gaussian and contaminated by a fraction of outlying data will allow the standard methods developed in Chave *et al.* (1987) to be utilized. This is easily implemented, and will need to be tested on real or simulated data with impulsive outliers to determine its effectiveness. However, given that a stable population inherently exhibits frequent data that are far from the distribution centre, outlier detection becomes a more difficult problem. Outlier detection in stable populations has received some attention in the financial statistics literature where stable distributions are very important. For example, Schluter & Trede (2008) treated the problem of multiple outlier detection in heavy tailed populations through outward testing using the k most extreme order statistics. It would be straightforward to implement an iterative outlier detection and removal scheme using the Schluter & Trede approach by estimating the MLE stable parameters, culling outliers one by one using their algorithm, and re-estimating the stable parameters at each step. Whether this is a better approach than the naïve one would have to be determined experimentally.

Finally, the MLE estimator does not solve the correlated cultural noise problem that plagues MT response estimation. Existing tools such as the two-stage method of Chave & Thomson (2004) are immediately applicable, but require that one or more clean reference sites are available. An alternate approach is the multivariate latent variable estimator of Egbert (1997) that, under some circumstances, can separate the natural source MT and cultural signals. However, this estimator is based on the incorrect robust model for MT data, and needs modification to account for their stable nature. Independent component analysis (ICA) uses an analogous latent variable model, except that the expansion basis is assumed to be statistically independent rather than uncorrelated, hence generalizes to include all of the cross dependencies and thus encompasses non-Gaussian data (Hyvärinen & Oja 2000; Hastie *et al.* 2008, section 14.7.2). The application of ICA to stable data is described by Sahmoudi *et al.* (2004) and Wang *et al.* (2009), but its use in MT remains to be investigated.

7 CONCLUSIONS

This paper constitutes a re-examination of the assumptions that underlie robust estimation of the MT response tensor, showing that the standard robust model of a Gaussian core contaminated by a fraction of outlying data is pervasively incorrect. Instead, MT data

are well described by the family of alpha stable distributions whose tails are algebraic rather than exponential, so that their variance, and in some cases mean, are undefined (i.e. infinite). The key tail thickness, skewness, scale and location parameters of the stable distribution family were specified.

The implementation of a MLE for the MT response that exploits the stable statistical nature of the data has been described. An iteratively reweighted form was used to emphasize the differences between the stable MLE and conventional robust estimators. The weights for the former are based on a stable distribution fit to the differences between the predicted and observed electric fields (i.e. the residuals), while robust estimators are based on an *a priori* robust model that was shown to be incorrect. The stable MLE weights result in an estimator that is inherently robust. A general MLE using an unconstrained non-linear multivariable function minimizer applied to the negative log likelihood function was elucidated. The consistency, Gaussianity and efficiency optimality characteristics of the MLE were established.

Proper and improper complex random variables were defined, with proper ones being those that are uncorrelated with their complex conjugates. Proper and improper complex Gaussian distributions were described, and the Fisher information matrix for each was derived from first principles. The Cramér-Rao lower bounds for the covariance matrix of proper and improper complex random variables were specified, and a generalized likelihood ratio test for propriety and its null distribution were established based on standard multivariate statistical methods.

These concepts were applied to exemplar long period and broadband MT data sets from South Africa. Both data sets are shown to be stably distributed with tail thickness parameters lying between $\sim$0.8 and $\sim$1.5, establishing their infinite variance, non-Gaussian nature. Kolmogorov–Smirnov and Ansari-Bradley goodness of fit tests were used to establish the significance of the stable model for the data. Non-linear MLE solutions for the MT response functions were obtained and compared to a robust estimator. The likelihood ratio test for propriety was applied, showing pervasive impropriety and necessitating the use of the improper Cramér-Rao bound to define the MLE covariance matrices. The MLE standard deviations were shown to be systematically smaller than the robust ones because the former achieves the Cramér-Rao lower bound that the latter cannot reach.

The differences between the MLE and robust solutions standardized to the MLE standard deviations were shown to exhibit extensive and large differences that suggest pervasive bias and non-Gaussianity. The MLE was shown to have low bias for the numbers of data typically used for MT estimation, and the bootstrap was used to further constrain MLE bias and show that the asymptotic Gaussian limit is typically reached with these data. Consequently, the bias is ascribed to the robust estimator whose statistics are shown to be described by a truncated stable distribution, accounting in large part for the large, frequent excursions observed in the standardized differences.

The establishment of a stable model for MT data is synonymous with the existence of fractional derivative or fractal processes in the governing physics that must originate in the magnetospheric/ionospheric source region. Further, the stable model explains the high variability and frequent large excursions from the mean that characterize MT data, and hence recourse to non-stationarity to explain their behaviour is not required. Finally, a number of improvements to the MLE algorithm were proposed as future work.

ACKNOWLEDGEMENTS

This work was supported by NSF grant EAR0809074. The author is grateful to Alan Jones and Xavier Garcia for providing the data used in this paper. Andreas Junge and an anonymous reviewer provided helpful comments. The STABLE software package originated by Robust Analysis at www.RobustAnalysis.com was used in this work.

REFERENCES

Anderson, T.W., 2003. *An Introduction to Multivariate Statistical Analysis,* 3rd edn, 721 pp., John Wiley and Sons.

Anh, V., Yu, Z.-G. & Wanliss, J., 2007. Analysis of global geomagnetic variability, *Nonlin. Proc. Geophys.,* **14,** 701–708.

Ansari, A.R. & Bradley, R.A., 1960. Rank-sum tests for dispersions, *Ann. Math. Stat.,* **31,** 1174–1189.

Chave, A.D., 2012. Estimation of the magnetotelluric response function, in *The Magnetotelluric Method: Theory and Practice,* pp. 165–218, eds Chave, A.D. & Jones, A.G., Cambridge Univ. Press.

Chave, A.D. & Thomson, D.J., 1989. Some comments on magnetotelluric response function estimation, *J. geophys. Res.,* **94,** 14 215–14 225.

Chave, A.D. & Thomson, D.J., 2004. Bounded influence estimation of magnetotelluric response functions, *Geophys. J. Int.,* **157,** 988–1006.

Chave, A.D., Thomson, D.J. & Ander, M.E., 1987. On the robust estimation of power spectra, coherences and transfer functions, *J. geophys. Res.,* **92,** 633–648.

Conn, N.R., Gould, N.I.M. & Toint, P.L., 2000. *Trust Region Methods,* 959 pp., SIAM & MPS.

Consolini, G., 2002. Self-organized criticality: a new paradigm for the magnetotail dynamics, *Fractals,* **10,** 275–283.

Consolini, G., Kretzschmar, M., Lui, A.T.Y., Zimbardo, G. & Macek, W.M., 2005. On the magnetic field fluctuations during magnetospheric tail current disruption: A statistical approach, *J. geophys. Res.,* **110,** A07202, doi:10.1029/2004JA010947.

Davison, A.C. & Hinkley, D.V., 1997. *Bootstrap Methods and Their Application,* Cambridge Univ. Press, 582 pp.

DuMouchel, W.H., 1973. On the asymptotic normality of the maximum likelihood estimate when sampling from a stable distribution, *Ann. Stat.,* **1,** 948–957.

DuMouchel, W.H., 1975. Stable distributions in statistical inference: 2, information from stably distributed samples, *J. Am. Stat. Assoc.,* **70,** 386–393.

Egbert, G.D., 1997. Robust multiple station magnetotelluric data processing, *Geophys. J. Int.,* **130,** 475–496.

Egbert, G.D & Booker, J.R., 1986. Robust estimation of geomagnetic transfer functions, *Geophys. J. R. Astr. Soc.,* **87,** 173–194.

Feller, W.E., 1971. *An Introduction to Probability Theory and Its Applications,* 2nd edn, John Wiley and Sons, 669 pp.

Hastie, T., Tibshirani, R. & Friedman, J., 2008. *The Elements of Statistical Learning,* 2nd edn, Springer-Verlag, 745 pp.

Hyvärinen, A. & Oja, E., 2000. Independent component analysis: algorithms and applications, *Neural Networks,* **13,** 411–430.

Khoza, T.D., Jones, A.G., Muller, M.R., Evans, R.L., Miensopust, M.P. & Webb, S.J., 2013. Lithospheric structure of an Archean craton and adjacent mobile belt revealed from 2-D and 3-D inversion of magnetotelluric data: example from southern Congo craton in northern Namibia, *J. geophys. Res. Solid Earth,* **118,** 4378–4397.

Kogon, S.M. & Williams, D.B., 1998. Characteristic function based estimation of stable distribution parameters, in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications,* pp. 311–338, eds Adler, R.J., Feldman, R.E. & Taqqu, M.S., Birkhäuser.

Koutrouvelis, I.A., 1980. Regression-type estimation of the parameters of stable laws, *J. Am. Stat. Assoc.,* **75,** 918–928.

Kvam, P.H. & Vidakovic, B., 2007. *Nonparametric Statistics with Applications to Science and Engineering,* John Wiley and Sons, 420 pp.

Lévy, P.P., 1925. *Calcul des Probabilités,* Gauthier-Villars, 350 pp.
McCulloch, J.H., 1986. Simple consistent estimators of stable distribution parameters, *Comm. Stat. Simul. Comp.,* **15,** 1109–1136.
McCulloch, J.H., 1998. Linear regression with stable disturbances, in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications,* pp. 359–378, eds Adler, R.J., Feldman, R.E. & Taqqu, M.S., Birkhäuser.
Meerschaert, M.M., 2012. Fractional calculus, anomalous diffusion and probability, in *Fractional Dynamics,* pp. 265–284, eds Lim, S.C., Klafter, J. & Metler, R., World Scientific Press.
Michael, J.R., 1983. The stabilized probability plot, *Biometrika,* **70,** 11–17.
Nolan, J.P., 1997. Numerical calculation of stable densities and distribution functions, *Comm. Stat. Stoch. Mod.,* **13,** 759–774.
Nolan, J.P., 1998. Parameterizations and modes of stable distributions, *Stat. Prob. Lett.,* **38,** 187–195.
Nolan, J.P., 2001. Maximum likelihood estimation and diagnostics for stable distributions, in *Lévy Processes: Theory and Applications,* pp. 379–400, eds Barndorff-Nielsen, O.E., Mikosch, T. & Resnick, S.I., Birkhäuser.
Nolan, J.P. & Ojeda-Revah, D., 2013. Linear and nonlinear regression with stable errors, *J. Econometrics,* **172,** 186–194.
Picinbono, B., 1996. Second order complex random vectors and normal distributions, *IEEE Trans. Sig. Proc.,* **44,** 2637–2640.
Rao, C.R., 1951. An asymptotic expansion of the distribution of Wilks' criterion, *Bull. Inter. Stat. Inst.,* **33,** 177–180.
Sahmoudi, M., Abed-Meraim, K. & Benidir, M., 2004. Blind separation of heavy-tailed signals using normalized statistics, in *Lecture Notes in Computer Science 3195,* pp. 113–120, eds Puntonet, C.G. & Prieto, A., Springer-Verlag.
Samorodnitsky, G. & Taqqu, M., 1994. *Stable Non-Gaussian Random Processes*, Chapman & Hall, 632 pp.
Schluter, C. & Trede, M., 2008. Identifying multiple outliers in heavy-tailed distributions with an application to market crashes, *J. Empir. Finance,* **15,** 700–713.
Schreier, P.J. & Scharf, L.L., 2010. *Statistical Signal Processing of Complex-Valued Data*, Cambridge Univ. Press, 309 pp.
Schreier, P.J., Scharf, L.L. & Hanssen, A., 2006. A generalized likelihood ratio test for impropriety of complex signals, *IEEE Sig. Proc. Lett.,* **13,** 433–436.
Sims, W.E., Bostick, F.X. & Smith, H.W., 1971. The estimation of magnetotelluric impedance tensor elements from measured data, *Geophysics,* **36,** 938–942.
Stuart, A., Ord, J.K. & Arnold, S., 1999. *Kendall's Advanced Theory of Statistics, vol 2A: Classical Inference and the Linear Model,* Arnold, 885 pp.
Thomson, D.J. & Chave, A.D., 1991. Jackknife error estimates for spectra, coherences, and transfer functions, in *Advances in Spectral Analysis and Array Processing,* vol. 1, pp. 58–113, ed Haykin, S., Prentice-Hall.
Tietze, K. & Ritter, O., 2013. Three-dimensional magnetotelluric inversion in practice – the electrical conductivity structure of the San Andreas Fault in Central California, *Geophys. J. Int.,* **195,** 130–147.
Uchaikin, V.V. & Zolotarev, V.M., 1999. *Chance and Stability,* VSP Press, 570 pp.
Van den Bos, A., 1995. A multivariate complex normal distribution-a generalization, *IEEE Trans. Inform. Theory,* **41,** 537–539.
Walden, A.T. & Rubin-Delanchy, P., 2009. On testing for impropriety of complex-valued Gaussian vectors, *IEEE Trans. Sig. Proc.*, **57,** 825–834.
Wang, B., Kuruoglu, E.E. & Zhang, J., 2009. ICA by maximizing non-stability, in *Lecture Notes in Computer Science 5441,* pp. 179–186, eds Adali, T. *et al.*, Springer-Verlag.
Yu, Z.-G., Anh, V. & Eastes, R., 2009. Multifractal analysis of geomagnetic storm and solar flare indices and their class dependence, *J. geophys Res.,* **114,** A05214, doi:10.1029/2008JA013854.
Yu, Z.-G., Anh, V., Wang, Y., Mao, D. & Wanliss, J., 2010. Modeling and simulation of the horizontal component of the geomagnetic field by fractional stochastic differential equations in conjunction with empirical mode decomposition, *J. geophys. Res.,* **115,** A102019, doi:10.1029/2009JA015206.
Zaslavsky, G.M., Guzdar, P.N., Edelman, M., Sitnov, M.I. & Sharma, A.S., 2007. Self similarity and fractional kinetics of solar wind-magnetosphere coupling, *Phys. A: Stat. Mech. Appl.,* **373,** 11–20.